Обучение распознаванию образов по методу опорных векторов:

Исходная идея, вероятностная интерпретация, алгоритмы

В.В. Моттль

Вычислительный центр РАН
Московский физико-технический институт
Тульский государственный университет

Типовая задача восстановления закономерностей в множествах объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$.

Некоторое множество значений скрытой характеристики объектов $y \in Y$.

Объективно существующая скрытая функция $y(\omega): \Omega \to Y$.

Желание наблюдателя:

Иметь инструмент оценивания скрытой характеристики для реальных объектов

$\hat{y}(\omega): \Omega \to Y; \quad \hat{y}(\omega) \neq y(\omega)$ – ошибка.

Обучение по прецедентам:

Подмножество наблюдаемых объектов, для которых измерено значение функции

$\Omega^* = \ (\omega_j, y_j),\ j = 1, \ldots, N \ ,\ y_j = y(\omega_j)$.

Задача: Продолжить функцию на все множество Ω, так чтобы можно было в дальнейшем оценивать значение рассматриваемой характеристики $\hat{y}(\omega)$ для новых объектов $\omega \in \Omega \setminus \Omega^*$.

Типовая задача восстановления закономерностей в множествах объектов реального мира

Некоторое множество реально существующих объектов $\omega \in \Omega$.

Некоторое множество значений скрытой характеристики объектов $y \in Y$.

Объективно существующая скрытая функция $y(\omega): \Omega \to Y$.

Желание наблюдателя:

Иметь инструмент оценивания скрытой характеристики для реальных объектов

$\hat{y}(\omega): \Omega \to Y$; $\hat{y}(\omega) \neq y(\omega)$ – ошибка.

Простейшие случаи:

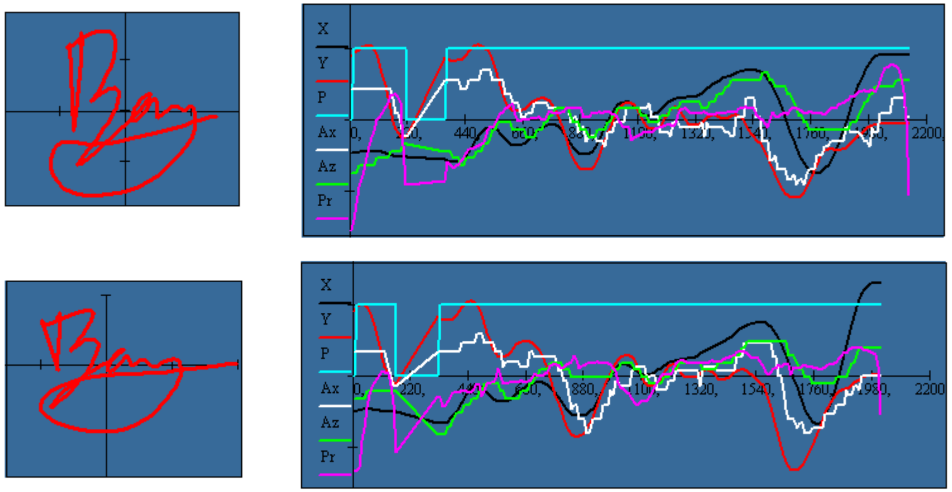
Задача восстановления числовой зависимости
$Y = \mathbb{R}$ – множество действительных чисел.

Задача распознавания образов
$Y = \{y_1, ..., y_m\}$ – конечное неупорядоченное множество.

$Y = \{-1, 1\}$ – в частности, если различению подлежат два класса объектов

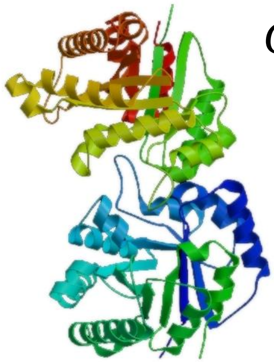
Практические задачи распознавания образов

Распознавание подписи в процессе ввода

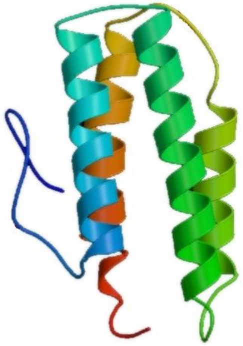


Идентификация личности по фотопортрету

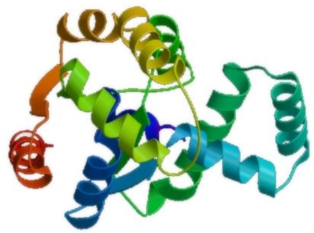
Распознавание классов пространственной структуры белка



Cytochrome C



Immunoglobin beta–sandwich



Four-helical bundle

1TIM:A 247 аминокислот

APRKFFVGGNWKMNGKRKSLGELIHTLDGAKLSADTEVVCGAPSIYLDFA
RQKLDAKIGVAAQNCYKVPKGAFTGEISPAMIKDIGAAWVILGHSERRHV
FGESDELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKVVFQETKAIA
DNVKDWSKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGWLKTHVSDAVA
VQSRIIYGGSVTGGNCKELASQHDVDGFLVGGASLKPEFVDIINAKH

2MHR:_ 118 аминокислот

GWEIPEPYVWDESFRVFYEQLDEEHKKIFKGIFDCIRDNSAPNLATLVKV
TTNHFTHEEAMMDAAKYSEVVPHKKMHKDFLEKIGGLSAPVDAKNVDYCK
EWLVNHIKGTDFKYKGKL

3ADK:_ 195 аминокислот

XMEEKLKKSKIIFVVGGPGSGKGTQCEKIVQKYGYTHLSTGDLLRAEVSS
GSARGKMLSEIMEKGQLVPLETVLDMLRDAMVAKVDTSKGFLIDGYPREV
KQGEEFERKIGQPTLLLYVDAGPETMTKRLLKRGETSGRVDDNEETIKKR
LETYYKATEPVIAFYEKRGIVRKVNAEGSVDDVFSQVCTHLDTLK

Концептуальная база восстановления зависимостей: гипотеза компактности

Множество объектов реального мира $\omega \in \Omega$

Скрытая характеристика объекта (целевая характеристика) $y(\omega): \Omega \to Y$

Искомое решающее правило $\hat{y}(\omega): \Omega \to Y$

Основная идея:

Выбрать в множестве объектов некоторую метрику

$\rho(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$

$\rho(\omega', \omega'') = \rho(\omega'', \omega') \geq 0$, $\rho(\omega', \omega'') > 0$, если $\omega' \neq \omega''$, $\rho(\omega', \omega'') + \rho(\omega'', \omega''') \geq \rho(\omega', \omega''')$

Принимать для близких объектов $\rho(\omega', \omega'') \cong 0$ близкие решения

$\hat{y}(\omega') = \hat{y}(\omega'')$ в задаче распознавания образов $Y = \{y_1, ..., y_m\}$

$\hat{y}(\omega') \cong \hat{y}(\omega'')$ в задаче восстановления числовой зависимости $Y = \mathbb{R}$

Выбор метрик удачен, если для них выполняется ***гипотеза компактности*** (Эммануил Маркович Браверман, 1961):

Для пар объектов $\omega', \omega'' \in \Omega$, похожих в смысле выбранной метрики $\rho(\omega', \omega'') \cong 0$, значения целевой характеристики также в большинстве случаев близки $y(\omega') \cong y(\omega'')$.

Диполь в метрическом пространстве

Метрическое пространство объектов реального мира: $\omega \in \Omega$, $\rho(\omega', \omega'')$ – метрика

Диполь в метрическом пространстве – упорядоченная пара: $< \alpha_{-1}, \alpha_1 > \in \Omega \times \Omega$

Простейшая реализация гипотезы компактности

Принадлежность произвольного объекта $\omega \in \Omega$ к одному из двух классов

$$\hat{y}(\omega) = \begin{cases} 1, \text{ если } \rho(\alpha_1, \omega) < \rho(\alpha_{-1}, \omega), \\ -1, \text{ если } \rho(\alpha_1, \omega) > \rho(\alpha_{-1}, \omega) \end{cases}$$

Как выбрать диполь?

В множестве объектов Ω слишком мало элементов. К тому же, наблюдатель располагает лишь конечной обучающей совокупностью объектов ω_j, $j = 1, \ldots, N$

Более «тонкая» реализация гипотезы компактности для непрерывного метрического пространства

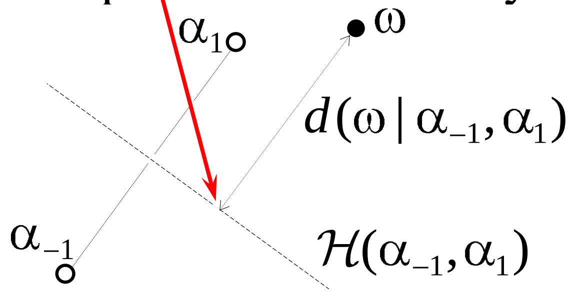
$\tilde{\Omega} \supset \Omega$ – воображаемое непрерывное метрическое пространство, в котором множество реальных объектов является подмножеством, быть может, изолированных элементов.

$\alpha_{-1}, \alpha_1 \in \tilde{\Omega}$, $\mathcal{H}(\alpha_{-1}, \alpha_1) = \left\{ \vartheta \in \tilde{\Omega} : \rho(\alpha_{-1}, \vartheta) = \rho(\alpha_1, \vartheta) \right\}$ – метрическая гиперплоскость в $\tilde{\Omega}$

$\omega_{\mathcal{H}}(\alpha_{-1}, \alpha_1) \in \mathcal{H}(\alpha_{-1}, \alpha_1)$ – проекция реального объекта $\omega \in \Omega$ на гиперплоскость в $\tilde{\Omega}$

Решающая функция (score function): расстояние точки от гиперплоскости в $\tilde{\Omega}$ с учетом знака

$$d(\omega \mid \alpha_{-1}, \alpha_1) = \begin{cases} \rho(\omega_{\mathcal{H}}, \omega), \text{ если } \rho(\alpha_1, \omega) \le \rho(\alpha_{-1}, \omega), \\ -\rho(\omega_{\mathcal{H}}, \omega), \text{ если } \rho(\alpha_1, \omega) > \rho(\alpha_{-1}, \omega). \end{cases}$$



Классификация:

$$\hat{y}(\omega) = \begin{matrix} 1, \text{ если } d(\omega \mid \alpha_{-1}, \alpha_1) + b > 0, \\ -1, \text{ если } d(\omega \mid \alpha_{-1}, \alpha_1) + b < 0. \end{matrix}$$

Числовая зависимость: $\hat{y}(\omega) = d(\omega \mid \alpha_{-1}, \alpha_1) + b$

Идеальные условия для реализации гипотезы компактности: Евклидова метрика в конечномерном линейном пространстве

Вектор действительных признаков $\mathbf{x}(\omega) \in \mathbb{R}^n$ погружает множество реальных объектов в $\mathbb{R}^n$

Естественная евклидова метрика в $\mathbb{R}^n$

$$\rho(\omega', \omega'') = \rho\big(\mathbf{x}(\omega'), \mathbf{x}(\omega'')\big) = \rho(\mathbf{x}', \mathbf{x}'') = \| \mathbf{x}' - \mathbf{x}'' \| = \big[(\mathbf{x}' - \mathbf{x}'')^T (\mathbf{x}' - \mathbf{x}'')\big]^{1/2}$$

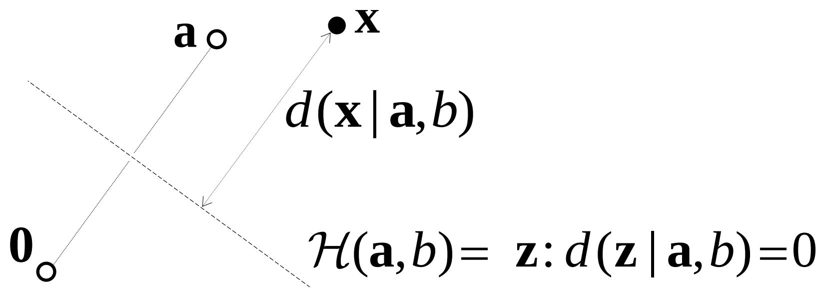
Диполь: $\alpha_1 = \mathbf{a} \in \mathbb{R}^n$, $\alpha_{-1} = -\mathbf{a} \in \mathbb{R}^n$; $\mathbf{a}$ – направляющий вектор гиперплоскости

Смещенная гиперплоскость, определяемая диполем: $\mathcal{H}(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} + b = 0\}$, $\mathbf{a} \in \mathbb{R}^n$

Решающая функция – decision (score) function:

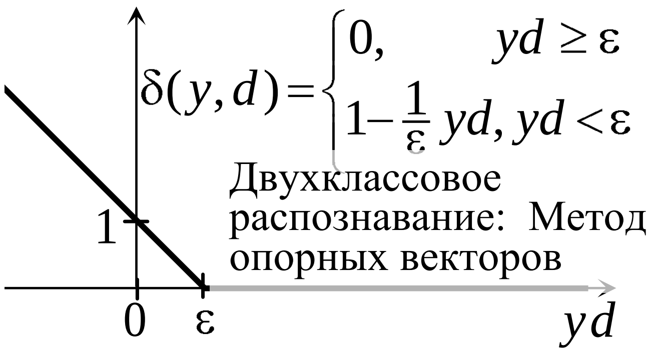
Расстояние от точки $\mathbf{x}(\omega) \in \mathbb{R}^n$ до гиперплоскости

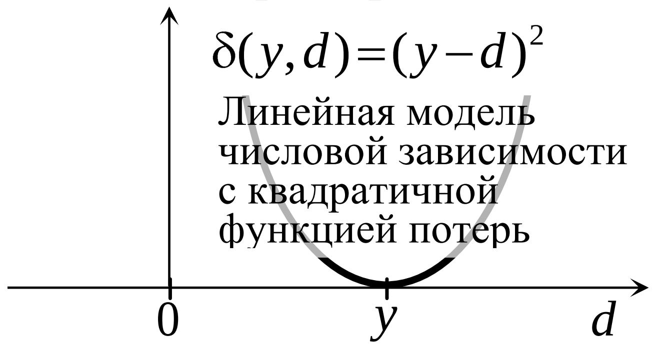
$$d(\mathbf{x} \mid \mathbf{a}, b) = \frac{\mathbf{a}^T \mathbf{x} + b}{(\mathbf{a}^T \mathbf{a})^{1/2}}, \quad d(\mathbf{x} \mid \mathbf{a}, b) = \mathbf{a}^T \mathbf{x} + b \text{ при } \|\mathbf{a}\| = 1$$



Функция потерь: Степень несоответствия значения решающей функции значению целевой характеристики объекта

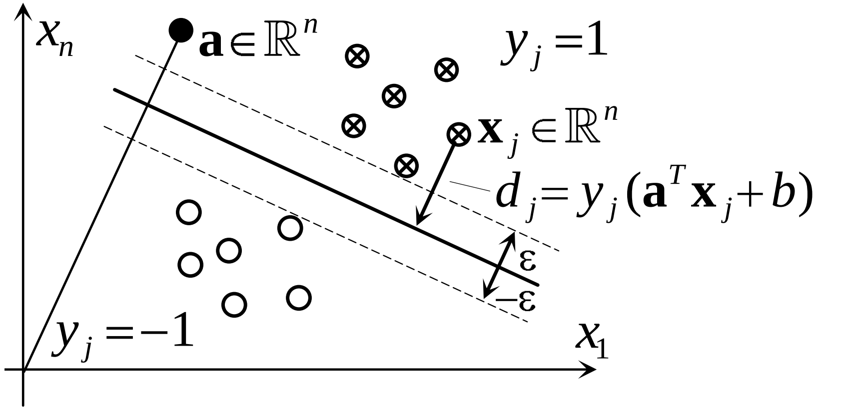
Индекс класса объекта $y \in \{-1, 1\}$ | Числовая характеристика $y \in \mathbb{R}$





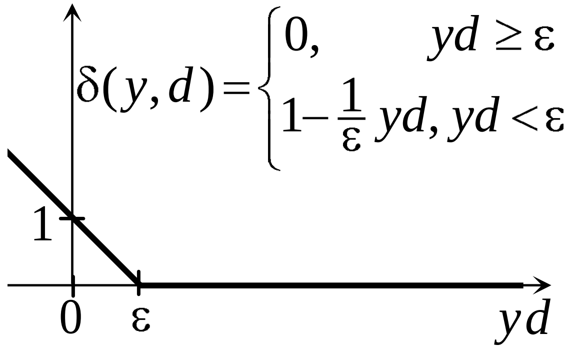
Метод опорных векторов (Support Vector Machine – SVM): Принцип максимального зазора (margin) между классами

Обучающая совокупность $(\mathbf{x}_j, y_j),\ j=1,...,N$, $\mathbf{x}_j \in \mathbb{R}^n,\ y_j \in \{-1, 1\}$

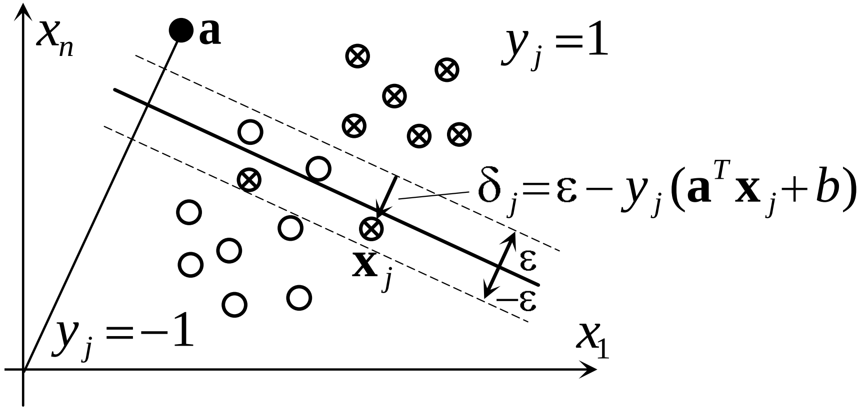


Требование максимизации зазора между линейно разделимыми классами

$$\begin{cases} 1/\varepsilon^2 \to \min(\varepsilon, \mathbf{a}, b) \\ \mathbf{a}^T \mathbf{a} = 1 \end{cases}$$



$$\delta(y,d) = \begin{cases} 0, & yd \ge \varepsilon \\ 1 - \frac{1}{\varepsilon} yd, & yd < \varepsilon \end{cases}$$



Требование минимизации суммы штрафов

$$\begin{cases} \sum_{j=1}^{N} \delta_j \to \min(\varepsilon, \mathbf{a}, b, \delta_1, ..., \delta_N),\ \mathbf{a}^T \mathbf{a} = 1 \\ y_j(\mathbf{a}^T \mathbf{x}_j + b) \ge \varepsilon(1 - \delta_j),\ \delta_j \ge 0,\ j = 1, ..., N \end{cases}$$

Компромисс:

$$\begin{cases} 1/\varepsilon^2 + C\sum_{j=1}^{N} \delta_j \to \min(\varepsilon, \mathbf{a}, b, \delta_1, ..., \delta_N),\ \mathbf{a}^T \mathbf{a} = 1 \\ y_j(\mathbf{a}^T \mathbf{x}_j + b) \ge \varepsilon(1 - \delta_j),\ \delta_j \ge 0,\ j = 1, ..., N \end{cases}$$

Оптимизация на сфере $\mathbf{a}^T \mathbf{a} = 1$:
Невыпуклый критерий обучения

Метод опорных векторов: Выпуклая форма критерия

Невыпуклый критерий

$$\begin{cases} 1/\varepsilon^2 + C\sum_{j=1}^{N}\delta_j \to \min(\varepsilon, \mathbf{a}, b, \delta_1, ..., \delta_N),\ \mathbf{a}^T\mathbf{a} = 1, \\ y_j(\mathbf{a}^T\mathbf{x}_j + b) \ge \varepsilon(1-\delta_j),\ \delta_j \ge 0,\ j = 1, ..., N. \end{cases}$$

Замена переменных

$$\mathbf{a} = \varepsilon\tilde{\mathbf{a}},\ \ b = \varepsilon\tilde{b}, \qquad \text{тогда } \tilde{\mathbf{a}}^T\tilde{\mathbf{a}} = 1/\varepsilon^2$$

