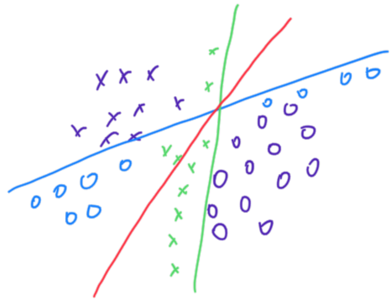
Метод опорных векторов

Support vector machine (SVM)

$$a(x) = \operatorname{sign}(\langle w, x\rangle + b)$$

① Линейно разделимый случай



$$\operatorname{sign}\Big(\underbrace{\langle \tfrac{w}{\alpha}, x\rangle + \tfrac{b}{\alpha}}\Big)$$

$\alpha \in \mathbb{R}$

$\alpha > 0$

$$\boxed{\min_{x \in X} |\langle w, x\rangle + b| = 1}$$

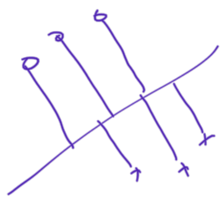
← новое требование к модели (*)

$$\rho(x_0, a) = \frac{|\langle w, x\rangle + b|}{\|w\|_2}$$

от x_0 до разд. гиперплоск., соотв. a

$$\min_{x \in X} \frac{|\langle w, x\rangle + b|}{\|w\|_2} = \frac{1}{\|w\|_2}$$

мин. расстояние от разд. гиперплоск. до объектов обуч. выборки



$\frac{2}{\|w\|_2}$ – ширина разделяющей полосы (margin – отступ классификатора)

$$\frac{1}{\|w\|_2} \to \max_w$$

$\|w\|_2 \to \min_w$ — нужны ещё условия

$y_i \in \{-1, +1\}$

$$\begin{cases} \|w\|_2^2 \to \min_w \\ y_i(\langle w, x_i\rangle + b) \geq 0 \\ |\langle w, x_i\rangle + b| \geq 1 \iff (*) \end{cases}$$

← требование корректной класиф.

$\exists\, w_*$ — решение и $\underline{\forall x_i \in X \quad |\langle w, x_i\rangle + b| > 1}$

тогда $\tilde{w}_* = \frac{w_*}{\alpha}$, $\tilde{b}_* = \frac{b_*}{\alpha}$ и $\|\tilde{w}_*\| < \|w_*\|$ $\qquad \alpha > 0$

$$\begin{cases} \|w\|_2^2 \to \min_w \\ y_i(\langle w, x_i\rangle + b) \geq 1, \quad i = 1, \dots, l \end{cases}$$

— SVM для лин. разд. случая

---

② Линейно неразделимый случай

$$\begin{cases} \|w\|_2^2 + C\frac{1}{l}\sum_{i=1}^{l} \xi_i \to \min_{w, b, \xi_i} \\ y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \\ \xi_i \geq 0 \end{cases}$$

← slack variables

— SVM в общем случае

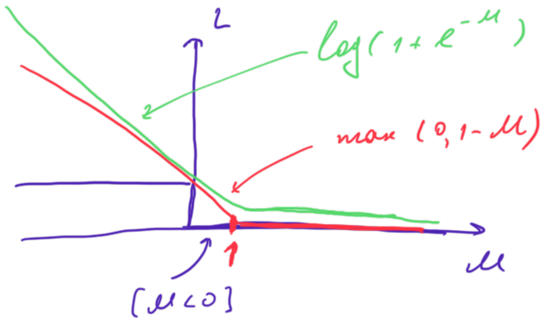
C - гиперпарам.

$$\begin{cases} \xi_i \geq \underline{1 - y_i(\langle w, x_i\rangle + b)} \\ \xi_i \geq 0 \end{cases} \Rightarrow \xi_i = \max(0, 1 - y_i(\langle w, x_i\rangle + b))$$

$$\|w\|_2^2 + C \cdot \frac{1}{l}\sum_{i=1}^{l} \max(0, 1 - \underbrace{y_i(\langle w, x_i\rangle + b)}_{M_i}) \to \min_{w, b}$$

