SVM

Пусть $X = \mathbb{R}$ - пространство объектов, $Y = \{-1, 1\}$ - классы, $X^l = \{(x_i, y_i)\}_{i=1}^l$. Требуется построить классификатор $a : X \to Y$.

Линейный пороговый классификатор:

$$a(x) = sign(\langle w, x \rangle - w_0)$$

Уравнение $\langle w, x \rangle = w_0$ описывает гиперплоскость, разделяющую классы. Функционал ошибок:

$$Q(w, w_0) = \sum_{i=1}^{n} \Big(y_i(\langle w, x_i \rangle - w_0) < 0 \Big)$$

Нормировка:
Пусть x_k - наиболее близкие к гиперплоскости из набора $\{x_i\}$. Тогда возьмем такие w, w_0, что $\langle w, x_k \rangle - w_0 = y_k$. Тогда для всех x_i:

$$\langle w, x_k \rangle - w_0 \begin{cases} \leq -1, & y_i = -1 \\ \geq 1, & y_i = 1 \end{cases} \Leftrightarrow m_i = y_i(\langle w, x_k \rangle - w_0) \geq 1$$

Необходимо найти такие параметры, при которых ширина разделяющей полосы максимальна. Пусть x_1 - ближайший к гиперплоскости элемент класса $\{1\}$, x_{-1} - класса $\{-1\}$. Тогда ширина полосы:

$$\langle (x_1 - x_{-1}), \frac{w}{||w||} \rangle = \frac{(w_0 + 1) - (w_0 - 1)}{||w||} = \frac{2}{||w||} \to sup$$

Необходимо решить следующую экстремальную задачу:

$$\begin{cases} ||w|| \to inf \\ y_i(\langle w, x_i \rangle - w_0) \geq 1, \quad i \in 1 : l \end{cases}$$

В результате получается, что w может быть представлена как линейная комбинация x_i, лежащих на границе разделяющей полосы.

Если выборка линейно неразделима:

$$\begin{cases} \frac{1}{2}||w|| + C \sum_{i=1}^{l} \varepsilon_i \to inf \\ y_i(\langle w, x_i \rangle - w_0) \geq 1 - \varepsilon_i, \quad i \in 1 : l \\ \varepsilon_i \geq 0 \end{cases}$$

$$Q = \sum_{i=1}^{l} \Big[m_i < 0 \Big]$$

Замеим пороговую функцию на ее верхней оценкой $\Big[m_i < 0 \Big] \leq (1 - m_i)_+$ и добавим слагаемое регуляризации:

$$Q = \sum_{i=1}^{l} (1 - m_i)_+ + \tau ||w||^2 \to inf$$

В результате снова получается, что w можнт быть представлена в виде линейной комбинации векторов на границе разделяющей полосы.

$$\begin{cases} -\sum_{i=1}^{l}\lambda_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j\langle x_i, x_j\rangle \to inf \\ 0 \le \lambda_i \le C \\ \sum_{i=1}^{l}\lambda_i y_i = 0 \end{cases}$$

Ядро

Ядром называется функция $K(x, x')$, если она симметрична и ноетрицательно определена.

Во всех приведенных выражениях скалярное произведение можно заменить ядром K.