Результат замены:

$$\begin{cases} 1/\varepsilon^2 + C\sum_{j=1}^{N}\delta_j \to \min(\varepsilon, \tilde{\mathbf{a}}, \tilde{b}, \delta_1, ..., \delta_N),\ \tilde{\mathbf{a}}^T\tilde{\mathbf{a}} = 1/\varepsilon^2, \\ y_j(\tilde{\mathbf{a}}^T\mathbf{x}_j + \tilde{b}) \ge 1-\delta_j,\ \delta_j \ge 0,\ j = 1, ..., N. \end{cases}$$

Выпуклый критерий обучения
SVM – Support Vector Machine:
$N + n + 1$ переменных, $2N$ ограничений

$$\boxed{\begin{cases} \mathbf{a}^T\mathbf{a} + C\sum_{j=1}^{N}\delta_j \to \min(\mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}, \delta_1, ..., \delta_N \in \mathbb{R}), \\ y_j(\mathbf{a}^T\mathbf{x}_j + b) \ge 1-\delta_j,\ \delta_j \ge 0,\ j = 1, ..., N. \end{cases}}$$

Итак, надо минимизировать выпуклую функцию $J\underbrace{(\mathbf{a}, b, \delta_1, ..., \delta_N)}_{N+n+1\ \text{переменных}} = \mathbf{a}^T\mathbf{a} + C\sum_{j=1}^{N}\delta_j$

при линейных ограничениях типа неравенств

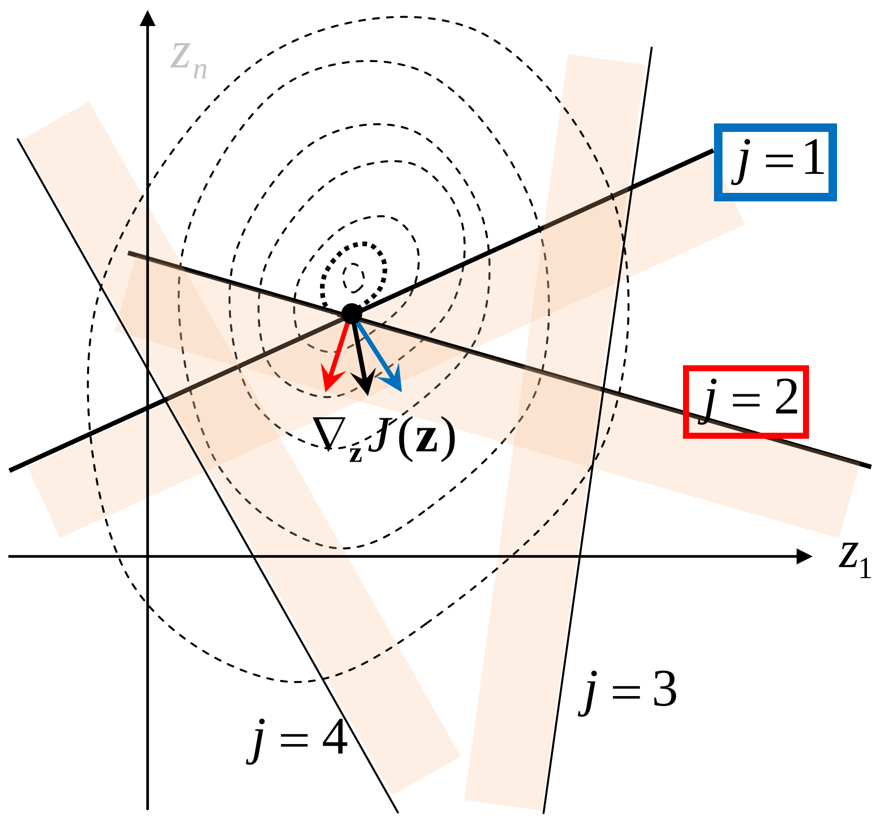
$$\begin{cases} \varphi_j\underbrace{(\mathbf{a}, b, \delta_1, ..., \delta_N)}_{N+n+1\ \text{переменных}} = y_j(\mathbf{a}^T\mathbf{x}_j + b) - 1 + \delta_j \ge 0 \\ \eta_j\underbrace{(\mathbf{a}, b, \delta_1, ..., \delta_N)}_{N+n+1\ \text{переменных}} = \delta_j \ge 0 \end{cases}$$

Это частный вид задачи выпуклого программирования

Задача выпуклого программирования

Минимизировать выпуклую функцию $J(z_1,\dots,z_m)=J(\mathbf{z}),\ \mathbf{z}\in\mathbb{R}^n$

при линейных ограничениях типа неравенств $\varphi_j(z_1,\dots,z_m)=\varphi_j(\mathbf{z})\ge 0,\ \ j=1,\dots,m$



Активные ограничения в точке $\mathbf{z}\in\mathbb{R}^n$:

те ограничения, для которых $\varphi_j(\mathbf{z})=0$

Теорема Куна-Таккера:

Точка $\mathbf{z}\in\mathbb{R}^n$ является точкой минимума тогда и только тогда, когда в этой точке

$$\nabla_{\mathbf{z}}J(\mathbf{z})=\sum_{j\in\left\{\substack{\text{активные}\\ \text{ограничения}}\right\}}\lambda_j\nabla_{\mathbf{z}}\varphi_j(\mathbf{z})\ \text{при}\ \lambda_j\ge 0.$$

Упрощенная формулировка:
Метод множителей Лагранжа:

$$L(\mathbf{z},\lambda_1,\dots,\lambda_m)=J(\mathbf{z})-\sum_{j=1}^{m}\lambda_j\varphi_j(\mathbf{z})\to$$

$$\to\begin{cases}\min(\mathbf{z})\\ \max(\lambda_1\ge 0,\dots,\lambda_m\ge 0)\end{cases}\qquad (\textit{седловая точка})$$

Двойственная задача:

$$\begin{cases}W(\lambda_1,\dots,\lambda_m)=\min\limits_{\mathbf{z}}L(\mathbf{z},\lambda_1,\dots,\lambda_m)\to\max,\\ \lambda_1\ge 0,\dots,\lambda_m\ge 0.\end{cases}$$

Это опять же задача выпуклого программирования, но, как правило, более легкая

Метод опорных векторов: Выпуклый критерий обучения

Задача квадратичного программирования

$N+n+1$ переменных, $2N$ ограничений

$$\begin{cases} \underbrace{\mathbf{a}^T\mathbf{a}+C\sum_{j=1}^{N}\delta_j}_{\text{квадратичная функция}} \to \min(\mathbf{a}\in\mathbb{R}^n, b\in\mathbb{R}, \delta_1,\ldots,\delta_N\in\mathbb{R}), \\ \underbrace{y_j(\mathbf{a}^T\mathbf{x}_j+b)\geq 1-\delta_j}_{\substack{\text{множители Лагранжа}\\ \lambda_1,\ldots,\lambda_N}}, \quad \underbrace{\delta_j\geq 0}_{\substack{\text{множители Лагранжа}\\ \mu_1,\ldots,\mu_N}}, \quad j=1,\ldots,N. \end{cases}$$

две группы линейных ограничений

Функция Лагранжа: $L(\mathbf{a},b,\delta_1,\ldots,\delta_N;\lambda_1,\ldots,\lambda_N,\mu_1,\ldots,\mu_N)=$

$$\frac{1}{2}\left(\mathbf{a}^T\mathbf{a}+C\sum_{j=1}^{N}\delta_j\right)-\sum_{j=1}^{N}\lambda_j\ y_j(\mathbf{a}^T\mathbf{x}_j+b)-1+\delta_j\ -\sum_{j=1}^{N}\mu_j\delta_j\to \begin{array}{l}\min(\mathbf{a},b,\delta_1,\ldots,\delta_N),\\ \max(\lambda_1\geq 0,\ldots,\lambda_N\geq 0,\mu_1\geq 0,\ldots,\mu_N\geq 0)\end{array}$$

Минимум по $\mathbf{a}$: $\nabla_{\mathbf{a}}L(\mathbf{a},b,\delta_1,\ldots,\delta_N;\lambda_1,\ldots,\lambda_N,\mu_1,\ldots,\mu_N)=\mathbf{0}\Rightarrow \boxed{\mathbf{a}=\sum_{j=1}^{N}y_j\lambda_j\mathbf{x}_j}$

Минимум по b: $\partial/\partial b\ \ L(\mathbf{a},b,\delta_1,\ldots,\delta_N;\lambda_1,\ldots,\lambda_N,\mu_1,\ldots,\mu_N)=0\Rightarrow \boxed{\sum_{j=1}^{N}y_j\lambda_j=0}$

Минимум по $\delta_1,\ldots,\delta_N$: $\partial/\partial\delta_j\ \ L(\mathbf{a},b,\delta_1,\ldots,\delta_N;\lambda_1,\ldots,\lambda_N,\mu_1,\ldots,\mu_N)=0\Rightarrow \boxed{\begin{array}{l}C/2-\lambda_j-\mu_j=0,\\ j=1,\ldots,N\end{array}}$

Подстановка этих равенств функцию Лагранжа приводит к двойственной задаче:

$\begin{cases}\sum_{j=1}^{N}\lambda_j-\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l(\mathbf{x}_j^T\mathbf{x}_l)\lambda_j\lambda_l\to\max,\\ \sum_{j=1}^{N}y_j\lambda_j=0,\ \ 0\leq\lambda_j\leq C/2,\ \ j=1,\ldots,N.\end{cases}$	Задача квадратичного программирования SVM. Решения λ_j соответствуют обучающим объектам и распадаются на три группы 1) $\lambda_j=0$; 2) $0<\lambda_j<C/2$; 3) $\lambda_j=C/2$.

Решение задачи обучения SVM

Еще раз двойственная задача квадратичного программирования
Решения λ_j соответствуют обучающим объектам и распадаются на три группы

1) $\lambda_j = 0$; $\underbrace{\text{2) } 0 < \lambda_j < C/2; \quad \text{3) } \lambda_j = C/2}_{\textit{опорные объекты}}$.

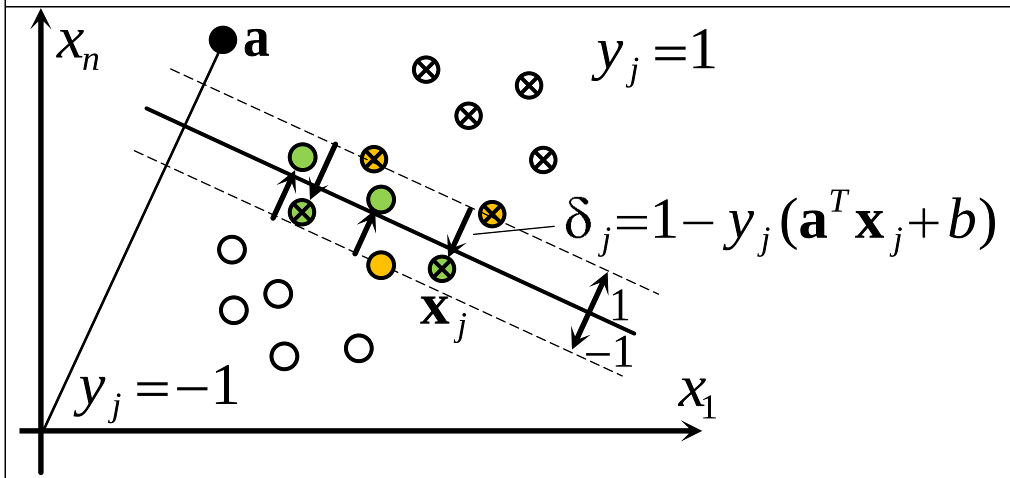
$$\begin{cases} \sum_{j=1}^{N} \lambda_j + \frac{1}{2} \sum_{j=1}^{N} \sum_{l=1}^{N} y_j y_l (\mathbf{x}_j^T \mathbf{x}_l) \lambda_j \lambda_l \to \max, \\ \sum_{j=1}^{N} y_j \lambda_j = 0, \ \ 0 \le \lambda_j \le C/2, \ \ j = 1, \dots, N. \end{cases}$$

Направляющий вектор разделяющей гиперплоскости:

$$\mathbf{a} = \sum_{j=1}^{N} y_j \lambda_j \mathbf{x}_j = \underbrace{\sum_{j:\, \lambda_j > 0} y_j \lambda_j \mathbf{x}_j}_{\textit{опорные объекты}}$$

Качество разделения классов в обучающей совокупности:

$$\delta_j = \begin{cases} 0, \ \lambda_j < C/2 \\ \delta_j > 0, \ \lambda_j = C/2 \end{cases}$$



⊗ ○ неопорные объекты

⊗ ○ опорные несдвинутые объекты

⊗ ○ опорные сдвинутые объекты

Результат обучения

Исходное правило классификации нового объекта $d(\mathbf{x}) = \dfrac{\mathbf{a}^T\mathbf{x}+b}{\mathbf{a}^T\mathbf{a}} \gtrless 0$

$$\mathbf{a} = \sum_{j:\lambda_j>0} y_j\lambda_j\mathbf{x}_j\,, \qquad b = -\frac{\sum_{j:0<\lambda_j<C/2}\lambda_j \sum_{l:\lambda_l>0} y_l\lambda_l(\mathbf{x}_j^T\mathbf{x}_l) + (C/2)\sum_{j:\lambda_j=C/2} y_j}{\sum_{j:0<\lambda_j<C/2}\lambda_j}$$

Эквивалентное правило классификации нового объекта

$$d(\mathbf{x}) = \frac{\sum_{j:\lambda_j>0} y_j\lambda_j(\mathbf{x}_j^T\mathbf{x}) + b}{\sum_{k:\lambda_k>0}\sum_{l:\lambda_l>0} y_k y_l(\mathbf{x}_k^T\mathbf{x}_l)\lambda_k\lambda_l} = \sum_{j:\lambda_j>0} c_j(\mathbf{x}_j^T\mathbf{x}) + g \gtrless 0$$

$$d(\mathbf{x}) = \frac{\sum_{j:\lambda_j>0} {\color{red}y_j\lambda_j}(\mathbf{x}_j^T\mathbf{x}) + b}{\color{red}\sum_{k:\lambda_k>0}\sum_{l:\lambda_l>0} y_k y_l(\mathbf{x}_k^T\mathbf{x}_l)\lambda_k\lambda_l} = \sum_{j:\lambda_j>0} {\color{red}c_j}(\mathbf{x}_j^T\mathbf{x}) + g \gtrless 0$$

$$d(\mathbf{x}) = \frac{\sum_{j:\lambda_j>0} y_j\lambda_j(\mathbf{x}_j^T\mathbf{x}) + {\color{green}b}}{\color{green}\sum_{k:\lambda_k>0}\sum_{l:\lambda_l>0} y_k y_l(\mathbf{x}_k^T\mathbf{x}_l)\lambda_k\lambda_l} = \sum_{j:\lambda_j>0} c_j(\mathbf{x}_j^T\mathbf{x}) + {\color{green}g} \gtrless 0$$

Перейдем теперь к вероятностной интерпретации метода опорных векторов.

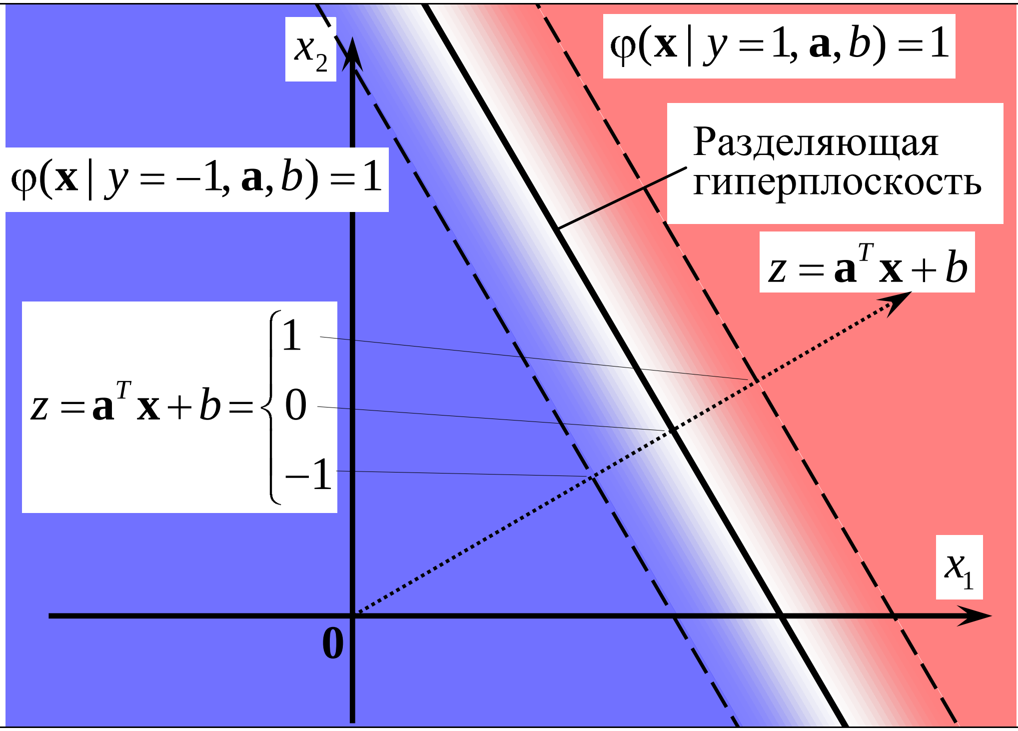
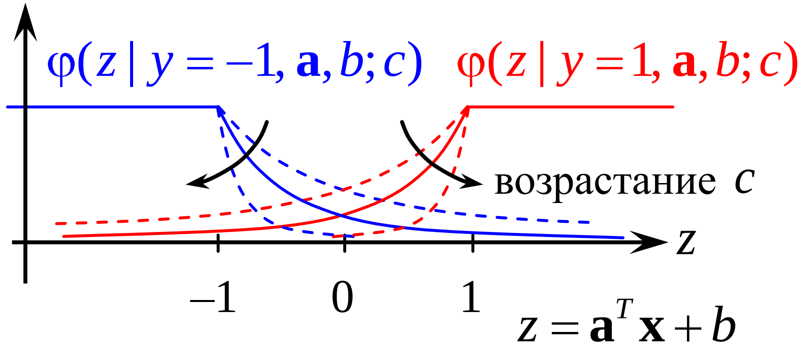
Начнем с простейшей вероятностной постановка задачи обучения распознаванию образов вообще, безотносительно к методу опорных векторов.

Простейшая вероятностная байесовская постановка задачи обучения распознаванию образов

Вектор признаков объекта и индекс его класса	$\mathbf{x} = (x_1 \cdots x_n)^T \in \mathbb{R}^n$ $y \in \{1,\ldots,m\}$
Параметрическое семейство плотностей распределения вектора признаков случайного объекта, зависящее от класса	$\varphi(\mathbf{x} \mid y, \mathbf{a}),\ \mathbf{a} \in \mathbb{R}^m,\ \int_{\mathbb{R}^n} \varphi(\mathbf{x} \mid y; \mathbf{a}) d\mathbf{x} = 1$
Обучающая совокупность	$(X, Y) = \ (\mathbf{x}_j, y_j),\ j = 1,\ldots,N$
Совместная условная плотность распределения обучающей совокупности с известными классами объектов	$\Phi(X \mid Y, \mathbf{a}) = \prod_{j=1}^{N} \varphi(\mathbf{x}_j \mid y_j, \mathbf{a})$
Априорная плотность распределения неизвестного случайного параметра	$\Psi(\mathbf{a}),\ \int_{\mathbb{R}^k} \Psi(\mathbf{a}) d\mathbf{a} = 1,\ \ln \Psi(\mathbf{a})$ строго выпуклая функция
Апостериорная плотность распределения параметра	$P(\mathbf{a} \mid X, Y) = \dfrac{\Psi(\mathbf{a})\Phi(X \mid Y, \mathbf{a})}{F(X \mid Y)} \propto \Psi(\mathbf{a})\Phi(X \mid Y, \mathbf{a})$
Байесовский критерий обучения	$\hat{\mathbf{a}}(X, Y) = \arg\max_{\mathbf{a}} \Psi(\mathbf{a})\Phi(X \mid Y, \mathbf{a})$
Априорные вероятности классов	$q(y),\ \sum_y q(y) = 1$
Апостериорные вероятности каждого класса для нового объекта	$p(y \mid \mathbf{x}, \hat{\mathbf{a}}) \propto q(y)\varphi(\mathbf{x} \mid y, \hat{\mathbf{a}})$
Решение о его классе	$\hat{y}(\mathbf{x}, \hat{\mathbf{a}}) = \arg\max_{y \in \{1,\ldots,m\}} q(y)\varphi(\mathbf{x} \mid y, \hat{\mathbf{a}})$

Параметрическое семейство пары несобственных плотностей распределения в пространстве признаков объектов двух классов

Первая идея: Единственное предположение о двух классах объектов заключается в том, что они преимущественно отображаются по разные стороны некоторой гиперплоскости в пространстве признаков.