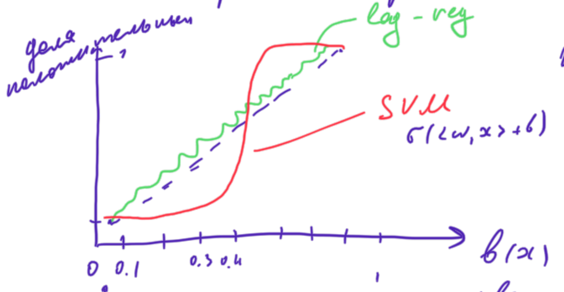
$$\underbrace{\frac{1}{l}\sum_{i=1}^{l} \max(0,\ 1-M_i)}_{\text{ошибка}} + \underbrace{\frac{1}{C}\|w\|_2^2}_{\text{регуляриз.}} \to \min_{w,b}$$

$$[M<0] \le \max(0,\ 1-M)$$



логрег	SVM
$M_i \to \max$	$M_i \ge 1$
лучше для log-loss, оценивания вер-тей	лучше для AUC-ROC, F-меры

Калибровка вероятностей



калибровочная кривая

(вер-ть из модели)

все объекты из X

$0 \le b(x) < 0.1$

можно калибровать - преобразовывать $b(x)$ так, чтобы кривая была близка к идеал.

Многоклассовая классификация

$\mathbb{Y} = \{1, \dots K\}$

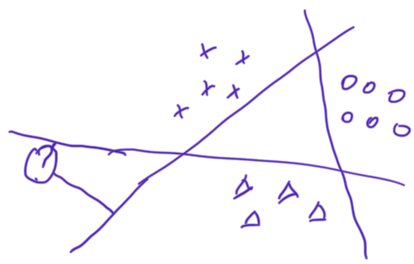
① Один против всех (one-vs-all)
one-vs-rest

$X_k = \{(x_i,\ 2[y_i = k] - 1)\}_{i=1}^{l}$

$a_k(x) = \operatorname{sign} b_k(x)$ - обуч. на X_k

$b_k(x) = \langle w_k, x\rangle + w_{0k}$

$a(x) = \arg\max_{k \in \{1,\dots K\}} b_k(x)$



как получить вероятности:

$$p_k(x) = \frac{\exp(b_k(x))}{\sum_{j=1}^{K} \exp(b_j(x))} \quad \text{— softmax}$$

② All-vs-all

one-vs-one

C_K^2 классиф.

a_{ij} — относит к i или к j

$$a(x) = \arg\max_{k \in \{1, \dots K\}} \sum_{i=1}^{K} \sum_{j \neq i} [a_{ij}(x) = k]$$

③ Многоклассовая логистич. регрессия

$$\sum_{i=1}^{\ell} \sum_{k=1}^{K} [y_i = k] \log p_k(x_i) \to \max_{\{w_k, w_{0k}\}}$$

Метрики качества в многокласс. классиф.

$$\text{accuracy} = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i] \quad \text{— без изменений}$$

задача: $y_i^k = 2[y_i = k] - 1$
$a^k(x) = [a(x) = k]$ $\to TP_k,\ FP_k,\ FN_k,\ TN_k$

микро-усреднение

$$\overline{TP} = \frac{1}{K} \sum_{k=1}^{K} TP_k$$

$\overline{FP}, \dots$

$$\text{precision} = \frac{\overline{TP}}{\overline{TP} + \overline{FP}}$$

если класс маленький,

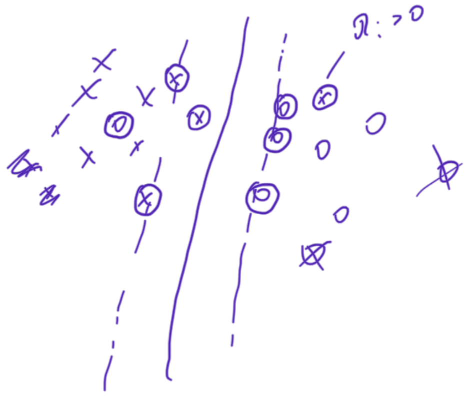
макро-усреднение

$$\text{precision}_k = \frac{TP_k}{TP_k + FP_k}$$

$$\text{precision} = \frac{1}{K} \sum_{k=1}^{K} \text{precision}_k$$

все классы вносят

то вклад тоже меньший / равный нулю



$$W = \sum_{i=1}^{l} \lambda_i y_i x_i$$

$$\lambda_i \geq 0$$