

Дано:

Обучающая выборка $X^\ell = (x_i, y_i)_{i=1}^\ell$,

x_i — объекты, векторы из множества $X = \mathbb{R}^n$,

y_i — метки классов, элементы множества $Y = \{-1, +1\}$.

Найти:

Параметры $w \in \mathbb{R}^n$, $w_0 \in \mathbb{R}$ линейной модели классификации

$$a(x; w, w_0) = \text{sign}(\langle x, w \rangle - w_0).$$

Критерий — минимизация эмпирического риска:

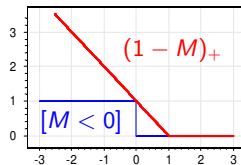
$$\sum_{i=1}^{\ell} [a(x_i; w, w_0) \neq y_i] = \sum_{i=1}^{\ell} [M_i(w, w_0) < 0] \rightarrow \min_{w, w_0}.$$

где $M_i(w, w_0) = (\langle x_i, w \rangle - w_0) y_i$ — отступ (margin) объекта x_i ,
 $b(x) = \langle x, w \rangle - w_0$ — дискриминантная функция.

Эмпирический риск — это кусочно-постоянная функция.
Заменим его оценкой сверху, непрерывной по параметрам:

$$\begin{aligned} Q(w, w_0) &= \sum_{i=1}^{\ell} [M_i(w, w_0) < 0] \leq \\ &\leq \sum_{i=1}^{\ell} (1 - M_i(w, w_0))_+ + \frac{1}{2C} \|w\|^2 \rightarrow \min_{w, w_0}. \end{aligned}$$

- Аппроксимация штрафует объекты за приближение к границе классов, увеличивая зазор между классами
- Регуляризация штрафует неустойчивые решения в случае мультиколлинеарности



Линейный классификатор:

$$a(x, w) = \text{sign}(\langle w, x \rangle - w_0), \quad w, x \in \mathbb{R}^n, \quad w_0 \in \mathbb{R}.$$

Пусть выборка $X^\ell = (x_i, y_i)_{i=1}^\ell$ линейно разделима:

$$\exists w, w_0 : \quad M_i(w, w_0) = y_i(\langle w, x_i \rangle - w_0) > 0, \quad i = 1, \dots, \ell$$

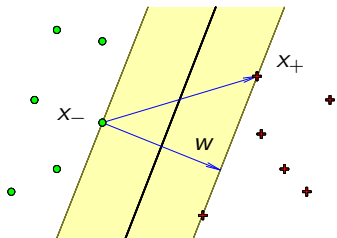
Нормировка: $\min_{i=1, \dots, \ell} M_i(w, w_0) = 1.$

Разделяющая полоса:

$$\{x : -1 \leq \langle w, x \rangle - w_0 \leq 1\}.$$

Ширина полосы:

$$\frac{\langle x_+ - x_-, w \rangle}{\|w\|} = \frac{2}{\|w\|} \rightarrow \max.$$



Постановка задачи в случае линейно разделимой выборки:

$$\begin{cases} \frac{1}{2} \|w\|^2 \rightarrow \min_{w, w_0}; \\ M_i(w, w_0) \geq 1, \quad i = 1, \dots, \ell. \end{cases}$$

Общий случай — линейно неразделимая выборка:

$$\begin{cases} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{w, w_0, \xi}; \\ M_i(w, w_0) \geq 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geq 0, \quad i = 1, \dots, \ell. \end{cases}$$

Исключая ξ_i , получаем задачу безусловной минимизации:

$$C \sum_{i=1}^{\ell} (1 - M_i(w, w_0))_+ + \frac{1}{2} \|w\|^2 \rightarrow \min_{w, w_0}.$$

Задача математического программирования:

$$\begin{cases} f(x) \rightarrow \min_x; \\ g_i(x) \leq 0, & i = 1, \dots, m; \\ h_j(x) = 0, & j = 1, \dots, k. \end{cases}$$

Необходимые условия. Если x — точка локального минимума, то существуют множители $\mu_i, i = 1, \dots, m, \lambda_j, j = 1, \dots, k$:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x} = 0, & \mathcal{L}(x; \mu, \lambda) = f(x) + \sum_{i=1}^m \mu_i g_i(x) + \sum_{j=1}^k \lambda_j h_j(x); \\ g_i(x) \leq 0; & h_j(x) = 0; \text{ (исходные ограничения)} \\ \mu_i \geq 0; & \text{ (двойственные ограничения)} \\ \mu_i g_i(x) = 0; & \text{ (условие дополняющей нежёсткости)} \end{cases}$$

Функция Лагранжа:

$$\begin{aligned}\mathcal{L}(w, w_0, \xi; \lambda, \eta) = \\ = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1) - \sum_{i=1}^{\ell} \xi_i (\lambda_i + \eta_i - C),\end{aligned}$$

λ_i — переменные, двойственные к ограничениям $M_i \geq 1 - \xi_i$;

η_i — переменные, двойственные к ограничениям $\xi_i \geq 0$.

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 0, & \frac{\partial \mathcal{L}}{\partial w_0} = 0, & \frac{\partial \mathcal{L}}{\partial \xi} = 0; \\ \xi_i \geq 0, & \lambda_i \geq 0, & \eta_i \geq 0, \quad i = 1, \dots, \ell; \\ \lambda_i = 0 \text{ либо } M_i(w, w_0) = 1 - \xi_i, & i = 1, \dots, \ell; \\ \eta_i = 0 \text{ либо } \xi_i = 0, & i = 1, \dots, \ell; \end{cases}$$

$$\begin{cases} -\sum_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \rightarrow \min_{\lambda}; \\ 0 \leq \lambda_i \leq C, \quad i = 1, \dots, \ell; \\ \sum_{i=1}^{\ell} \lambda_i y_i = 0. \end{cases}$$

Решив эту задачу численно относительно λ_i ,
получаем линейный классификатор:

$$a(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^{\ell} \lambda_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle - w_0 \right),$$

где $w_0 = \sum_{i=1}^{\ell} \lambda_i y_i \langle \mathbf{x}_i, \mathbf{x}_j \rangle - y_j$ для такого j , что $\lambda_j > 0$, $M_j = 1$

Определение

Объект \mathbf{x}_i называется *опорным*, если $\lambda_i \neq 0$.

Преимущества:

- Задача выпуклого квадратичного программирования имеет единственное решение.
- Выделяется множество опорных объектов.
- Имеются эффективные численные методы для SVM.
- Изящное обобщение на нелинейные классификаторы.

Недостатки:

- Опорными объектами могут становиться выбросы.
- Нет отбора признаков в исходном пространстве X .
- Приходится подбирать константу C .