$\varphi(\mathbf{x}\mid y,\mathbf{a},b;c)\propto\begin{cases}1, & y(\mathbf{a}^T\mathbf{x}+b)\geq 1\\ \exp\left[-c\ \ 1-y(\mathbf{a}^T\mathbf{x}+b)\ \ \right], & y(\mathbf{a}^T\mathbf{x}+b)<1\end{cases}$	$\mathbf{x}\in\mathbb{R}^n,\ y\in\{-1,1\}$ Параметры гиперплоскости: $\mathbf{a}\in\mathbb{R}^n,\ b\in\mathbb{R}$ Structural parameter $c>0$ is responsible for intersection rate of the two distributions
	

Критерий обучения

Апостериорная совместная плотность распределения параметров гиперплоскости	$P(\mathbf{a},b\mid X,Y;c)\propto\Psi(\mathbf{a})\prod_{j=1}^{N}\varphi(\mathbf{x}_j\mid y_j,\mathbf{a},b;c)$
Критерий обучения: Оптимальная (апостериори наиболее вероятная) разделяющая гиперплоскость	$(\hat{\mathbf{a}},\hat{b}\mid X,Y;c)=\arg\max_{\mathbf{a},b}\left[\ln\Psi(\mathbf{a})+\sum_{j=1}^{N}\ln\varphi(\mathbf{x}_j\mid y_j,\mathbf{a},b;c)\right]=$ $\arg\min_{\mathbf{a},b}\left\{-\ln\Psi(\mathbf{a})+\underbrace{c\left[\sum_{\substack{j:\,y_j=1\\ \mathbf{a}^T\mathbf{x}_j+b<1}}1-(\mathbf{a}^T\mathbf{x}_j+b)+\sum_{\substack{j:\,y_j=-1\\ \mathbf{a}^T\mathbf{x}_j+b>-1}}1+(\mathbf{a}^T\mathbf{x}_j+b)\right]}_{\textit{неверно классифицированные объекты обучающей совокупности}}\right\}$
Теорема: Эквивалентный критерий обучения	$\begin{cases}-\ln\Psi(\mathbf{a})+c\sum_{j=1}^{N}\delta_j\to\min(\mathbf{a},b,\delta_1,,\dots,\delta_N),\\ y_j(\mathbf{a}^T\mathbf{x}_j+b)\geq 1-\delta_j,\ \delta_j\geq 0,\ j=1,\dots,N.\end{cases}$
Для сравнения: Критерий SVM, построенный из геометрических соображений	$\begin{cases}\mathbf{a}^T\mathbf{a}+C\sum_{j=1}^{N}\delta_j\to\min(\mathbf{a},b,\delta_1,,\dots,\delta_N),\\ y_j(\mathbf{a}^T\mathbf{x}_j+b)\geq 1-\delta_j,\ \delta_j\geq 0,\ j=1,\dots,N.\end{cases}$

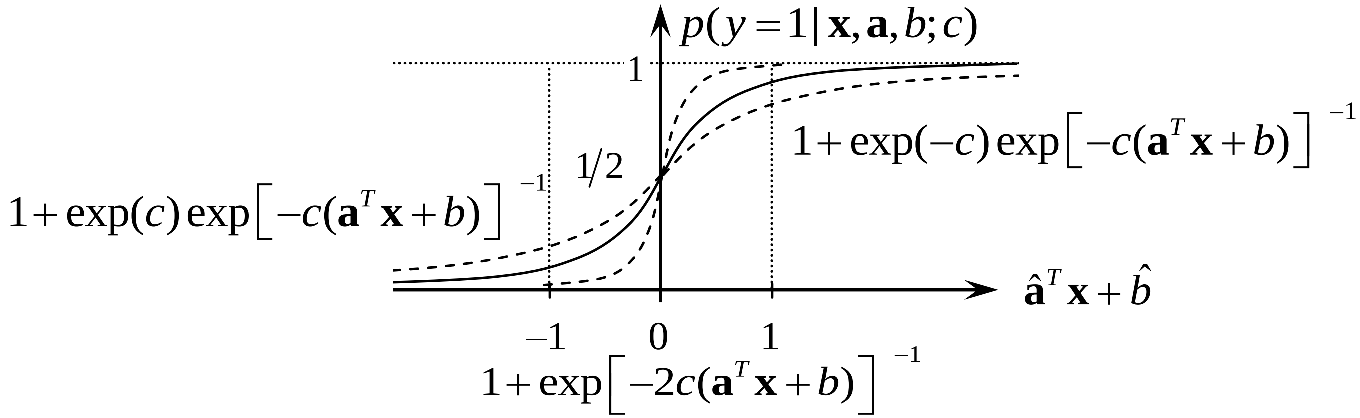
Критерий обучения

Апостериорная совместная плотность распределения параметров гиперплоскости	$P(\mathbf{a},b \mid X,Y;c) \propto \Psi(\mathbf{a}) \prod_{j=1}^{N} \varphi(\mathbf{x}_j \mid y_j, \mathbf{a}, b; c)$
Критерий обучения: Оптимальная (апостериори наиболее вероятная) разделяющая гиперплоскость	$(\hat{\mathbf{a}},\hat{b} \mid X,Y;c) = \arg\max_{\mathbf{a},b} \left[\ln \Psi(\mathbf{a}) + \sum_{j=1}^{N} \ln \varphi(\mathbf{x}_j \mid y_j, \mathbf{a}, b; c) \right] =$ $\arg\min_{\mathbf{a},b} \left\{ -\ln \Psi(\mathbf{a}) + \underbrace{c \left[\sum_{\substack{j:\, y_j = 1 \\ \mathbf{a}^T\mathbf{x}_j + b < 1}} 1 - (\mathbf{a}^T\mathbf{x}_j + b) + \sum_{\substack{j:\, y_j = -1 \\ \mathbf{a}^T\mathbf{x}_j + b > -1}} 1 + (\mathbf{a}^T\mathbf{x}_j + b) \right]}_{\text{неверно классифицированные объекты обучающей совокупности}} \right\}$
Теорема: Эквивалентный критерий обучения	$\begin{cases} -\ln \Psi(\mathbf{a}) + c \sum_{j=1}^{N} \delta_j \to \min(\mathbf{a}, b, \delta_1, ..., \delta_N), \\ y_j(\mathbf{a}^T\mathbf{x}_j + b) \geq 1 - \delta_j,\ \delta_j \geq 0,\ j = 1,...,N. \end{cases}$
Апостериорные вероятности для априори равновероятных классов	$p(y=1 \mid \mathbf{x}, \mathbf{a}, b; c) = \begin{cases} \left(1 + \exp(c) \exp\left[-c(\mathbf{a}^T\mathbf{x} + b)\right]\right)^{-1}, \mathbf{a}^T\mathbf{x} + b < -1, \\ \left(1 + \exp\left[-2c(\mathbf{a}^T\mathbf{x} + b)\right]\right)^{-1}, -1 \leq \mathbf{a}^T\mathbf{x} + b \leq 1, \\ \left(1 + \exp(-c) \exp\left[-c(\mathbf{a}^T\mathbf{x} + b)\right]\right)^{-1}, \mathbf{a}^T\mathbf{x} + b > 1. \end{cases}$

Апостериорные вероятности классов для нового объекта: Графическое представление

$$p(y=1\,|\,\mathbf{x},\mathbf{a},b;c) = 1 - p(y=-1\,|\,\mathbf{x},\mathbf{a},b;c) = \begin{cases} \left(1+\exp(c)\exp\left[-c(\mathbf{a}^T\mathbf{x}+b)\right]\right)^{-1}, \mathbf{a}^T\mathbf{x}+b<-1, \\ \left(1+\exp\left[-2c(\mathbf{a}^T\mathbf{x}+b)\right]\right)^{-1}, -1\le\mathbf{a}^T\mathbf{x}+b\le 1, \\ \left(1+\exp(-c)\exp\left[-c(\mathbf{a}^T\mathbf{x}+b)\right]\right)^{-1}, \mathbf{a}^T\mathbf{x}+b>1. \end{cases}$$

Это логистическая функция:



Одинаковые независимые нормальные распределения компонент направляющего вектора

Совместная априорная плотность распределения направляющего вектора	$\Psi(\mathbf{a}) = \Psi(a_1,...,a_n) = \prod_{i=1}^{n} \mathcal{N}(a_i \mid 0, r) = \frac{1}{r^{n/2}(2\pi)^{n/2}} \exp\left(-\frac{1}{2r}\mathbf{a}^T\mathbf{a}\right)$
Логарифмическая априорная плотность	$-\ln\Psi(\mathbf{a}) = \frac{1}{2} n \ln(2\pi) + \frac{1}{2} n \ln r + \frac{1}{2r}\mathbf{a}^T\mathbf{a}$
Критерий обучения: Обычный критерий SVM	$\begin{cases} \mathbf{a}^T\mathbf{a} + C\sum_{j=1}^{N}\delta_j \to \min(\mathbf{a}, b), \quad C = 2rc, \\ y_j(\mathbf{a}^T\mathbf{x}_j + b) \ge 1 - \delta_j,\ \delta_j \ge 0,\ j = 1,...,N. \end{cases}$
Двойственная задача	$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}(y_j y_l \mathbf{x}_j^T\mathbf{x}_l)\lambda_j\lambda_l \to \max(\lambda_1,...,\lambda_N), \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0 \le \lambda_j \le C/2,\ \ j = 1,...,N. \end{cases}$ опорные векторы $\mathbf{x}_j : \lambda_j > 0$, если $\lambda_j < C/2$, то $\delta_j = 0$.
Параметры оптимальной разделяющей гиперплоскости	$\mathbf{a} = \sum_{j:\lambda_j>0} y_j\lambda_j\mathbf{x}_j,\ \ b = -\frac{\sum_{j:\lambda_j>0}\sum_{l:0<\lambda_l<C/2} y_j y_l \mathbf{x}_j^T\mathbf{x}_l\lambda_l\lambda_l + (C/2)\sum_{j:\lambda_j=C/2} y_j}{\sum_{j:0<\lambda_j<C/2} y_j\lambda_j}$
Решение о классе нового объекта	$\sum_{j:\lambda_j>0} y_j\lambda_j\mathbf{x}_j^T\mathbf{x} + b \gtrless 0$

Метод опорных векторов: Выпуклая форма критерия

Выпуклый критерий обучения:
Задача квадратичного программирования
$N+n+1$ переменных, $2N$ ограничений

$$\begin{cases} \mathbf{a}^T\mathbf{a} + C\sum_{j=1}^{N}\delta_j \to \min(\mathbf{a}\in\mathbb{R}^n, b\in\mathbb{R}, \delta_1,\dots,\delta_N\in\mathbb{R}), \\ y_j(\mathbf{a}^T\mathbf{x}_j + b)\ge 1-\delta_j,\ \delta_j\ge 0,\ j=1,\dots,N. \end{cases}$$

Двойственная форма задачи:
Задача квадратичного программирования
Переменные $\{\lambda_1,\dots,\lambda_N\}$ соответствуют обучающим объектам $\{\mathbf{x}_1,\dots,\mathbf{x}_N\}$

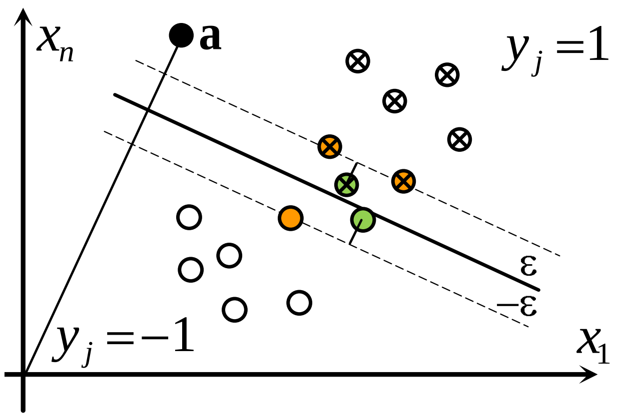
$$\begin{cases} \sum_{j=1}^{N}\lambda_j + \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l(\mathbf{x}_j^T\mathbf{x}_l)\lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\le\lambda_j\le C/2,\ j=1,\dots,N. \end{cases}$$

$$\mathbf{a} = \sum_{j:\lambda_j>0} y_j\lambda_j\mathbf{x}_j,\quad b = -\frac{\sum_{j:0<\lambda_j<C/2}\lambda_j\sum_{l:\lambda_l>0} y_l\lambda_l(\mathbf{x}_j^T\mathbf{x}_l) + (C/2)\sum_{j:\lambda_j=C/2} y_j}{\sum_{j:0<\lambda_j<C/2}\lambda_j},\quad \delta_j = \begin{cases} 0,\ \lambda_j<C/2 \\ \delta_j>0,\ \lambda_j=C/2 \end{cases}$$

опорные векторы

Правило классификации нового объекта $d(\mathbf{x}) = \frac{\mathbf{a}^T\mathbf{x}+b}{\mathbf{a}^T\mathbf{a}} \gtrless 0$

Эквивалентная запись $d(\mathbf{x}) = \frac{\sum_{j:\lambda_j>0} y_j\lambda_j(\mathbf{x}_j^T\mathbf{x}) + b}{\sum_{k:\lambda_k>0}\sum_{l:\lambda_l>0} y_k y_l(\mathbf{x}_k^T\mathbf{x}_l)\lambda_k\lambda_l} = \sum_{j:\lambda_j>0} c_j(\mathbf{x}_j^T\mathbf{x}) + g \gtrless 0$



Метод опорных векторов в пространстве векторов действительных признаков объектов Support Vector Machine (SVM)

Но ведь мы начинали наши рассуждения с некоторой метрики на множестве объектов, т.е. без предположения, что на каждом объекте измерен вектор действительных признаков

Повтор: Диполь в метрическом пространстве

Метрическое пространство объектов реального мира: $\omega \in \Omega$, $\rho(\omega', \omega'')$ – метрика

Диполь в метрическом пространстве – упорядоченная пара: $< \alpha_{-1}, \alpha_1 > \in \Omega \times \Omega$

Простейшая реализация гипотезы компактности

Принадлежность произвольного объекта $\omega \in \Omega$ к одному из двух классов

$$\hat{y}(\omega) = \begin{cases} 1, \text{ если } \rho(\alpha_1, \omega) < \rho(\alpha_{-1}, \omega), \\ -1, \text{ если } \rho(\alpha_1, \omega) > \rho(\alpha_{-1}, \omega). \end{cases}$$

Как выбрать диполь?

В множестве объектов Ω лишком мало элементов. К тому же, наблюдатель располагает лишь конечной обучающей совокупностью объектов ω_j, $j = 1, ..., N$

Более «тонкая» реализация гипотезы компактности для непрерывного метрического пространства

$\tilde{\Omega} \supset \Omega$ – воображаемое непрерывное метрическое пространство, в котором множество реальных объектов является подмножеством, быть может, изолированных элементов.

$\alpha_{-1}, \alpha_1 \in \tilde{\Omega}$, $\mathcal{H}(\alpha_{-1}, \alpha_1) = \{ \vartheta \in \tilde{\Omega} : \rho(\alpha_{-1}, \vartheta) = \rho(\alpha_1, \vartheta) \}$ – метрическая гиперплоскость в $\tilde{\Omega}$

$\omega_{\mathcal{H}}(\alpha_{-1}, \alpha_1) \in \mathcal{H}(\alpha_{-1}, \alpha_1)$ – проекция реального объекта $\omega \in \Omega$ на гиперплоскость в $\tilde{\Omega}$

Решающая функция (score function): расстояние точки от гиперплоскости в $\tilde{\Omega}$ с учетом знака

$$d(\omega \,|\, \alpha_{-1}, \alpha_1) = \begin{cases} \rho(\omega_{\mathcal{H}}, \omega), \text{ если } \rho(\alpha_1, \omega) \le \rho(\alpha_{-1}, \omega), \\ -\rho(\omega_{\mathcal{H}}, \omega), \text{ если } \rho(\alpha_1, \omega) > \rho(\alpha_{-1}, \omega). \end{cases}$$

Средство построения такого непрерывного метрического пространства: Линейное пространство векторов признаков. Как обойтись без необходимости измерять признаки?

Соосность элементов метрического пространства

Метрическое пространство Ω с метрикой $\rho(\omega',\omega'')$:

$\rho(\omega,\omega)=0$, $\rho(\omega',\omega'')>0$, если $\omega'\neq\omega''$;

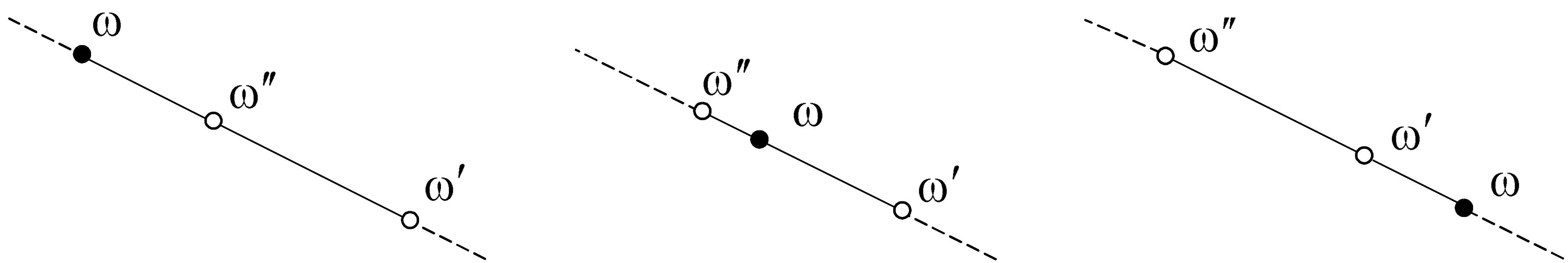
$\rho(\omega',\omega'')=\rho(\omega'',\omega')$;

$\rho(\omega',\omega'')+\rho(\omega'',\omega''')\geq\rho(\omega',\omega''')$ – неравенство треугольника (равенство для некоторых троек)

Упорядоченная пара элементов $<\omega',\omega''>\in\Omega\times\Omega$, действительное число $c\in\mathbb{R}$.

Элемент, соосный паре $<\omega',\omega''>$ с коэффициентом c

$$\omega=\text{Coax}\ <\omega',\omega''>;c\ :\qquad \rho(\omega',\omega)=|c|\,\rho(\omega',\omega''),\ \ \rho(\omega'',\omega)=|c-1|\,\rho(\omega',\omega'')$$



$c>1$ $\qquad$ $0\leq c\leq 1$ $\qquad$ $c<0$

Метрическое пространство Ω называется ординарным, если для каждой пары $<\omega',\omega''>\in\Omega\times\Omega$ и коэффициента $c\in\mathbb{R}$ существует не более одного элемента $\omega=\text{Coax}\ <\omega',\omega''>;c$.

Неограниченно выпуклое метрическое пространство

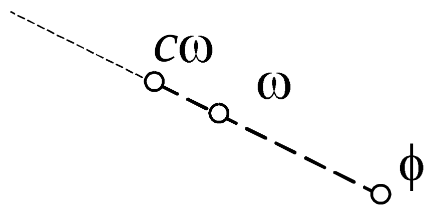
Метрическое пространство Ω называется неограниченно выпуклым, если для любой пары $<\omega',\omega''>\in\Omega\times\Omega$ и любого коэффициента $c\in\mathbb{R}$ в нем существует элемент $\omega=\mathrm{Coax}\ <\omega',\omega''>;c$.

Идея введения линейных операций в ординарном неограниченно выпуклом метрическом пространстве

Центр метрического пространства $\phi\in\Omega$.

