3.3 kernel trick $\quad x_1^2 + 2x_2^2 = 3$

Квадратичное ядро: $k(x,y) = \langle x,y\rangle^2 = (x_1y_1 + x_2y_2)^2 =$

$= x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2 = \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2),$

$(y_1^2, y_2^2, \sqrt{2}\, y_1y_2)$

Отображение: $\psi(x_1, x_2) \to (x_1^2, x_2^2, \sqrt{2}\, x_1 x_2)$

Линейная поверхность в H:

$\langle (x_1^2, x_2^2, \sqrt{2} x_1x_2), (w_1, w_2, w_3)\rangle + w_0 =$

$= w_1 x_1^2 + w_2 x_2^2 + w_3\sqrt{2}\, x_1x_2 =$

При $\cancel{w_1 = 1}$ $\cancel{w_2 = 2}$ $w = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ $w_0 = -3$

$=$ $x_1^2 + 2x_2^2 - 3 = 0$

(3.2) Задача классификации на два непересекающихся класса, в которой объекты описываются n-мерными вещественными векторами: $X = \mathbb{R}^n$, $Y = \{-1, 1\}$.

Линейный пороговый классификатор:

$$a(x) = \mathrm{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right) = \mathrm{sign}(\langle w, x\rangle - w_0)$$

где $(x^1 \dots x^n)$ - признаковое описание объекта x, вектор $w = (w^1 \dots w^n) \in \mathbb{R}^n$ и скалярный порог $w_0 \in \mathbb{R}$ - параметры алгоритма.

Предположим, что выборка линейно разделима: найдутся w, w_0, задающие разделяющую гиперплоскость $\langle w, x\rangle = w_0$, при которых функционал числа ошибок

$$Q(w, w_0) = \sum_{i=1}^{\ell} \left[y_i(\langle w, x_i\rangle - w_0) \leq 0\right] = 0$$

Для удобства домножим w и w_0 на константу, чтобы

$$\min_{i=\overline{1,\ell}} y_i(\langle w, x_i\rangle - w_0) = 1$$

Посмотрим на ширину полосы. На её границе лежат точки $x_+ \in \{1\}$ и $x_- \in \{-1\}$.

Ширина:

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|}\right\rangle = \frac{\langle w, x_+\rangle - \langle w, x_-\rangle}{\|w\|} =$$

$$= \left| y_i(\langle w_i, x_i\rangle - w_0) = 1, i \in \{+, -\}\right| = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

Получается, что ширина полосы максимальна, когда норма вектора w ~~[illegible]~~ минимальна. Задача оптимизации:

$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min\limits_{w, w_0} \\ y_i(\langle w, x_i\rangle - w_0) \geq 1, \quad i = 1 \dots \ell \end{cases}$$

Пусть теперь выборка неразделима. $y_i(\langle w, x_i\rangle - w_0)$ не обязательно будет не меньше 1. Ослабим ограничение, введем минимизируемый ~~функционал~~ штраф за суммарную ошибку:

$$\begin{cases} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \to \min\limits_{w, w_0, \xi} \\ M_i(w, w_0) = y_i(\langle w, x_i\rangle - w_0) \geq 1 - \xi_i, \; i = 1 \ldots l \\ \xi_i \geq 0, \; i = 1 \ldots l \end{cases}$$

ξ_i в этих условиях будет показывать величину ошибки на x_i объекте. Преобразуем условия на ξ_i:

$$\begin{cases} \xi_i \geq 1 - M_i(w, w_0) \\ \xi_i \geq 0 \end{cases}$$

Значит, минимум ξ_i будет при $\xi_i = (1 - M_i(w, w_0))_+$

Эквивалентная задача безусловной оптимизации:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(1 - M_i(w, w_0))_+ \to \min\limits_{w, w_0}$$

(3.5) l_1-регуляризация

С помощью теоремы Куна-Таккера докажем, что

$\begin{cases} \|\langle w, x\rangle - y\|^2 \to \min \\ \sum|w_i| - a \le 0 \end{cases}$ ~~эквивалентно~~ и $\left(\|\langle w, x\rangle - y\|^2 + \lambda(\sum|w_i| - a)\right) \to \min$

приводят к построению одного и того же алгоритма для задачи линейной регрессии

По теореме Куна-Таккера:

Если $\exists$ решение $w = (\hat{w}_1 \dots \hat{w}_n)$ для задачи с ограничениями:

$$\begin{cases} \|\langle w, x\rangle - y\|^2 \to \min \\ \sum\limits_{i=1}^{n}|\hat{w}_i| - a \le 0 \end{cases}$$

то найдется вектор $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$, такой что выполнены условия:

~~$\min \|\langle w, x\rangle - y\|^2 = \|\langle \hat{w}, x\rangle - y\|^2$~~

1) $\min\left(\|\langle w, x\rangle - y\|^2 + \lambda(\sum|w_i| - a)\right) =$

$= \|\langle \hat{w}, x\rangle - y\|^2 + \lambda(\sum|\hat{w}_i| - a)$

2) $\lambda(\sum|w_i| - a) = 0$

3) $\lambda_i \ge 0$

Следовательно, ~~мы можем сказать~~ две задачи эквивалентны и приводят к одному и тому же решению