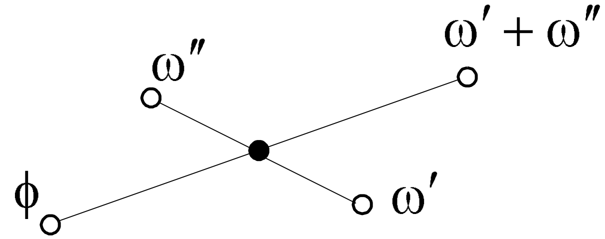
Умножение элемента $\omega\in\Omega$ на коэффициент

$$c\omega=\mathrm{Coax}\ <\phi,\omega>;c$$



Сложение элементов $\omega',\omega''\in\Omega$

$$\omega'+\omega''=2\mathrm{Coax}(<\omega',\omega''>;1/2)$$



Пример «наивной» реализации линейных операций над изображениями

Пара изображений лица человека

изображение $\mathbf{x}'$

изображение $\mathbf{x}'$

Пример «наивной» реализации линейных операций над изображениями

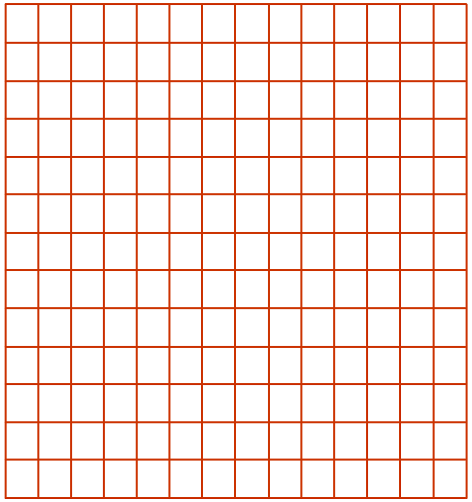
Пара изображений лица человека

изображение $\mathbf{x}'$

$\mathbf{x}' = \; x'_{\mathbf{s}'}, \mathbf{s}' \in \mathbf{T}'$

изображение $\mathbf{x}'$

$\mathbf{x}'' = \; x''_{\mathbf{s}''}, \mathbf{s}'' \in \mathbf{T}''$

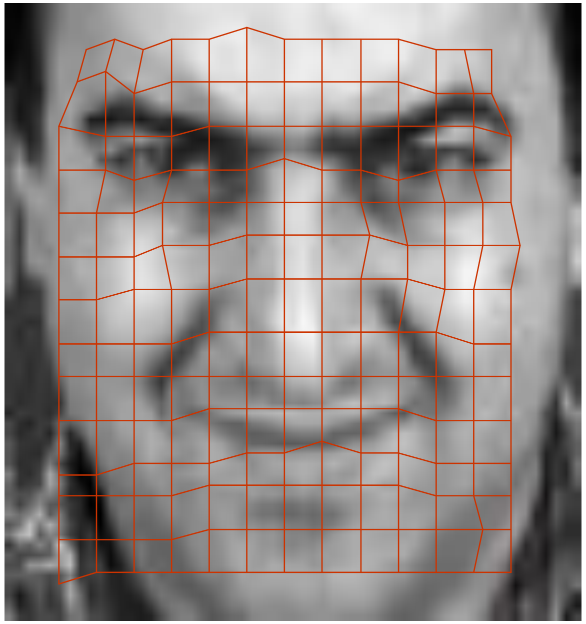


Пример «наивной» реализации линейных операций над изображениями

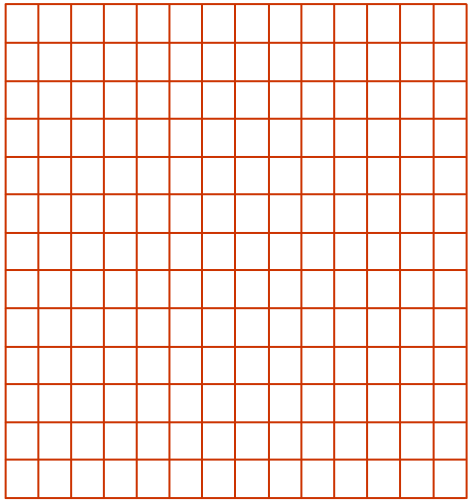
Пара изображений лица человека

изображение $\mathbf{x}'$

$\mathbf{x}' = \quad x'_{\mathbf{s}'}, \mathbf{s}' \in \mathbf{T}'$

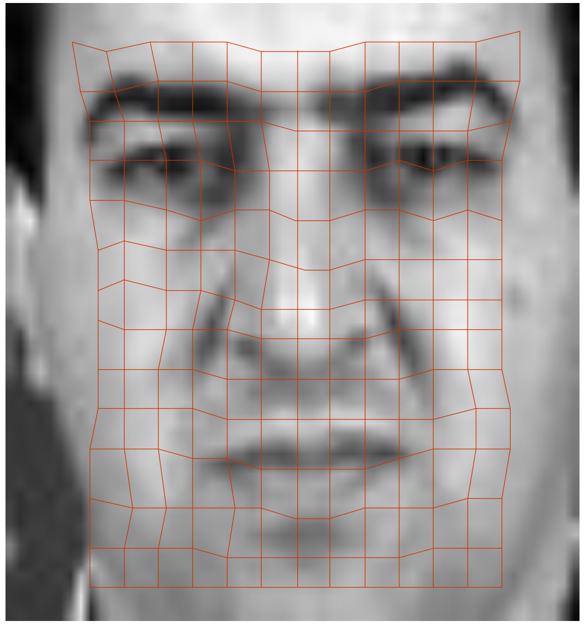


«среднее» изображение



изображение $\mathbf{x}'$

$\mathbf{x}'' = \quad x''_{\mathbf{s}''}, \mathbf{s}'' \in \mathbf{T}''$

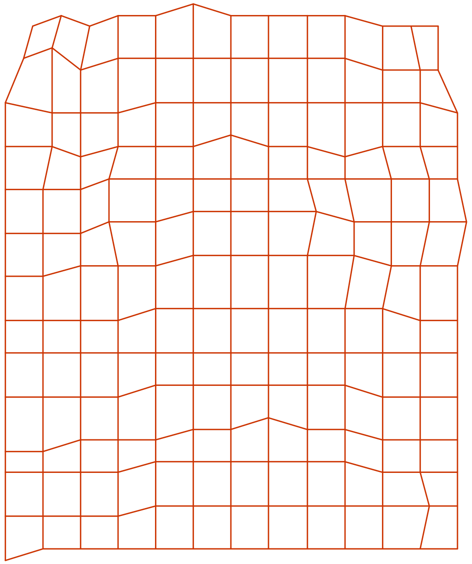


Пример «наивной» реализации линейных операций над изображениями

Пара изображений лица человека

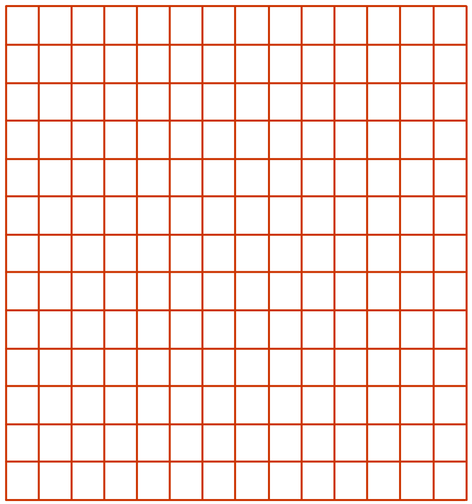
изображение $\mathbf{x}'$

$\mathbf{x}' = x'_{\mathbf{s}'},\ \mathbf{s}' \in \mathbf{T}'$



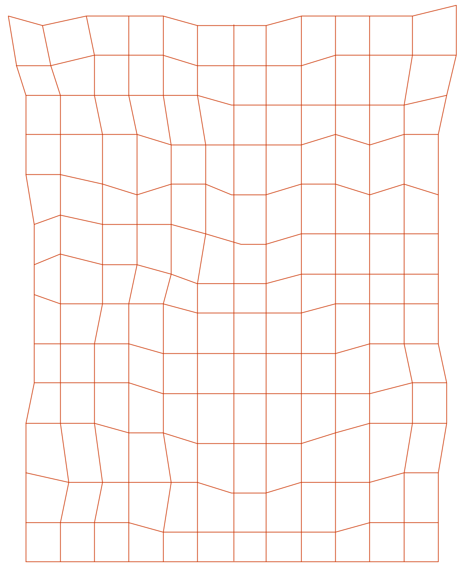
«среднее» изображение

$\mathbf{x} = x_{\mathbf{t}},\ \mathbf{t} \in \mathbf{T}$



изображение $\mathbf{x}'$

$\mathbf{x}'' = x''_{\mathbf{s}''},\ \mathbf{s}'' \in \mathbf{T}''$



$$\mathbf{s}'(\mathbf{t}) = \mathbf{t} + \mathbf{v}_{\mathbf{t}}/2 \in \mathbf{T}' \quad \leftarrow \quad \mathbf{t} \in \mathbf{T} \quad \rightarrow \quad \mathbf{s}''(\mathbf{t}) = \mathbf{t} - \mathbf{v}_{\mathbf{t}}/2 \in \mathbf{T}''$$

$$J(\mathbf{V} \mid \mathbf{x}', \mathbf{x}'') = \sum_{\mathbf{t}\in\mathbf{T}} \left(x'_{\mathbf{t}+\mathbf{v}_{\mathbf{t}}/2} - x''_{\mathbf{t}-\mathbf{v}_{\mathbf{t}}/2}\right)^2 + \beta \sum_{(\mathbf{t},\mathbf{u})\in G} \| \mathbf{v}_{\mathbf{t}} - \mathbf{v}_{\mathbf{u}} \|^2 \to \min$$

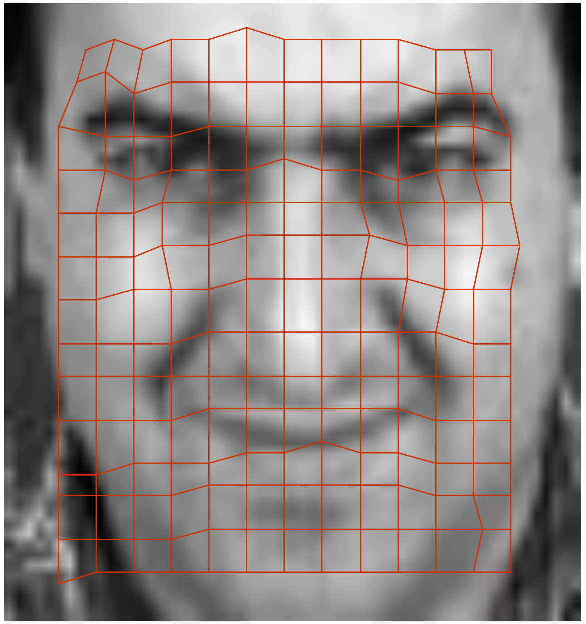
Метрика: $\rho(\mathbf{x}', \mathbf{x}'') = \left(\sum_{\mathbf{t}\in\mathbf{T}} \left(x'_{\mathbf{t}+\hat{\mathbf{v}}_{\mathbf{t}}/2} - x''_{\mathbf{t}-\hat{\mathbf{v}}_{\mathbf{t}}/2}\right)^2 \right)^{1/2}$

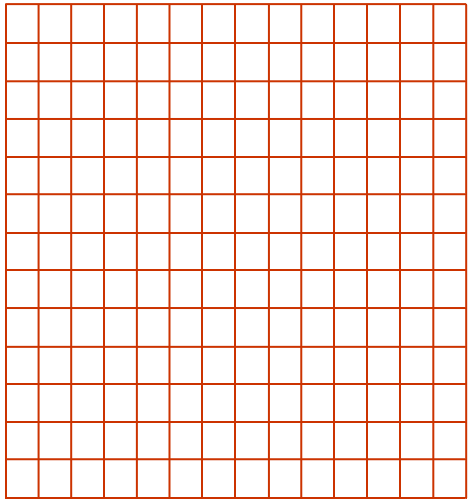
Пример «наивной» реализации линейных операций над изображениями

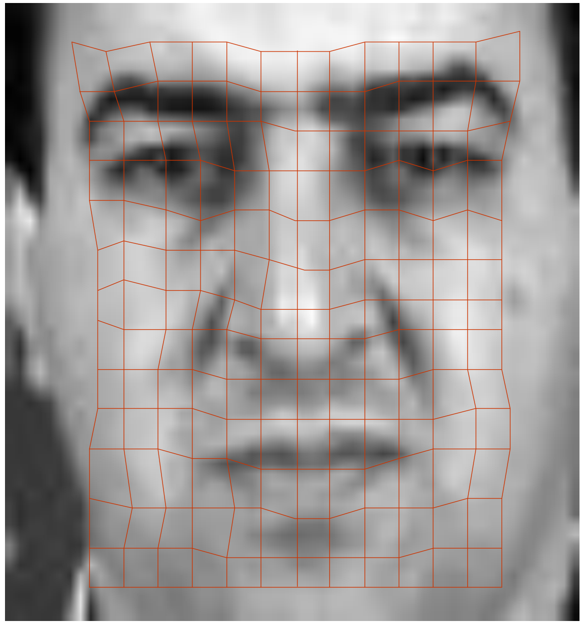
Пара изображений лица человека

изображение $\mathbf{x}'$ «среднее» изображение изображение $\mathbf{x}'$

$\mathbf{x}' = (x'_{\mathbf{s}'}, \mathbf{s}' \in \mathbf{T}')$ $\mathbf{x} = (x_{\mathbf{t}}, \mathbf{t} \in \mathbf{T})$ $\mathbf{x}'' = (x''_{\mathbf{s}''}, \mathbf{s}'' \in \mathbf{T}'')$







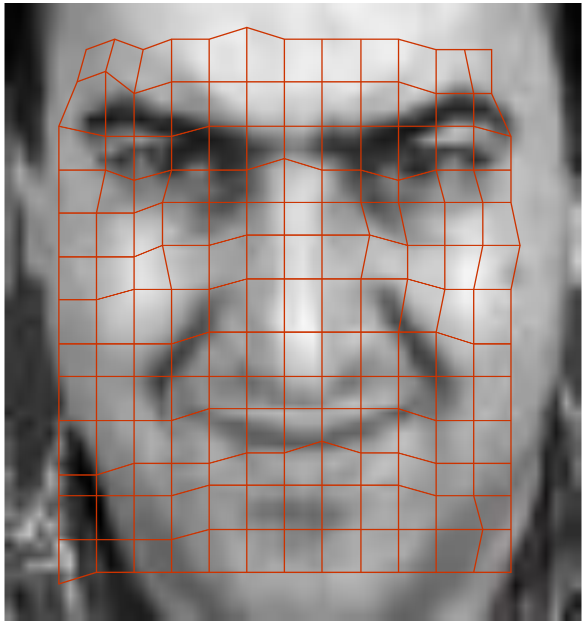
Оптимальная эластичная деформация растров пары изображений

Пример «наивной» реализации линейных операций над изображениями

Пара изображений лица человека

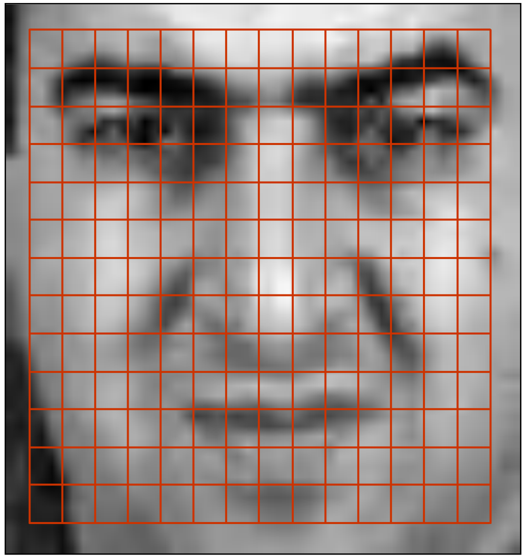
изображение $\mathbf{x}'$

$\mathbf{x}' = \;\; x'_{\mathbf{s}'}, \mathbf{s}' \in \mathbf{T}'$



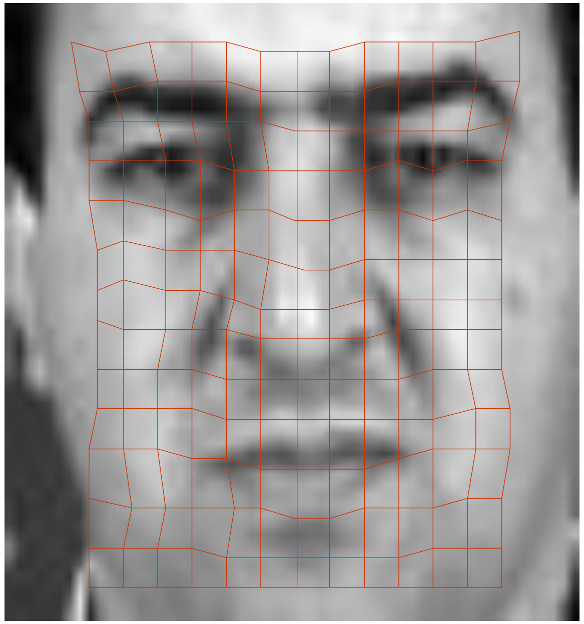
«среднее» изображение

$\mathbf{x} = \;\; x_{\mathbf{t}}, \mathbf{t} \in \mathbf{T}$



изображение $\mathbf{x}'$

$\mathbf{x}'' = \;\; x''_{\mathbf{s}''}, \mathbf{s}'' \in \mathbf{T}''$



Среднее арифметическое изображение

$$\mathbf{x} = \frac{1}{2}(\mathbf{x}' + \mathbf{x}'')$$

Пример «наивной» реализации линейных операций над изображениями

Пара изображений лица человека

изображение $\mathbf{x}'$

$\mathbf{x}' = \ x'_{\mathbf{s}'}, \mathbf{s}' \in \mathbf{T}'$

«среднее» изображение

$\mathbf{x} = \ x_{\mathbf{t}}, \mathbf{t} \in \mathbf{T}$

изображение $\mathbf{x}'$

$\mathbf{x}'' = \ x''_{\mathbf{s}''}, \mathbf{s}'' \in \mathbf{T}''$

Среднее арифметическое изображение

$$\mathbf{x} = \frac{1}{2}(\mathbf{x}' + \mathbf{x}'')$$

Требования, предъявляемые к линейным операциям

- сложение коммутативно $\omega' + \omega'' = \omega'' + \omega'$

  и ассоциативно $(\omega' + \omega'') + \omega''' = \omega' + (\omega'' + \omega''')$;
- существует нулевой элемент $\omega + \phi = \omega$, $c\phi = \phi$;
- для каждого элемента существует обратный ему элемент $(-\omega) + \omega = \phi$;
- умножение ассоциативно $c'(c''\omega) = (c'c'')\omega$;
- умножение на единицу оставляет элемент без изменения $1\omega = \omega$;
- сложение и умножение дистрибутивны $(c' + c'')\omega = c'\omega + c''\omega$, $c(\omega' + \omega'') = c\omega' + c\omega''$;

Требования, предъявляемые к линейным операциям

- сложение коммутативно $\omega' + \omega'' = \omega'' + \omega'$
  и ассоциативно $(\omega' + \omega'') + \omega''' = \omega' + (\omega'' + \omega''')$;
- существует нулевой элемент $\omega + \phi = \omega$, $c\phi = \phi$;
- для каждого элемента существует обратный ему элемент $(-\omega) + \omega = \phi$;
- умножение ассоциативно $c'(c''\omega) = (c'c'')\omega$;
- умножение на единицу оставляет элемент без изменения $1\omega = \omega$;
- сложение и умножение дистрибутивны $(c' + c'')\omega = c'\omega + c''\omega$, $c(\omega' + \omega'') = c\omega' + c\omega''$;

Однако не все эти требования выполняются для произвольного неограниченно выпуклого метрического пространства

Не гарантирована ассоциативность сложения.

Требования, предъявляемые к линейным операциям

- сложение коммутативно $\omega' + \omega'' = \omega'' + \omega'$
  и ассоциативно $\boxed{(\omega' + \omega'') + \omega''' = \omega' + (\omega'' + \omega''')}$;
- существует нулевой элемент $\omega + \phi = \omega$, $c\phi = \phi$;
- для каждого элемента существует обратный ему элемент $(-\omega) + \omega = \phi$;
- умножение ассоциативно $c'(c''\omega) = (c'c'')\omega$;
- умножение на единицу оставляет элемент без изменения $1\omega = \omega$;
- сложение и умножение дистрибутивны $(c' + c'')\omega = c'\omega + c''\omega$, $c(\omega' + \omega'') = c\omega' + c\omega''$;

Однако не все эти требования выполняются для произвольного неограниченно выпуклого метрического пространства

Не гарантирована ассоциативность сложения.

Требования, предъявляемые к линейным операциям

- сложение коммутативно $\omega' + \omega'' = \omega'' + \omega'$
  и ассоциативно $\boxed{(\omega' + \omega'') + \omega''' = \omega' + (\omega'' + \omega''')}$;
- существует нулевой элемент $\omega + \phi = \omega$, $c\phi = \phi$;
- для каждого элемента существует обратный ему элемент $(-\omega) + \omega = \phi$;
- умножение ассоциативно $c'(c''\omega) = (c'c'')\omega$;
- умножение на единицу оставляет элемент без изменения $1\omega = \omega$;
- сложение и умножение дистрибутивны $(c' + c'')\omega = c'\omega + c''\omega$, $c(\omega' + \omega'') = c\omega' + c\omega''$;

Однако не все эти требования выполняются для произвольного неограниченно выпуклого метрического пространства

Не гарантирована ассоциативность сложения.

Необходимое условие: Ординарность метрического пространства, т.е. единственность соосного элемента $x = \mathrm{Coax} < \omega', \omega'' >; c$.

Требования, предъявляемые к линейным операциям

- сложение коммутативно $\omega' + \omega'' = \omega'' + \omega'$
  и ассоциативно $\boxed{(\omega' + \omega'') + \omega''' = \omega' + (\omega'' + \omega''')}$;
- существует нулевой элемент $\omega + \phi = \omega$, $c\phi = \phi$;
- для каждого элемента существует обратный ему элемент $(-\omega) + \omega = \phi$;
- умножение ассоциативно $c'(c''\omega) = (c'c'')\omega$;
- умножение на единицу оставляет элемент без изменения $1\omega = \omega$;
- сложение и умножение дистрибутивны $(c' + c'')\omega = c'\omega + c''\omega$, $c(\omega' + \omega'') = c\omega' + c\omega''$;

Однако не все эти требования выполняются для произвольного неограниченно выпуклого метрического пространства

Не гарантирована ассоциативность сложения.

Необходимое условие: Ординарность метрического пространства, т.е. единственность соосного элемента $x = \mathrm{Coax} < \omega', \omega'' >; c$.

Доказанное достаточное условие базируется на понятии евклидовой метрики.

Евклидовы метрики

Множество объектов реального мира Ω с метрикой $\rho(\omega',\omega'')$.

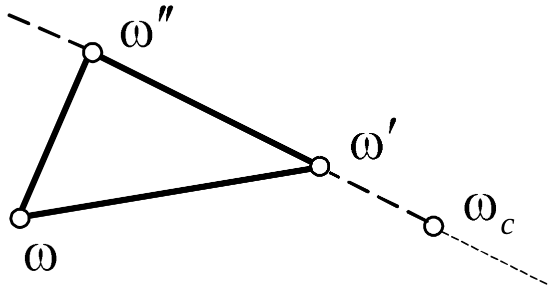
Метрика $\rho(\omega',\omega'')$ на Ω называется евклидовой, если для любого конечного подмножества элементов $\omega_j,\ j=1,\dots,m$ матрица $\left[-\rho^2(\omega_j,\omega_l),\ j,l=1,\dots,m\right]$ условно неотрицательно определена.

Квадратичная форма, образуемая этой матрицей в $\mathbb{R}^m$, неотрицательна на гиперплоскости с нулевой суммой аргументов $\sum_{j=1}^{N}\sum_{l=1}^{N} -\rho^2(\omega_j,\omega_l)\ a_j a_l \geq 0$, если $\sum_{j=1}^{N} a_j = 0$

Теорема 1. Евклидова метрика ординарна – для каждой пары $<\omega',\omega''>\in\Omega\times\Omega$ и коэффициента $c\in\mathbb{R}$ существует не более одного элемента $\omega_c=\mathrm{Coax}\ <\omega',\omega''>;c\ \in\Omega$.

Теорема 2. В метрическом пространстве с евклидовой метрикой для всякого элемента $\omega_c=\mathrm{Coax}\ <\omega',\omega''>;c$ определены значения расстояний до всех элементов $\omega\in\Omega$:

$$\rho^2(\omega,\omega_c)=c^2\rho^2(\omega,\omega')+(1-c)^2\rho^2(\omega,\omega'')-c(1-c)\rho^2(\omega',\omega'')$$



Евклидовы метрики

Множество объектов реального мира Ω с метрикой $\rho(\omega', \omega'')$.

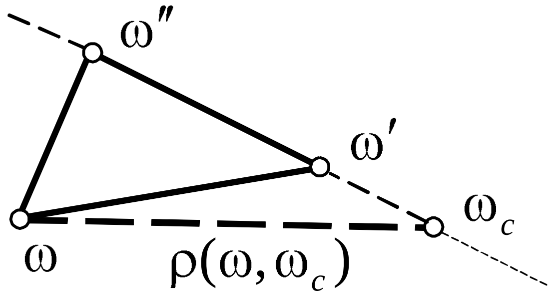
Метрика $\rho(\omega', \omega'')$ на Ω называется евклидовой, если для любого конечного подмножества элементов $\omega_j,\ j = 1, ..., m$ матрица $\left[-\rho^2(\omega_j, \omega_l),\ j, l = 1, ..., m\right]$ условно неотрицательно определена.

Квадратичная форма, образуемая этой матрицей в $\mathbb{R}^m$, неотрицательна на гиперплоскости с нулевой суммой аргументов $\sum_{j=1}^{m}\sum_{l=1}^{m} -\rho^2(\omega_j, \omega_l)\ a_j a_l \geq 0$, если $\sum_{j=1}^{m} a_j = 0$

Теорема 1. Евклидова метрика ординарна – для каждой пары $< \omega', \omega'' > \in \Omega \times \Omega$ и коэффициента $c \in \mathbb{R}$ существует не более одного элемента $\omega_c = \text{Coax} < \omega', \omega'' >; c \in \Omega$.

Теорема 2. В метрическом пространстве с евклидовой метрикой для всякого элемента $\omega_c = \text{Coax} < \omega', \omega'' >; c$ определены значения его расстояний до всех элементов $\omega \in \Omega$:

$$\boxed{\rho^2(\omega, \omega_c)} = c^2 \rho^2(\omega, \omega') + (1-c)^2 \rho^2(\omega, \omega'') - c(1-c)\rho^2(\omega', \omega'')$$



Евклидовы метрики

Множество объектов реального мира Ω с метрикой $\rho(\omega', \omega'')$.

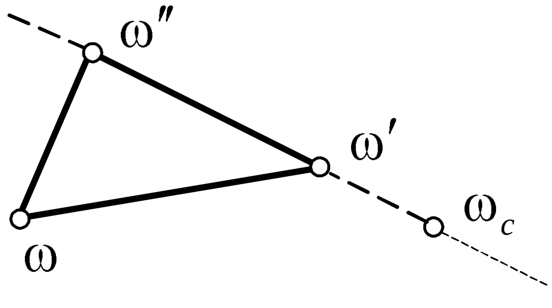
Метрика $\rho(\omega', \omega'')$ на Ω называется евклидовой, если для любого конечного подмножества элементов $\omega_j,\ j = 1, ..., m$ матрица $\left[-\rho^2(\omega_j, \omega_l),\ j, l = 1, ..., m\right]$ условно неотрицательно определена.

Квадратичная форма, образуемая этой матрицей в $\mathbb{R}^m$, неотрицательна на гиперплоскости с нулевой суммой аргументов $\sum_{j=1}^{N}\sum_{l=1}^{N} -\rho^2(\omega_j, \omega_l)\ a_j a_l \geq 0$, если $\sum_{j=1}^{N} a_j = 0$

Теорема 1. Евклидова метрика ординарна – для каждой пары $< \omega', \omega'' > \in \Omega \times \Omega$ и коэффициента $c \in \mathbb{R}$ существует не более одного элемента $\omega_c = \text{Coax} < \omega', \omega'' >; c \in \Omega$.

Теорема 2. В метрическом пространстве с евклидовой метрикой для всякого элемента $\omega_c = \text{Coax} < \omega', \omega'' >; c$ определены значения расстояний до всех элементов $\omega \in \Omega$:

$$\rho^2(\omega, \omega_c) = c^2 \rho^2(\omega, \omega') + (1-c)^2 \rho^2(\omega, \omega'') - c(1-c)\rho^2(\omega', \omega'')$$



Евклидовы метрики

Множество объектов реального мира Ω с метрикой $\rho(\omega',\omega'')$.

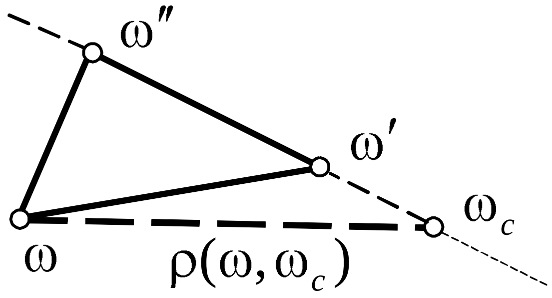
Метрика $\rho(\omega',\omega'')$ на Ω называется евклидовой, если для любого конечного подмножества элементов $\omega_j,\ j=1,\dots,m$ матрица $\left[-\rho^2(\omega_j,\omega_l),\ j,l=1,\dots,m\right]$ условно неотрицательно определена.

Квадратичная форма, образуемая этой матрицей в $\mathbb{R}^m$, неотрицательна на гиперплоскости с нулевой суммой аргументов $\sum_{j=1}^{N}\sum_{l=1}^{N} -\rho^2(\omega_j,\omega_l)\ a_j a_l \ge 0$, если $\sum_{j=1}^{N} a_j = 0$

Теорема 1. Евклидова метрика ординарна – для каждой пары $<\omega',\omega''>\in\Omega\times\Omega$ и коэффициента $c\in\mathbb{R}$ существует не более одного элемента $\omega_c=\text{Coax}\ <\omega',\omega''>;c\ \in\Omega$.

Теорема 2. В метрическом пространстве с евклидовой метрикой для всякого элемента $\omega_c=\text{Coax}\ <\omega',\omega''>;c$ определены значения расстояний до всех элементов $\omega\in\Omega$:

$$\rho^2(\omega,\omega_c)=c^2\rho^2(\omega,\omega')+(1-c)^2\rho^2(\omega,\omega'')-c(1-c)\rho^2(\omega',\omega'')$$



Евклидовы метрики

Множество объектов реального мира Ω с метрикой $\rho(\omega', \omega'')$.

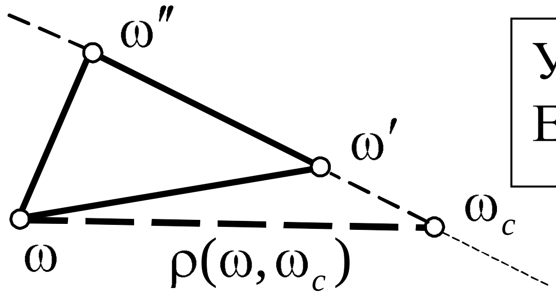
Метрика $\rho(\omega', \omega'')$ на Ω называется евклидовой, если для любого конечного подмножества элементов ω_j, $j = 1, ..., m$ матрица $\left[-\rho^2(\omega_j, \omega_l),\ j, l = 1, ..., m \right]$ условно неотрицательно определена.

Квадратичная форма, образуемая этой матрицей в $\mathbb{R}^m$, неотрицательна на гиперплоскости с нулевой суммой аргументов $\sum_{j=1}^{N} \sum_{l=1}^{N} -\rho^2(\omega_j, \omega_l)\ a_j a_l \geq 0$, если $\sum_{j=1}^{N} a_j = 0$

Теорема 1. Евклидова метрика ординарна – для каждой пары $< \omega', \omega'' > \in \Omega \times \Omega$ и коэффициента $c \in \mathbb{R}$ существует не более одного элемента $\omega_c = \text{Coax } < \omega', \omega'' >; c\ \in \Omega$.

Теорема 2. В метрическом пространстве с евклидовой метрикой для всякого элемента $\omega_c = \text{Coax } < \omega', \omega'' >; c$ определены значения расстояний до всех элементов $\omega \in \Omega$:

$$\rho^2(\omega, \omega_c) = c^2 \rho^2(\omega, \omega') + (1-c)^2 \rho^2(\omega, \omega'') - c(1-c) \rho^2(\omega', \omega'')$$



Условность этого утверждения: Если $\omega_c \in \Omega$ существует!

Пополнение метрического пространства с евклидовой метрикой

Множество объектов реального мира Ω с евклидовой метрикой $\rho(\omega', \omega'')$.

Если для какой-либо пары $< \omega', \omega'' > \Omega$ и какого-либо коэффициента $c \in \mathbb{R}$ в множестве Ω не существует соосного элемента $\omega_c = \mathrm{Coax} \; < \omega', \omega'' >; c$, то расширим это множество, добавив в него такой элемент $\tilde{\Omega} = \Omega \bigcup \{\omega_c\}$.

Так же поступим со всеми парами $< \omega', \omega'' > \Omega$, всеми числами $c \in \mathbb{R}$, и со всеми парами, образуемыми полученными соосными элементами.

Результат: Гипотетическое метрическое пространство $\tilde{\Omega} \supset \Omega$, неограниченное и выпуклое, содержащее исходное множество объектов реального мира.

Неограниченное выпуклое метрическое пространство с евклидовой метрикой будем называть евклидовым метрическим пространством.

Евклидово метрическое пространство $\tilde{\Omega} \supset \Omega$ вместе с любой парой его элементов $< \omega', \omega'' > \tilde{\Omega}$ содержит также и всю определяемую ими ось.

Множество всех элементов, соосных паре элементов $< \omega', \omega'' > \tilde{\Omega}$, $\omega' \neq \omega''$, будем называть осью в $\tilde{\Omega}$, определяемой этой парой:

$\tilde{\Omega}_2(\omega', \omega'') = \;\; \omega_c = \mathrm{Coax} \; < \omega', \omega'' >; c \;\; , c \in \mathbb{R}$

Пополнение метрического пространства с евклидовой метрикой

Множество объектов реального мира Ω с евклидовой метрикой $\rho(\omega', \omega'')$.

Если для какой-либо пары $< \omega', \omega'' > \Omega$ и какого-либо коэффициента $c \in \mathbb{R}$ в множестве Ω не существует соосного элемента $\omega_c = \text{Coax} \;\, < \omega', \omega'' >; c$, то расширим это множество, добавив в него такой элемент $\tilde{\Omega} = \Omega \bigcup \{\omega_c\}$.

Так же поступим со всеми парами $< \omega', \omega'' > \Omega$, всеми числами $c \in \mathbb{R}$, и со всеми парами, образуемыми полученными соосными элементами.

Результат: Гипотетическое метрическое пространство $\tilde{\Omega} \supset \Omega$, неограниченное и выпуклое, содержащее исходное множество объектов реального мира.

Неограниченное выпуклое метрическое пространство с евклидовой метрикой будем называть евклидовым метрическим пространством.

Множество всех элементов, соосных паре элементов $< \omega', \omega'' > \in \tilde{\Omega} \times \tilde{\Omega}$, $\omega' \neq \omega''$, будем называть осью в $\tilde{\Omega}$, определяемой этой парой:

$\tilde{\Omega}_2(\omega', \omega'') = \;\; \omega_c = \text{Coax} \;\, < \omega', \omega'' >; c \;\;, c \in \mathbb{R}$

Евклидово метрическое пространство $\tilde{\Omega} \supset \Omega$ вместе с любой парой его элементов $< \omega', \omega'' > \in \tilde{\Omega} \times \tilde{\Omega}$ содержит также и всю определяемую ими ось.

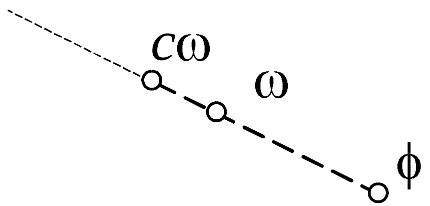
Минимальное евклидово метрическое пространство $\tilde{\Omega} \supset \Omega$, содержащее данное метрическое пространство Ω, будем называть его неограниченным выпуклым замыканием.

Линейные операции в евклидовом метрическом пространстве

Произвольный выбор центра метрического пространства $\phi \in \tilde{\Omega}$.

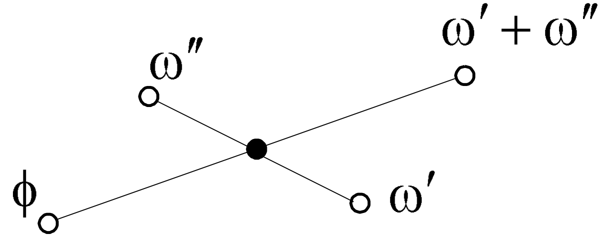
Умножение элемента $\omega \in \tilde{\Omega}$ на коэффициент

$$c\omega = \mathrm{Coax}\ \ <\phi, \omega>; c$$



Сложение элементов $\omega', \omega'' \in \tilde{\Omega}$

$$\omega' + \omega'' = 2\mathrm{Coax}(<\omega', \omega''>; 1/2)$$



Скалярное произведение двух элементов (потенциальная функция, кернел)

$$K_\phi(\omega', \omega'') = \frac{1}{2}\left[\rho^2(\omega', \phi) + \rho^2(\omega'', \phi) - \rho^2(\omega', \omega'')\right]$$

- Сложение коммутативно $\omega' + \omega'' = \omega'' + \omega'$
  и ассоциативно $(\omega' + \omega'') + \omega''' = \omega' + (\omega'' + \omega''')$;
- существует нулевой элемент $\omega + \phi = \omega$, $c\phi = \phi$;
- для каждого элемента существует обратный ему элемент $(-\omega) + \omega = \phi$;
- умножение ассоциативно $c'(c''\omega) = (c'c'')\omega$;
- умножение на единицу оставляет элемент без изменения $1\omega = \omega$;
- сложение и умножение дистрибутивны $(c' + c'')\omega = c'\omega + c''\omega$, $c(\omega' + \omega'') = c\omega' + c\omega''$;
- скалярное произведение линейно, $K_\phi(c'\omega' + c''\omega'', \omega''') = c'K_\phi(\omega', \omega''') + c''K_\phi(\omega'', \omega''')$;
- $K_\phi(\omega, \omega) \geq 0$, причем $K_\phi(\omega, \omega) = 0$ тогда и только тогда, когда $\omega = \phi$;
- $\rho(\omega', \omega'') = \left[\underbrace{K_\phi(\omega', \omega') + K_\phi(\omega'', \omega'') - 2K_\phi(\omega', \omega'')}_{\geq 0}\right]^{1/2}$.

Потенциальная функция (кернел) на множестве объектов, определяемая евклидовой метрикой

Множество объектов реального мира $\omega \in \Omega$

$\Downarrow$

евклидова метрика $\rho(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$:

условно положительно определенные матрицы $\left[-\rho^2(\omega_j, \omega_l),\ j, l = 1, \dots, m \right]$

$\Downarrow$

пополнение метрического пространства Ω всеми соосными элементами $\tilde{\Omega} \supseteq \Omega$

и продолжение евклидовой метрики

$\Downarrow$

выбор центра $\phi \in \tilde{\Omega}$

$\Downarrow$

линейные операции в $\tilde{\Omega}$ с нулевым элементом $\phi \in \tilde{\Omega}$ и скалярным произведением

$K_\phi(\omega', \omega'')$: положительно определенные матрицы $\left[K_\phi(\omega_j, \omega_l),\ j, l = 1, \dots, m \right]$

$\Downarrow$

преобразование кернела при изменении центра $\tilde{\phi} \in \tilde{\Omega}$

$$K_{\tilde{\phi}}(\omega', \omega'') = K_\phi(\omega', \omega'') - K_\phi(\omega', \tilde{\phi}) - K_\phi(\omega'', \tilde{\phi}) + K_\phi(\tilde{\phi}, \tilde{\phi})$$

Евклидова метрика в Ω порождает класс линейных пространств $\tilde{\Omega} \supseteq \Omega$ с разными нулевыми элементами, разными линейными операциями, разными скалярными произведениями, но с одной и той же евклидовой метрикой $\rho(\omega', \omega'')$.

Евклидова метрика на множестве объектов, определяемая потенциальной функцией (кернелом)

Множество объектов реального мира $\omega \in \Omega$

$\Downarrow$

кернел $K(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$:

положительно определенные матрицы $\left[K(\omega_j, \omega_l),\ j, l = 1, ..., m \right]$

$\Downarrow$

евклидова метрика $\rho(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$

$\Downarrow$

пополнение метрического пространства Ω всеми соосными элементами $\tilde{\Omega} \supseteq \Omega$

и продолжение евклидовой метрики

$\Downarrow$

центр автоматически определен как $\phi \in \tilde{\Omega}$: $K(\phi, \phi) = 0$

$\Downarrow$

линейные операции в $\tilde{\Omega}$ с нулевым элементом $\phi \in \tilde{\Omega}$ и скалярным произведением $K(\omega', \omega'')$

$\Downarrow$

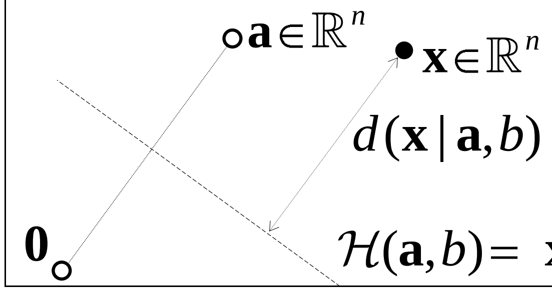
преобразование кернела при изменении центра $\tilde{\phi} \in \tilde{\Omega}$

$$K_{\tilde{\phi}}(\omega', \omega'') = K(\omega', \omega'') - K(\omega', \tilde{\phi}) - K(\omega'', \tilde{\phi}) + K(\tilde{\phi}, \tilde{\phi})$$

Класс эквивалентных кернелов в Ω порождает класс линейных пространств $\tilde{\Omega} \supseteq \Omega$ с разными нулевыми элементами, разными линейными операциями, разными скалярными произведениями, но с одной и той же евклидовой метрикой $\rho(\omega', \omega'')$.

Линейный принцип восстановления зависимостей на основе потенциальной функции (Kernel-based Dependence Estimation)

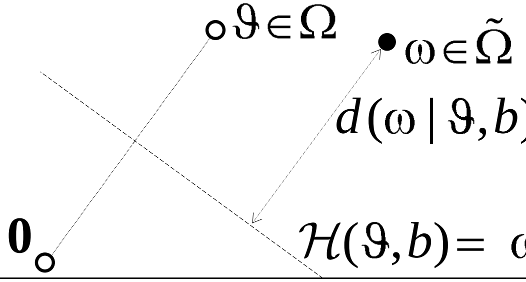
Объекты $\mathbf{x}\in\mathbb{R}^n$ Линейный принцип в $\mathbb{R}^n$:



$$d(\mathbf{x}\,|\,\mathbf{a},b)=\frac{\mathbf{a}^T\mathbf{x}+b}{(\mathbf{a}^T\mathbf{a})^{1/2}}$$

$$\mathcal{H}(\mathbf{a},b)=\ \mathbf{x}':d(\mathbf{x}'\,|\,\mathbf{a},b)=0$$

Объекты $\omega\in\Omega$. Линейный принцип в $\tilde{\Omega}\supset\Omega$:



$$d(\omega\,|\,\vartheta,b)=\frac{K(\vartheta,\omega)+b}{K(\vartheta,\vartheta)^{1/2}}$$

$$\mathcal{H}(\vartheta,b)=\ \omega':d(\omega'\,|\,\vartheta,b)=0$$

Распознавание объектов двух классов, метод опорных векторов

Критерий обучения в $\mathbb{R}^n$:

$$\begin{cases}\boxed{\mathbf{a}^T\mathbf{a}}+C\sum_{j=1}^{N}\delta_j\to\min(\mathbf{a}\in\mathbb{R}^n,b\in\mathbb{R},\delta_1,\ldots,\delta_N\in\mathbb{R}),\\ y_j(\boxed{\mathbf{a}^T\mathbf{x}_j}+b)\ge1-\delta_j,\ \delta_j\ge0,\ j=1,\ldots,N.\end{cases}$$

Двойственная задача:

$$\begin{cases}\sum_{j=1}^{N}\lambda_j-\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l\boxed{\mathbf{x}_j^T\mathbf{x}_l}\lambda_j\lambda_l\to\max,\\ \sum_{j=1}^{N}y_j\lambda_j=0,\ 0\le\lambda_j\le C/2,\ \ j=1,\ldots,N.\end{cases}$$

Правило классификации нового объекта:

$$d(\mathbf{x})\propto\sum_{j:\lambda_j>0}y_j\lambda_j\boxed{\mathbf{x}_j^T\mathbf{x}}+b=\sum_{j:\lambda_j>0}c_j\boxed{\mathbf{x}_j^T\mathbf{x}}+b\gtrless0$$

Критерий обучения в $\tilde{\Omega}$:

$$\begin{cases}\boxed{K(\vartheta,\vartheta)}+C\sum_{j=1}^{N}\delta_j\to\min(\vartheta\in\tilde{\Omega},b\in\mathbb{R},\delta_1,\ldots,\delta_N\in\mathbb{R}),\\ y_j(\boxed{K(\vartheta,\omega_j)}+b)\ge1-\delta_j,\ \delta_j\ge0,\ j=1,\ldots,N.\end{cases}$$

Двойственная задача:

$$\begin{cases}\sum_{j=1}^{N}\lambda_j-\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}y_jy_l\boxed{K(\omega_j,\omega_k)}\lambda_j\lambda_l\to\max,\\ \sum_{j=1}^{N}y_j\lambda_j=0,\ 0\le\lambda_j\le C/2,\ \ j=1,\ldots,N.\end{cases}$$

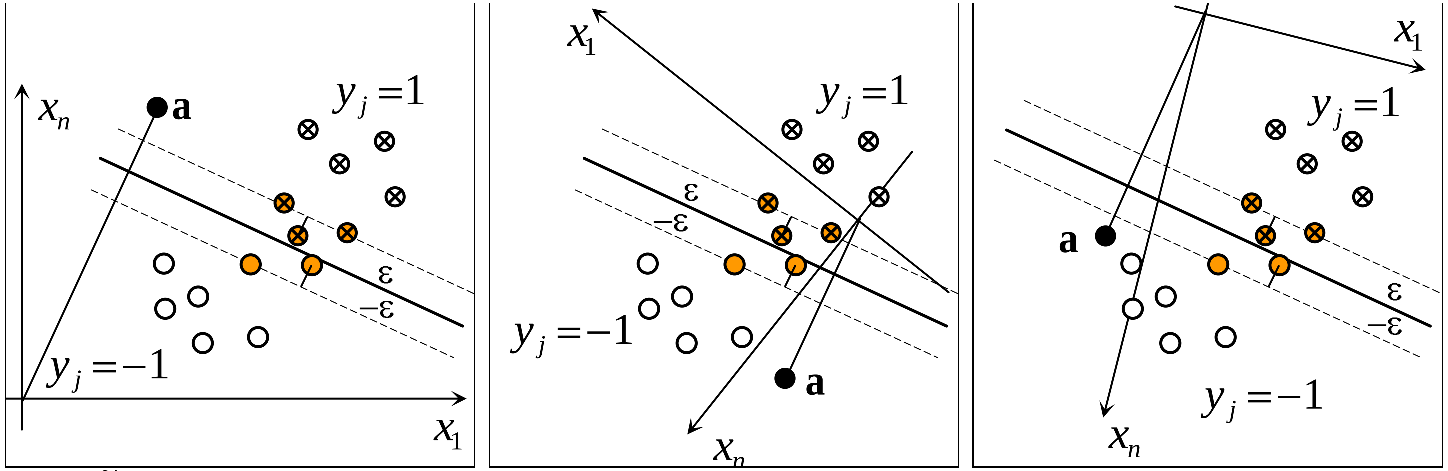
Правило классификации нового объекта:

$$d(\omega)\propto\sum_{j:\lambda_j>0}y_j\lambda_j\boxed{K(\omega_j,\omega)}+b=\sum_{j:\lambda_j>0}c_j\boxed{K(\omega_j,\omega)}+b\gtrless0$$

Напомним: Существует континуум линейных пространств и кернелов, выражающих одну и ту же евклидову метрику, различающихся лишь выбором нулевого элемента.

Правило классификации нового объекта, полученное по обучающей совокупности, определяется только взаимными евклидовыми расстояниями между объектами обучающей совокупности $\rho(\omega_j, \omega_l),\ y(\omega_j),\ \ j, l = 1, ..., N$

В частности, в $\mathbb{R}^n$:
$\rho(\omega_j, \omega_l) = \| \mathbf{x}_j - \mathbf{x}_l \|$



В линейном пространстве $\tilde{\Omega}$, определяемом кернелом:

$$\rho(\omega_j, \omega_l) = \left[K_\phi(\omega_j, \omega_j) + K_\phi(\omega_l, \omega_l) - 2K_\phi(\omega_j, \omega_l) \right]^{1/2}$$

Правило классификации нового объекта:

$$d(\mathbf{x}) \propto \sum_{j:\ опорные\ объекты} c_j \mathbf{x}_j^T \mathbf{x} + b \gtrless 0 \qquad d(\omega) \propto \sum_{j:\ опорные\ объекты} c_j K(\omega_j, \omega) + b \gtrless 0$$

При переносе нуля линейного пространства коэффициенты c_j, определяющие расстояние объекта от метрической гиперплоскости, будут изменяться, но состав множества опорных объектов и само значение расстояния остается неизменным.

Как обойтись без выбора нулевого элемента при погружении множества объектов с евклидовой метрикой в линейное пространство?

Как обеспечить возможность интерполяции непосредственно между элементами евклидова метрического пространства при выборе метрической гиперплоскости?

Аффинные операции в евклидовом метрическом пространстве

Пусть Ω – множество объектов реального мира с евклидовой метрикой $\rho(\omega', \omega'')$:
$\sum_{j=1}^{m}\sum_{l=1}^{m} -\rho^2(\omega_j, \omega_l)\ a_j a_l \geq 0$, если $\sum_{j=1}^{m} a_j = 0$.

Пусть $\tilde{\Omega} \supset \Omega$ – евклидово метрическое пространство, являющееся его неограниченным выпуклым замыканием.

Напомним, что евклидово метрическое пространство $\tilde{\Omega} \supset \Omega$ вместе с любой парой его элементов $< \omega', \omega'' > \in \tilde{\Omega} \times \tilde{\Omega}$ содержит также и всю определяемую ими ось $\tilde{\Omega}_2(\omega', \omega'') = \ \omega_c = \mathrm{Coax}\ < \omega', \omega'' >; c\ \ , c \in \mathbb{R}\ \subset \tilde{\Omega}$.

Обобщим понятие соосности на произвольную конечную неупорядоченную совокупность элементов евклидова метрического пространства $\ \omega_1, ..., \omega_m\ \subset \tilde{\Omega}$.

Пусть $\mathbf{c} = (c_1 \cdots c_m)^T \in \mathbb{R}^m$ – числовой вектор, такой что условию $\sum_{j=1}^{m} c_j = 1$, т.е. $\mathbf{1}^T \mathbf{c} = 1$.

Рассмотрим семейство функций $f(\omega \mid \omega_1, ..., \omega_m; c_1, ..., c_m) = \sum_{j=1}^{m} c_j \rho^2(\omega_j, \omega)$, $\tilde{\Omega} \underset{c_1,...,c_m}{\overset{\omega_1,...,\omega_m}{\rightarrow}} \mathbb{R}$.

Теорема. В евклидовом метрическом пространстве существует единственный элемент $\omega_{\mathbf{c}} = \arg\min\limits_{\omega \in \tilde{\Omega}} \sum_{j=1}^{m} c_j \rho^2(\omega_j, \omega)$.

Определение. Элемент $\omega_{\mathbf{c}} \in \tilde{\Omega}$ называется аффинной комбинацией элементов $\omega_1, ..., \omega_m\ \subset \tilde{\Omega}$ с коэффициентами $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{1}^T \mathbf{c} = 1$, и обозначается $\omega_{\mathbf{c}} = \mathrm{A}_{j=1}^{m} c_j \omega_j$.

Аффинные операции в евклидовом метрическом пространстве

Пусть $\tilde{\Omega} \supset \Omega$ – евклидово метрическое пространство, являющееся неограниченным выпуклым замыканием множества объектов реального мира Ω.

Напомним, что евклидово метрическое пространство $\tilde{\Omega} \supset \Omega$ вместе с любой парой его элементов $<\omega', \omega''> \in \tilde{\Omega} \times \tilde{\Omega}$ содержит также и всю определяемую ими ось $\tilde{\Omega}_2(\omega', \omega'') = \{\omega_c = \mathrm{Coax}(<\omega', \omega''>; c), c \in \mathbb{R}\} \subset \tilde{\Omega}$.

Обобщим понятие соосности на произвольную конечную неупорядоченную совокупность элементов евклидова метрического пространства $\{\omega_1, ..., \omega_m\} \subset \tilde{\Omega}$.

Пусть $\mathbf{c} = (c_1 \cdots c_m)^T \in \mathbb{R}^m$ – числовой вектор, такой что условию $\sum_{j=1}^m c_j = 1$, т.е. $\mathbf{1}^T \mathbf{c} = 1$.

Рассмотрим семейство функций $f(\omega \mid \omega_1, ..., \omega_m; c_1, ..., c_m) = \sum_{j=1}^m c_j \rho^2(\omega_j, \omega)$, $\tilde{\Omega} \overset{\omega_1, ..., \omega_m}{\underset{c_1, ..., c_m}{\rightarrow}} \mathbb{R}$.

Теорема. В евклидовом метрическом пространстве существует единственный элемент $\omega_{\mathbf{c}} = \arg\min_{\omega \in \tilde{\Omega}} \sum_{j=1}^m c_j \rho^2(\omega_j, \omega)$.

Определение. Элемент $\omega_{\mathbf{c}} \in \tilde{\Omega}$ называется аффинной комбинацией элементов $\{\omega_1, ..., \omega_m\} \subset \tilde{\Omega}$ с коэффициентами $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{1}^T \mathbf{c} = 1$, и обозначается $\omega_{\mathbf{c}} = \mathrm{A}_{j=1}^m c_j \omega_j$.

В частности, соосный элемент $\omega_c = \mathrm{Coax}(<\omega', \omega''>; c)$ является аффинной комбинацией двух элементов ω_1 и ω_2 с коэффициентами $c_1 = c$ и $c_2 = 1 - c$.

Аффинные операции в евклидовом метрическом пространстве

Пусть $\tilde{\Omega}\supset\Omega$ – евклидово метрическое пространство, являющееся неограниченным выпуклым замыканием множества объектов реального мира Ω.

Напомним, что евклидово метрическое пространство $\tilde{\Omega}\supset\Omega$ вместе с любой парой его элементов $<\omega',\omega''>\in\tilde{\Omega}\times\tilde{\Omega}$ содержит также и всю определяемую ими ось $\tilde{\Omega}_2(\omega',\omega'')=\ \omega_c=\mathrm{Coax}\ <\omega',\omega''>;c\ \ ,c\in\mathbb{R}\ \ \subset\tilde{\Omega}$.

Обобщим понятие соосности на произвольную конечную неупорядоченную совокупность элементов евклидова метрического пространства $\omega_1,...,\omega_m\ \ \subset\tilde{\Omega}$.

Пусть $\mathbf{c}=(c_1\cdots c_m)^T\in\mathbb{R}^m$ – числовой вектор, такой что условию $\sum_{j=1}^m c_j=1$, т.е. $\mathbf{1}^T\mathbf{c}=1$.

Рассмотрим семейство функций $f(\omega\,|\,\omega_1,...,\omega_m;c_1,...,c_m)=\sum_{j=1}^m c_j\rho^2(\omega_j,\omega)$, $\tilde{\Omega}\underset{c_1,...,c_m}{\overset{\omega_1,...,\omega_m}{\rightarrow}}\mathbb{R}$.

Теорема. В евклидовом метрическом пространстве существует единственный элемент $\omega_{\mathbf{c}}=\arg\min\limits_{\omega\in\tilde{\Omega}}\sum_{j=1}^m c_j\rho^2(\omega_j,\omega)$.

Определение. Элемент $\omega_{\mathbf{c}}\in\tilde{\Omega}$ называется аффинной комбинацией элементов $\omega_1,...,\omega_m\ \ \subset\tilde{\Omega}$ с коэффициентами $\mathbf{c}\in\mathbb{R}^m$, $\mathbf{1}^T\mathbf{c}=1$, и обозначается $\omega_{\mathbf{c}}=\mathrm{A}_{j=1}^m c_j\omega_j$.

Теорема. Евклидово расстояние между произвольным элементом $\omega\in\tilde{\Omega}$ и заданной аффинной комбинацией $\omega_{\mathbf{c}}=\mathrm{A}_{j=1}^m c_j\omega_j$, $\omega_1,...,\omega_m\ \ \subset\tilde{\Omega}$, $\mathbf{1}^T\mathbf{c}=1$, определяется выражением

$$\rho^2(\omega_{\mathbf{c}},\omega)=\sum_{j=1}^m c_j\rho^2(\omega_j,\omega)-(1/2)\sum_{j=1}^m\sum_{l=1}^m c_jc_l\rho^2(\omega_j,\omega_l)$$

Евклидово аффинное пространство

Аффинная комбинация конечной совокупности элементов евклидова метрического пространства $\omega_1,...,\omega_m \subset \tilde{\Omega}$ с коэффициентами $\mathbf{c}=(c_1 \cdots c_m)^T \in \mathbb{R}^m$, $\sum_{j=1}^m c_j = 1$:

$$\omega_{\mathbf{c}} = \mathrm{A}_{j=1}^m c_j \omega_j = \arg\min_{\omega \in \tilde{\Omega}} \sum_{j=1}^m c_j \rho^2(\omega_j, \omega) \in \tilde{\Omega}.$$

Евклидово расстояние между произвольным элементом $\omega \in \tilde{\Omega}$ и заданной аффинной комбинацией $\omega_{\mathbf{c}} = \mathrm{A}_{j=1}^m c_j \omega_j$ определяется выражением

$$\rho^2(\omega_{\mathbf{c}}, \omega) = \sum_{j=1}^m c_j \rho^2(\omega_j, \omega) - \frac{1}{2} \sum_{j=1}^m \sum_{l=1}^m c_j c_l \rho^2(\omega_j, \omega_l)$$

Наличие операции аффинной комбинации любого конечного числа элементов позволяет называть евклидово метрическое пространство $\tilde{\Omega}$ евклидовым аффинным пространством. От евклидова линейного пространства его отличает только отсутствие нулевого элемента. Если дополнительно назначить любой элемент в качестве нулевого $\phi \in \tilde{\Omega}$, то евклидова метрика $\rho(\omega', \omega'')$ порождает в $\tilde{\Omega}$, во-первых, линейные операции сложения двух элементов и умножения элемента на действительный коэффициент, и, во-вторых, скалярное произведение (кернел): $K_\phi(\omega', \omega'') = (1/2)\left[\rho^2(\omega', \phi) + \rho^2(\omega'', \phi) - \rho^2(\omega', \omega'')\right]$.

Евклидова метрика $\rho(\omega', \omega'')$ порождает континуум разных линейных пространств и кернелов, но все они определяют одну и ту же исходную метрику

$$\rho(\omega', \omega'') = \left[K_\phi(\omega', \omega') + K_\phi(\omega'', \omega'') - 2K_\phi(\omega', \omega'')\right]^{1/2}.$$

Диполь и аффинная гиперплоскость в евклидовом метрическом (аффинном) пространство

Евклидово метрическое (аффинное) пространство $\tilde{\Omega} \supset \Omega$ с метрикой $\rho(\omega', \omega'')$.

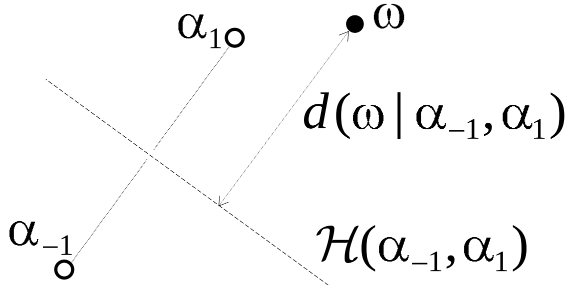
Диполь $< \alpha_{-1}, \alpha_1 > \in \tilde{\Omega} \times \tilde{\Omega}$, $\quad \alpha_{-1}, \alpha_1 \in \tilde{\Omega}$ – узлы диполя.

$\tilde{\Omega}(\alpha_{-1}, \alpha_1) = \ \vartheta \in \tilde{\Omega} : \rho(\alpha_{-1}, \vartheta) = \rho(\alpha_1, \vartheta)$ – аффинная гиперплоскость.

$\omega \in \tilde{\Omega}$ – произвольный элемент евклидова метрического пространства,
$d(\omega \,|\, \alpha_{-1}, \alpha_1)$ – его расстояние до аффинной гиперплоскости с учетом знака.

Теорема.

$$d(\omega \,|\, \alpha_{-1}, \alpha_1) = \frac{\rho^2(\alpha_{-1}, \omega) - \rho^2(\alpha_1, \omega)}{2\rho^2(\alpha_{-1}, \alpha_1)} \rho(\alpha_{-1}, \alpha_1).$$



Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\quad \rho(\omega_j, \omega_l),\ y(\omega_j) = \pm 1;\ \ j, l = 1, \dots, N$

$$\begin{cases} \dfrac{1}{\varepsilon^2} + C \sum_{j=1}^{N} \delta_j \to \min(\alpha_{-1}, \alpha_1 \in \tilde{\Omega}, \varepsilon, \delta_1, \dots, \delta_N), \\ d(\omega \,|\, \alpha_{-1}, \alpha_1) \geq \varepsilon - \delta_j,\ \delta_j \geq 0,\ j = 1, \dots, N,\ \rho(\alpha_{-1}, \alpha_1) = 1. \end{cases}$$

Представление искомых узлов диполя:

$$\alpha_{-1} = \mathrm{A}_{j=1}^{N} q_{-1,j} \omega_j, \qquad \alpha_1 = \mathrm{A}_{j=1}^{N} q_{1,j} \omega_j,$$

$$q_{-1,j} = g_j - (\tfrac{1}{2} - \boxed{c}) \boxed{a_j},\ q_{1,j} = g_j + (\tfrac{1}{2} + \boxed{c}) \boxed{a_j},$$

$\sum_{j=1}^{N} g_j = 1$ – любые числа, $\sum_{j=1}^{N} \boxed{a_j} = 0.$

Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\rho(\omega_j,\omega_l),\ y(\omega_j)=\pm 1;\ \ j,l=1,\dots,N$

$$\begin{cases}\dfrac{1}{\varepsilon^2}+C\sum_{j=1}^{N}\delta_j \to \min(\alpha_{-1},\alpha_1\in\tilde{\Omega},\varepsilon,\delta_1,\dots,\delta_N),\\ d(\omega\,|\,\alpha_{-1},\alpha_1)\ge\varepsilon-\delta_j,\ \delta_j\ge 0,\ j=1,\dots,N,\ \rho(\alpha_{-1},\alpha_1)=1.\end{cases}$$

Представление искомых узлов диполя:

$$\alpha_{-1}=\mathrm{A}_{j=1}^{N}q_{-1,j}\omega_j,\qquad \alpha_1=\mathrm{A}_{j=1}^{N}q_{1,j}\omega_j,$$

$$q_{-1,j}=g_j-(\tfrac{1}{2}-\boxed{c})\boxed{a_j},\ q_{1,j}=g_j+(\tfrac{1}{2}+\boxed{c})\boxed{a_j},$$

$$\sum_{j=1}^{N}g_j=1 \text{ – любые числа, } \sum_{j=1}^{N}\boxed{a_j}=0.$$

В условиях принятых предположений достаточно искать дискриминантную функцию в виде

$$d(\omega\,|\,a_1,\dots,a_N,b)=\frac{1}{2}\left(\sum_{j=1}^{N}a_j\ \ -\rho^2(\omega_j,\omega)\ \ +b\right)$$

Эквивалентная формулировка задачи:

$$\begin{cases}\dfrac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\ -\rho^2(\omega_j,\omega_l)\ \ a_j a_l+C\sum_{j=1}^{N}\delta_j\to\min(a_1,\dots,a_N,b,\delta_1,\dots,\delta_N),\\ y_j\left[\sum_{l=1}^{N}a_l\ \ -\rho^2(\omega_j,\omega_l)\ \ +b\right]\ge 1-\delta_j,\ \delta_j\ge 0,\ j=1,\dots,N.\end{cases}$$

Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\rho(\omega_j,\omega_l),\ y(\omega_j)=\pm 1;\ \ j,l=1,...,N$

$$\begin{cases} \dfrac{1}{\varepsilon^2}+C\sum_{j=1}^{N}\delta_j \to \min(\alpha_{-1},\alpha_1\in\tilde{\Omega},\varepsilon,\delta_1,...,\delta_N), \\ d(\omega\,|\,\alpha_{-1},\alpha_1)\ge\varepsilon-\delta_j,\ \delta_j\ge 0,\ j=1,...,N,\ \rho(\alpha_{-1},\alpha_1)=1. \end{cases}$$

Представление искомых узлов диполя:

$$\alpha_{-1}=\mathrm{A}_{j=1}^{N}q_{-1,j}\omega_j,\qquad \alpha_1=\mathrm{A}_{j=1}^{N}q_{1,j}\omega_j,$$

$$q_{-1,j}=g_j-(\tfrac{1}{2}-c)\boxed{a_j},\ q_{1,j}=g_j+(\tfrac{1}{2}+c)\boxed{a_j},$$

$$\sum_{j=1}^{N}g_j=1 \text{ – любые числа},\ \ \sum_{j=1}^{N}\boxed{a_j}=0.$$

В условиях принятых предположений достаточно искать дискриминантную функцию в виде

$$d(\omega\,|\,a_1,...,a_N,b)=\frac{1}{2}\left(\sum_{j=1}^{N}a_j\ \ -\rho^2(\omega_j,\omega)\ \ +b\right)$$

Эквивалентная формулировка задачи:

$$\begin{cases} \dfrac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\ -\rho^2(\omega_j,\omega_l)\ \ a_j a_l+C\sum_{j=1}^{N}\delta_j\to\min(a_1,...,a_N,b,\delta_1,...,\delta_N), \\ y_j\left[\sum_{l=1}^{N}a_l\ \ -\rho^2(\omega_j,\omega_l)\ \ +b\right]\ge 1-\delta_j,\ \delta_j\ge 0,\ j=1,...,N. \end{cases}$$

Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\rho(\omega_j,\omega_l),\ y(\omega_j)=\pm 1;\ \ j,l=1,...,N$

Эквивалентная формулировка задачи:

$$\begin{cases} \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\ -\rho^2(\omega_j,\omega_l)\ \ a_j a_l + C\sum_{j=1}^{N}\delta_j \to \min(a_1,...,a_N,b,\delta_1,...,\delta_N), \\ y_j\left[\sum_{l=1}^{N} a_l\ \ -\rho^2(\omega_j,\omega_l)\ \ +b\right] \geq 1-\delta_j,\ \delta_j \geq 0,\ j=1,...,N. \end{cases}$$

Это задача квадратичного программирования

Двойственная задача:

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l\ \ -\rho^2(\omega_j,\omega_l)\ \ \lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\leq\lambda_j\leq C/2,\ j=1,...,N. \end{cases}$$

Для сравнения:
двойственная задача Kernel-based SVM

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l K(\omega_j,\omega_k)\ \lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\leq\lambda_j\leq C/2,\ j=1,...,N. \end{cases}$$

Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\rho(\omega_j,\omega_l),\ y(\omega_j)=\pm 1;\ \ j,l=1,...,N$

Эквивалентная формулировка задачи:

$$\begin{cases} \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\ -\rho^2(\omega_j,\omega_l)\ \ a_j a_l + C\sum_{j=1}^{N}\delta_j \to \min(a_1,...,a_N,b,\delta_1,...,\delta_N), \\ y_j\left[\sum_{l=1}^{N} a_l\ \ -\rho^2(\omega_j,\omega_l)\ \ +b\right]\geq 1-\delta_j,\ \delta_j \geq 0,\ j=1,...,N. \end{cases}$$

Это задача квадратичного программирования

Двойственная задача:

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l \boxed{\ -\rho^2(\omega_j,\omega_l)\ } \lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\leq\lambda_j\leq C/2,\ j=1,...,N. \end{cases}$$

Для сравнения:
двойственная задача Kernel-based SVM

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l \boxed{K(\omega_j,\omega_k)} \lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\leq\lambda_j\leq C/2,\ j=1,...,N. \end{cases}$$

Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\rho(\omega_j,\omega_l),\ y(\omega_j)=\pm 1;\ \ j,l=1,\dots,N$

Эквивалентная формулировка задачи:

$$\begin{cases} \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\ -\rho^2(\omega_j,\omega_l)\ \ a_j a_l + C\sum_{j=1}^{N}\delta_j \to \min(a_1,\dots,a_N,b,\delta_1,\dots,\delta_N), \\ y_j\left[\sum_{l=1}^{N} a_l\ \ -\rho^2(\omega_j,\omega_l)\ \ +b\right] \geq 1-\delta_j,\ \delta_j \geq 0,\ j=1,\dots,N. \end{cases}$$

Это задача квадратичного программирования

Двойственная задача:

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l\ \ -\rho^2(\omega_j,\omega_l)\ \ \lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0 \leq \lambda_j \leq C/2,\ j=1,\dots,N. \end{cases}$$

Правило классификации нового объекта:

$$d(\omega) \propto \sum_{j:\lambda_j>0} y_j\lambda_j\ \ -\rho^2(\omega_j,\omega)\ \ +b \gtrless 0$$

Для сравнения:
двойственная задача Kernel-based SVM

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l K(\omega_j,\omega_k)\,\lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0 \leq \lambda_j \leq C/2,\ j=1,\dots,N. \end{cases}$$

Правило классификации нового объекта:

$$d(\omega) \propto \sum_{j:\lambda_j>0} y_j\lambda_j K(\omega_j,\omega) + b \gtrless 0$$

Задача обучения распознаванию объектов двух классов: Принцип максимизации зазора (аналог задачи SVM)

Обучающая совокупность: $\rho(\omega_j,\omega_l),\ y(\omega_j)=\pm 1;\ \ j,l=1,\dots,N$

Эквивалентная формулировка задачи:

$$\begin{cases} \dfrac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} \ \ -\rho^2(\omega_j,\omega_l)\ \ a_j a_l + C\sum_{j=1}^{N}\delta_j \to \min(a_1,\dots,a_N,b,\delta_1,\dots,\delta_N), \\ y_j\left[\sum_{l=1}^{N} a_l \ \ -\rho^2(\omega_j,\omega_l)\ \ +b\right]\ge 1-\delta_j,\ \delta_j\ge 0,\ j=1,\dots,N. \end{cases}$$

Это задача квадратичного программирования

Двойственная задача:

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \dfrac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l \ \ -\rho^2(\omega_j,\omega_l)\ \ \lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\le\lambda_j\le C/2,\ j=1,\dots,N. \end{cases}$$

Правило классификации нового объекта:

$$d(\omega)\propto \sum_{j:\lambda_j>0} y_j\lambda_j \ \ -\rho^2(\omega_j,\omega)\ \ +b \ = \\ = \sum_{j=1}^{N} c_j\rho^2(\omega_j,\omega)+b \gtrless 0$$

Для сравнения:
двойственная задача Kernel-based SVM

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - \dfrac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l K(\omega_j,\omega_k)\,\lambda_j\lambda_l \to \max, \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0\le\lambda_j\le C/2,\ j=1,\dots,N. \end{cases}$$

Правило классификации нового объекта:

$$d(\omega)\propto \sum_{j:\lambda_j>0} y_j\lambda_j K(\omega_j,\omega)+b \ = \\ = \sum_{j=1}^{N} c_j K(\omega_j,\omega)+b \gtrless 0$$

Задача обучения для произвольной функции попарного сравнения объектов: *Relational Dependence Estimation*

Доступна лишь функция попарного сравнения объектов $S(\omega',\omega''): \Omega\times\Omega\to\mathbb{S}\subseteq\mathbb{R}$

Обучающая совокупность: $\Omega^* = \ \omega_j,\ j=1,\ldots,N\ \ :\ \ S(\omega_j,\omega_l),\ y_j = y(\omega_j),\ j,l=1,\ldots,N$

Вектор вторичных признаков объекта $\omega\in\Omega$ относительно обучающей совокупности:
$\mathbf{x}(\omega)=\ x_k(\omega)=S(\omega_k,\omega),\ k=1,\ldots,N\ \in\mathbb{R}^N$.

Классический критерий обучения SVM:

$$\begin{cases}\sum_{k=1}^{N} a_k^2 + C\sum_{j=1}^{N}\delta_j \to \min(a_1,\ldots,a_{N^0},b,\delta_1,\ldots,\delta_N),\\ y_j\ \ \sum_{k=1}^{N} a_k S(\omega_k,\omega_j)+b\ \ \geq 1-\delta_j,\ \delta_j\geq 0,\ j=1,\ldots,N.\end{cases}$$

Обучение проводится по N объектам обучающей совокупности, представленным N вторичными признаками относительно обучающих (базисных) объектов.

$$d(\omega)=\sum_{k=1}^{N}\underbrace{\sum_{j:\lambda_j>0} y_j\lambda_j S(\omega_k,\omega_j)}_{a_k}\ S(\omega_k,\omega)+b=$$

$$=\sum_{k=1}^{N} a_k S(\omega_k,\omega)+b\gtrless 0$$

Результат обучения: Дискриминантная функция, применимая к произвольному объекту $\omega\in\Omega$

В дискриминантной функции участвуют все объекты обучающей совокупности
(R. Duin: Relational Discriminant Analysis)

Обучение для заданной функции попарного сравнения с отбором базисных объектов: *Relevance Object Machine*

Michael Tipping, Christopher Bishop: Оставить лишь наиболее информативные (релевантные) базисные объекты $a_k \neq 0$, игнорируя остальные $a_k = 0$.

Искомое решающее правило: $d(\omega) = \sum_{k:\ \textit{релевантные объекты}} a_k S(\omega_k, \omega) + b \gtrless 0$

Идея Tipping&Bishop:
Априорное вероятностное представление об искомом векторе коэффициентов $\mathbf{a}$

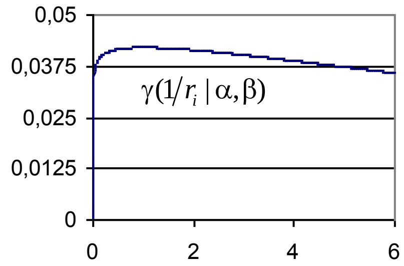
$$\mathcal{N}(\mathbf{a} \mid \mathbf{r}) = \prod_{k=1}^{N} \mathcal{N}(a_k \mid 0, 1/r_k)$$

чем меньше r_k, тем ближе априори a_k к нулю

Дополнительное предположение Tipping&Bishop: Меры точности $1/r_k$ также априори независимы, случайны и распределены по гамма закону

$$G(\mathbf{r} \mid \alpha, \beta) = \prod_{k=1}^{N} \gamma(1/r_k \mid \alpha, \beta) = \prod_{k=1}^{N} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1/r_k)^{\alpha-1} \exp\left(-\beta(1/r_k)\right)$$

Если $\alpha, \beta \to \infty$ и $\alpha/\beta \to \infty$, то гамма-распределение $\gamma(1/r \mid \alpha, \beta) \to$ к равномерному на $(0, \infty)$



Полное априорное распределение вектора коэффициентов $\mathbf{a} = (a_1, \ldots, a_N)$:

$$\Psi(\mathbf{a} \mid \alpha, \beta) = \int_{\mathbb{R}^N} \mathcal{N}(\mathbf{a} \mid \mathbf{r}) G(\mathbf{r} \mid \alpha, \beta) d\mathbf{r}$$

Обучение для заданной функции попарного сравнения с отбором базисных объектов: *Relevance Object Machine*

Michael Tipping, Christopher Bishop: Оставить лишь наиболее информативные (релевантные) базисные объекты $a_k \neq 0$, игнорируя остальные $a_k = 0$.

Искомое решающее правило: $d(\omega) = \sum\limits_{k:\ \textit{релевантные объекты}} a_k S(\omega_k, \omega) + b \gtrless 0$

Идея Tipping&Bishop:
Априорное вероятностное представление об искомом векторе коэффициентов **a**

$\mathcal{N}(\mathbf{a} \mid \mathbf{r}) = \prod_{k=1}^{N} \mathcal{N}(a_k \mid 0, 1/r_k)$
чем меньше r_k, тем ближе априори a_k к нулю

Дополнительное предположение Tipping&Bishop: Меры точности $1/r_k$ также априори независимы, случайны и распределены по гамма закону

$G(\mathbf{r} \mid \alpha, \beta) = \prod_{k=1}^{N} \gamma(1/r_k \mid \alpha, \beta) =$
$\prod_{k=1}^{N} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1/r_k)^{\alpha-1} \exp\left(-\beta(1/r_k)\right)$

Полное априорное распределение вектора коэффициентов $\mathbf{a} = (a_1, ..., a_N)$:

$\Psi(\mathbf{a} \mid \alpha, \beta) = \int_{\mathbb{R}^N} \mathcal{N}(\mathbf{a} \mid \mathbf{r}) G(\mathbf{r} \mid \alpha, \beta) d\mathbf{r}$, $G(\mathbf{r} \mid \alpha, \beta)$ близко к равномерному распределению

Регуляризованный критерий обучения
$$\begin{cases} -\ln \int \mathcal{N}(\mathbf{a} \mid \mathbf{r}) G(\mathbf{r} \mid \alpha, \beta) d\mathbf{r} + C \sum_{j=1}^{N} \delta_j \to \min(a_1, ..., a_{N^0}, b, \delta_1, ..., \delta_N), \\ y_j \left(\sum_{k=1}^{N} a_k S(\omega_k, \omega_j) + b \right) \geq 1 - \delta_j,\ \delta_j \geq 0,\ j = 1, ..., N. \end{cases}$$

Далее Tipping&Bishop применяют ЕМ-алгоритм. При этом приходится прибегать к эвристикам из-за наличия ограничений типа неравенств. Задача невыпукла.

Обучение для заданной функции попарного сравнения с отбором базисных объектов: *Relevance Object Machine*

Michael Tipping, Christopher Bishop: Оставить лишь наиболее информативные (релевантные) базисные объекты $a_k \neq 0$, игнорируя остальные $a_k = 0$.

Искомое решающее правило: $d(\omega) = \sum_{k:\ \textit{релевантные объекты}} a_k S(\omega_k, \omega) + b \gtrless 0$

Идея Tipping&Bishop:
Априорное вероятностное представление об искомом векторе коэффициентов **a**

$\mathcal{N}(\mathbf{a}\,|\,\mathbf{r}) = \prod_{k=1}^{N} \mathcal{N}(a_k\,|\,0, 1/r_k)$
чем меньше r_k, тем ближе априори a_k к нулю

Дополнительное предположение Tipping&Bishop: Меры точности $1/r_k$ также априори независимы, случайны и распределены по гамма закону

$G(\mathbf{r}\,|\,\alpha,\beta) = \prod_{k=1}^{N} \gamma(1/r_k\,|\,\alpha,\beta) = \prod_{k=1}^{N} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1/r_k)^{\alpha-1} \exp\left(-\beta(1/r_k)\right)$

Полное априорное распределение вектора коэффициентов $\mathbf{a} = (a_1, ..., a_N)$:

$\Psi(\mathbf{a}\,|\,\alpha,\beta) = \int_{\mathbb{R}^N} \mathcal{N}(\mathbf{a}\,|\,\mathbf{r}) G(\mathbf{r}\,|\,\alpha,\beta) d\mathbf{r}$, $\quad G(\mathbf{r}\,|\,\alpha,\beta)$ близко к равномерному распределению

Регуляризованный критерий обучения
$$\begin{cases} -\ln \int \mathcal{N}(\mathbf{a}\,|\,\mathbf{r}) G(\mathbf{r}\,|\,\alpha,\beta) d\mathbf{r} + C \sum_{j=1}^{N} \delta_j \to \min(a_1, ..., a_{N^0}, b, \delta_1, ..., \delta_N), \\ y_j \left(\sum_{k=1}^{N} a_k S(\omega_k, \omega_j) + b \right) \geq 1 - \delta_j,\ \delta_j \geq 0,\ j = 1, ..., N. \end{cases}$$

Такой метод обучения крайне селективен по отношению к базисным объектам – в качестве релевантных остаются лишь очень малое число объектов.

Обучение для заданной функции попарного сравнения с отбором базисных объектов: *Relevance Object Machine*

Michael Tipping, Christopher Bishop: Оставить лишь наиболее информативные (релевантные) базисные объекты $a_k \neq 0$, игнорируя остальные $a_k = 0$.

Искомое решающее правило: $\boxed{d(\omega) = \sum_{k:\ \textit{релевантные объекты}} a_k S(\omega_k, \omega) + b \gtrless 0}$

Идея Tipping&Bishop:
Априорное вероятностное представление об искомом векторе коэффициентов $\mathbf{a}$

$\mathcal{N}(\mathbf{a} \,|\, \mathbf{r}) = \prod_{k=1}^{N} \mathcal{N}(a_k \,|\, 0, 1/r_k)$
чем меньше r_k, тем ближе априори a_k к нулю

Дополнительное предположение Tipping&Bishop: Меры точности $1/r_k$ также априори независимы, случайны и распределены по гамма закону

$G(\mathbf{r} \,|\, \alpha, \beta) = \prod_{k=1}^{N} \gamma(1/r_k \,|\, \alpha, \beta) =$
$\prod_{k=1}^{N} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1/r_k)^{\alpha-1} \exp\left(-\beta(1/r_k)\right)$

Полное априорное распределение вектора коэффициентов $\mathbf{a} = (a_1, ..., a_N)$:

$\Psi(\mathbf{a} \,|\, \alpha, \beta) = \int_{\mathbb{R}^N} \mathcal{N}(\mathbf{a} \,|\, \mathbf{r}) G(\mathbf{r} \,|\, \alpha, \beta) d\mathbf{r}$, $\quad G(\mathbf{r} \,|\, \alpha, \beta)$ близко к равномерному распределению

Регуляризованный критерий обучения

$$\begin{cases} -\ln \int \mathcal{N}(\mathbf{a} \,|\, \mathbf{r}) G(\mathbf{r} \,|\, \alpha, \beta) d\mathbf{r} + C \sum_{j=1}^{N} \delta_j \to \min(a_1, ..., a_{N^0}, b, \delta_1, ..., \delta_N), \\ y_j \left(\sum_{k=1}^{N} a_k S(\omega_k, \omega_j) + b \right) \geq 1 - \delta_j,\ \delta_j \geq 0,\ j = 1, ..., N. \end{cases}$$

Такой метод обучения крайне селективен по отношению к базисным объектам – в качестве релевантных остаются лишь очень малое число объектов.

Обучение для заданной функции попарного сравнения с отбором базисных объектов: *Relevance Object Machine*

Michael Tipping, Christopher Bishop: Оставить лишь наиболее информативные (релевантные) базисные объекты $a_k \neq 0$, игнорируя остальные $a_k = 0$.

Искомое решающее правило: $\boxed{d(\omega) = \sum_{k:\ \text{релевантные объекты}} a_k S(\omega_k, \omega) + b \gtrless 0}$

Идея Tipping&Bishop:
Априорное вероятностное представление об искомом векторе коэффициентов $\mathbf{a}$

$\mathcal{N}(\mathbf{a}\,|\,\mathbf{r}) = \prod_{k=1}^{N} \mathcal{N}(a_k\,|\,0, 1/r_k)$
чем меньше r_k, тем ближе априори a_k к нулю

Дополнительное предположение Tipping&Bishop: Меры точности $1/r_k$ также априори независимы, случайны и распределены по гамма закону

$G(\mathbf{r}\,|\,\alpha,\beta) = \prod_{k=1}^{N} \gamma(1/r_k\,|\,\alpha,\beta) = \prod_{k=1}^{N} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1/r_k)^{\alpha-1} \exp\left(-\beta(1/r_k)\right)$

Полное априорное распределение вектора коэффициентов $\mathbf{a} = (a_1, \ldots, a_N)$:

$\Psi(\mathbf{a}\,|\,\alpha,\beta) = \int_{\mathbb{R}^N} \mathcal{N}(\mathbf{a}\,|\,\mathbf{r}) G(\mathbf{r}\,|\,\alpha,\beta) d\mathbf{r}$, $\quad G(\mathbf{r}\,|\,\alpha,\beta)$ близко к равномерному распределению

Регуляризованный критерий обучения
$$\begin{cases} -\ln \int \mathcal{N}(\mathbf{a}\,|\,\mathbf{r}) G(\mathbf{r}\,|\,\alpha,\beta) d\mathbf{r} + C \sum_{j=1}^{N} \delta_j \to \min(a_1, \ldots, a_{N^0}, b, \delta_1, \ldots, \delta_N), \\ y_j \left(\sum_{k=1}^{N} a_k S(\omega_k, \omega_j) + b \right) \geq 1 - \delta_j,\ \delta_j \geq 0,\ j = 1, \ldots, N. \end{cases}$$

Relevance Vector Machine – RVM, поскольку в качестве функции парного сравнения объектов $S(\omega', \omega'')$ Tipping&Bishop использовали кернел, погружающий множество объектов в линейное пространство, где они рассматриваются как векторы.

Relevance Object Machine – выпуклый критерий обучения в линейном пространстве вторичных признаков

Wang, Zhu, Zou: Doubly Regularized SVM – применение идеи Elastic Net для регрессии (Zou, Hastie, позднее Tibshirani) к SVM

$$\begin{cases} \sum_{k=1}^{N} (1-\mu)a_k^2 + \mu|a_k| \; + C\sum_{j=1}^{N}\delta_j \to \min(a_1,...,a_N, b, \delta_1,...,\delta_N), \\ y_j \; \sum_{k=1}^{N} a_k S(\omega_k, \omega_j) + b \;\; \geq 1-\delta_j, \delta_j \geq 0, \; j=1,...,N. \end{cases}$$

$0 \leq \mu \leq 1$ – параметр селективности

Этот критерий является выпуклым в отличие от невыпуклого критерия Tipping&Bishop

Для его оптимизации в двойственной форме удобно использовать итерационный метод внутренней точки. Этот метод делает относительно мало итераций, каждая из которых легко распараллеливается (аспирант МФТИ Николай Разин).

Итоговое решающее правило:

$$d(\omega) = \underbrace{\sum_{\omega_k \in \hat{\Omega}^* \subset \Omega^*}}_{\textit{подмножество релевантных объектов}} \hat{a}_k S(\omega_k, \omega) + \hat{b} \; \begin{matrix} > 0 \\ < 0 \end{matrix}$$

$\Omega^* = \; \omega_j, j=1,...,N$ – вся обучающая совокупность

$S(\omega_j, \omega_l), \; y_j = y(\omega_j), \; j,l = 1,...,N$

Более общая концепция: Несколько разных функций парного сравнения объектов

Несколько функций попарного сравнения объектов $S_i(\omega', \omega''): \Omega \times \Omega \to \mathbb{S} \subseteq \mathbb{R},\ i = 1, ..., n$ (модальности попарного представления объектов).

Обучающая совокупность: $\Omega^* = \{\omega_j,\ j = 1, ..., N\ :\ S_i(\omega_j, \omega_l),\ y_j = y(\omega_j),\ j, l = 1, ..., N;\ i = 1, ..., n\}$.

Каждый объект характеризуется n векторами вторичных признаков относительно обучающей (базисной) совокупности: $\mathbf{x}_i(\omega) = \big(x_{ik}(\omega) = S_i(\omega_k, \omega),\ k = 1, ..., N\big) \in \mathbb{R}^N,\ i = 1, ..., n$.

Теперь отбирать надо не только релевантные объекты в базисной совокупности, но и релевантные модальности.

<u>Принцип тот же:</u> Wang, Zhu, Zou: Doubly Regularized SVM – применение к SVM идеи Elastic Net для регрессии (Zou, Hastie, позднее Tibshirani)

Выпуклый критерий обучения в линейном пространстве вторичных признаков:

$$\begin{cases} \sum_{i=1}^{n} \sum_{k=1}^{N} \Big[(1-\mu) a_{ik}^2 + \mu |a_{ik}| \Big] + C \sum_{j=1}^{N} \delta_j \to \min\ \ (a_{i1}, ..., a_{iN}), i = 1, ..., n, b, \delta_1, ..., \delta_N\ , \\ y_j \Big[\sum_{i=1}^{n} \sum_{k=1}^{N} a_{ik} S_i(\omega_k, \omega) + b \Big] \geq 1 - \delta_j,\ \delta_j \geq 0,\ j = 1, ..., N. \end{cases}$$

В результате обучения часть коэффициентов обращается в строгие нули $a_{ik} = 0$.

Параметр селективности комбинирования $0 \leq \mu \leq 1$.

Этот критерий одновременно отбирает модальности – Modality-Selective SVM, – и релевантные объекты – Relevance Vector (Object) Machine, RVM, – поэтому мы называем его RVM с отбором модальностей.

Более общая концепция: Несколько разных функций парного сравнения объектов

Несколько функций попарного сравнения объектов $S_i(\omega',\omega''):\Omega\times\Omega\to\mathbb{S}\subseteq\mathbb{R},\ i=1,...,n$ (модальности попарного представления объектов).

Обучающая совокупность: $\Omega^*=\ \omega_j,\ j=1,...,N\ :\ \ S_i(\omega_j,\omega_l),\ y_j=y(\omega_j),\ j,l=1,...,N;\ i=1,...,n$.

Каждый объект характеризуется n векторами вторичных признаков относительно обучающей (базисной) совокупности: $\mathbf{x}_i(\omega)=\ x_{ik}(\omega)=S_i(\omega_k,\omega),\ k=1,...,N\ \in\mathbb{R}^N,\ i=1,...,n.$

Теперь отбирать надо не только релевантные объекты в базисной совокупности, но и релевантные модальности.

<u>Принцип тот же:</u> Wang, Zhu, Zou: Doubly Regularized SVM – применение к SVM идеи Elastic Net для регрессии (Zou, Hastie, позднее Tibshirani)

Выпуклый критерий обучения в линейном пространстве вторичных признаков:

$$\begin{cases}\sum_{i=1}^{n}\sum_{k=1}^{N}\Big[(1-\mu)a_{ik}^2+\mu|a_{ik}|\Big]+C\sum_{j=1}^{N}\delta_j\to\min\ (a_{i1},...,a_{iN}),i=1,...,n,b,\delta_1,...,\delta_N\ ,\\ y_j\Big[\sum_{i=1}^{n}\sum_{k=1}^{N}a_{ik}S_i(\omega_k,\omega)+b\Big]\ge 1-\delta_j,\ \delta_j\ge 0,\ j=1,...,N.\end{cases}$$

В результате обучения часть коэффициентов обращается в строгие нули $a_{ik}=0$.

Параметр селективности комбинирования $0\le\mu\le 1$.

Результат обучения – решающе правило, применимое к новым объектам:

$d(\omega)=\sum\sum_{i,k:\,a_{ik}\neq 0}a_{ik}S_i(\omega_k,\omega)+b\gtrless 0$ – релевантные базисные объекты и релевантные модальности.

Типичная зависимость ошибки на контроле и числа релевантных объектов от уровня селективности вторичных признаков

N. Razin, D. Sungurov, V. Mottl, I. Torshin, V. Sulimova, O. Seredin, D. Windridge. Application of the Multi-modal Relevance Vector Machine to the problem of protein secondary structure prediction. Proceedings of the 7th IAPR International Conference on Pattern Recognition in Bioinformatics. November 8-10, 2012, Tokyo, Japan.

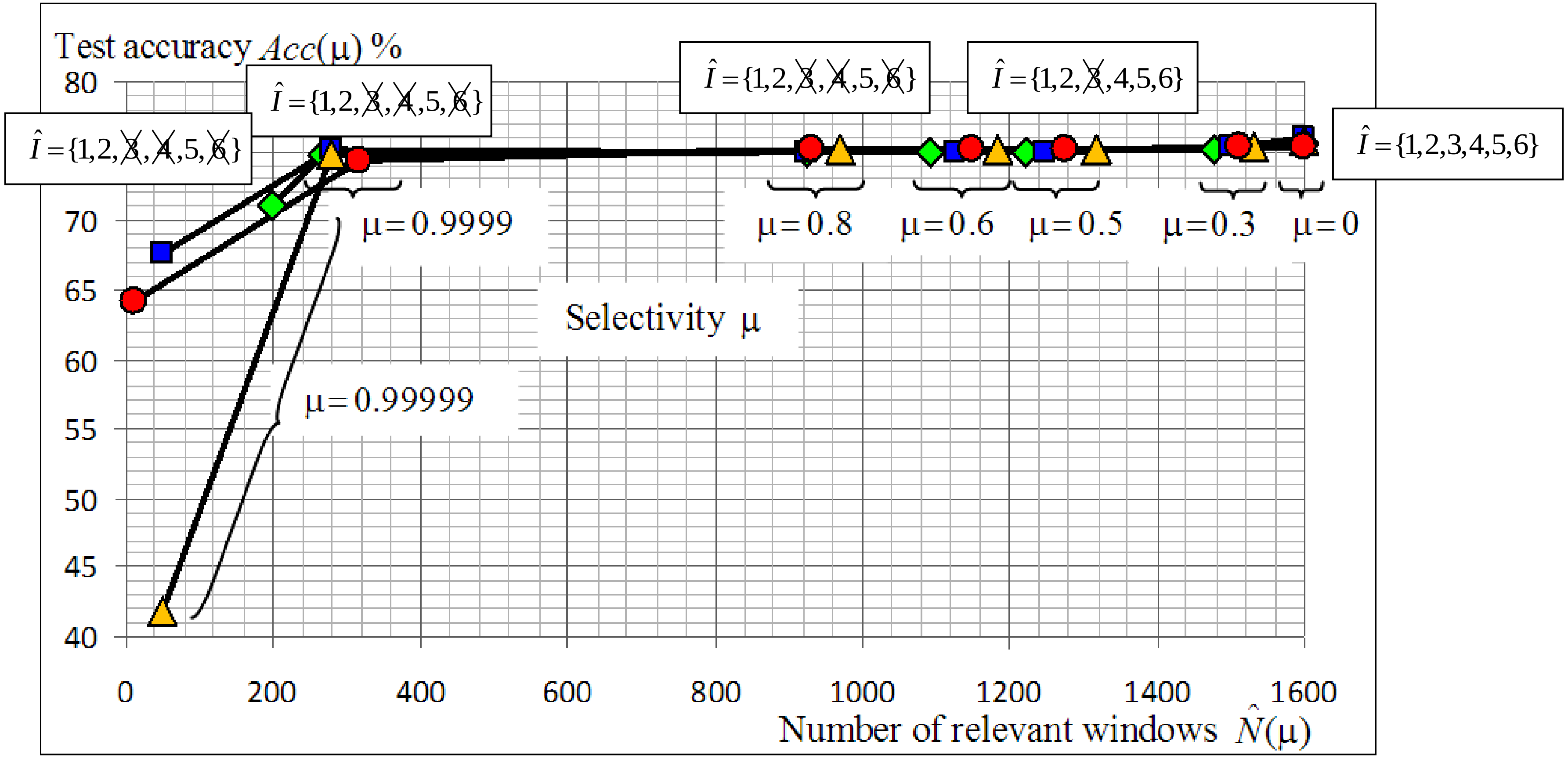


Figure 1. Experimental dependence of the number of relevant amino acid fragments $\hat{N}$ and the test-set accuracy of detecting strands Acc on the level of secondary feature selectivity μ .

Частный случай: Метрика как функция попарного сравнительного представления объектов

Метрика на множестве объектов Ω: $\rho(\omega',\omega'')\geq 0$, $\rho(\omega',\omega''')\leq\rho(\omega',\omega'')+\rho(\omega'',\omega''')$

Функция попарного сравнения: $S(\omega',\omega'')=S(\omega'',\omega')=\rho(\omega',\omega''):\Omega\times\Omega\to\mathbb{R}$

Обучающая совокупность: $\Omega^* = \ \omega_j,\ j=1,...,N \ :\ \rho(\omega_j,\omega_l),\ y_j = y(\omega_j),\ j,l=1,...,N$

Вектор вторичных признаков объекта $\omega\in\Omega$ относительно обучающей совокупности:
$\mathbf{x}(\omega)= \ x_k(\omega)=\rho(\omega_k,\omega),\ k=1,...,N \ \in\mathbb{R}^N$

Классический критерий SVM в линейном пространстве метрических признаков:

$$\begin{cases}\sum_{k=1}^{N} a_k^2 + C\sum_{j=1}^{N}\delta_j \to \min(a_1,...,a_N, b, \delta_1,...,\delta_N),\\ y_j \ \ \sum_{k=1}^{N} a_k\rho(\omega_k,\omega_j)+b \ \ \geq 1-\delta_j,\ \delta_j\geq 0,\ j=1,...,N.\end{cases}$$

Результат обучения: Дискриминантная функция, применимая к произвольному объекту

$$d(\omega)=\sum_{k=1}^{N} a_k\rho(\omega_k,\omega)+b \gtrless 0$$

Принципиально новое обстоятельство – базисная совокупность $\Omega^* = \ \omega_j,\ j=1,...,N$ снабжена метрикой.

Каждый коэффициент a_k соответствует объекту базисной совокупности, причем среди них разные пары по-разному отличаются друг от друга $\rho(\omega_k,\omega_l)$.

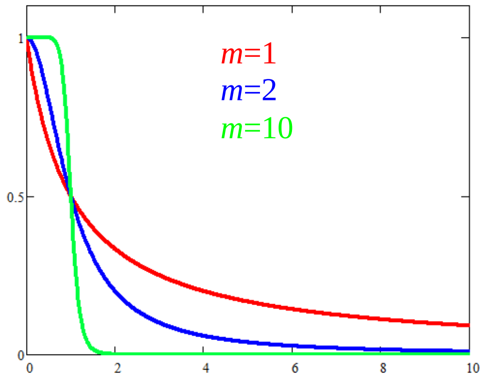
Естественное дополнительное требование:
Если $\rho(\omega_k,\omega_l)\to 0$, то $(a_k-a_l)^2\to 0$.

Регуляризованный критерий обучения в метрическом пространстве объектов

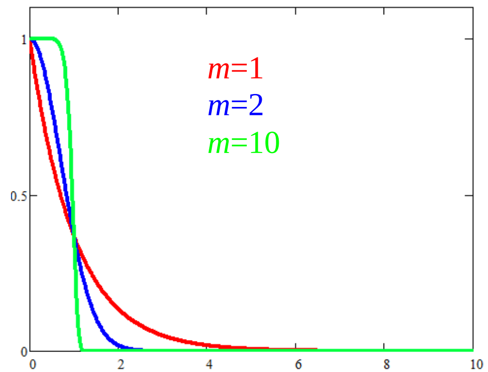
Исходный критерий обучения:	Критерий обучения с ***метрической регуляризацией***:
$\begin{cases} \sum_{k=1}^{N} a_k^2 + C\sum_{j=1}^{N}\delta_j \to \min(a_k, b, \delta_j), \\ y_j\left(\sum_{k=1}^{N} a_k \rho(\omega_k,\omega_j) + b\right) \ge 1-\delta_j, \\ \delta_j \ge 0,\ j=1,\ldots,N. \end{cases}$	$\begin{cases} \sum_{k=1}^{N^0} a_k^2 + \boxed{\sum_{k=1}^{N^0}\sum_{l=1}^{N^0}\eta(\omega_k,\omega_l)(a_k-a_l)^2} + C\sum_{j=1}^{N}\delta_j \to \min(a_k, b, \delta_j), \\ y_j\left(\sum_{k=1}^{N^0} a_k \rho(\omega_k,\omega_j) + b\right) \ge 1-\delta_j, \quad \delta_j \ge 0,\ j=1,\ldots,N. \end{cases}$

Здесь $\eta(\omega_k,\omega_l)\ge 0$ –убывающая функция значения метрики $\rho(\omega_k,\omega_l)$. Например:

$\eta(\omega_k,\omega_l\,|\,\beta,c)=\beta\big/\left[1+c\rho^m(\omega_k,\omega_l)\right]$



$\eta(\omega_k,\omega_l\,|\,\beta,c)=\beta\exp\left(-c\rho^m(\omega_k,\omega_l)\right)$



$\eta(\omega_k,\omega_l)=1/\rho(\omega_k,\omega_l)$