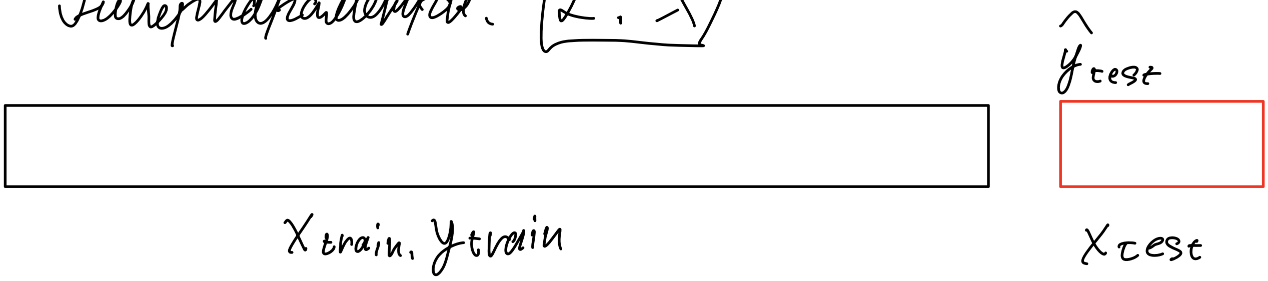
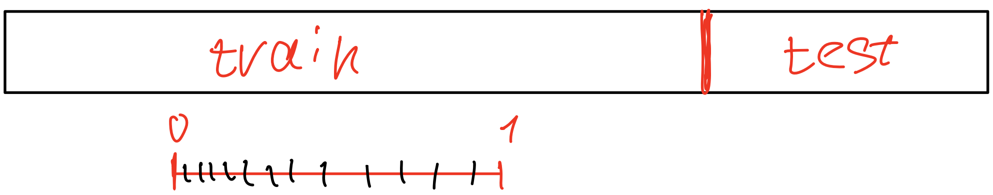
Кросс-валидация.

Параметры: w,

Гиперпараметры: $\boxed{L, \lambda}$



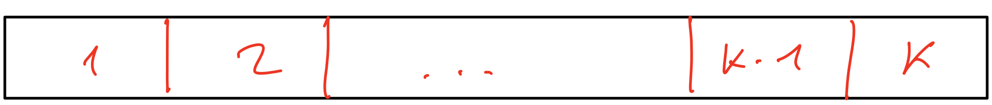
1) Отложенная выборка



Для каждого набора оцениваем модель на трейне, считаем метрику.

2) Кросс-валидация

а) K-fold



Для каждого набора параметров:

for k in K:

k-й фолд – тест

всё остальное – трейн

считаем метрики на тесте

усредняем по фолдам

2) LOOCV - Leave-one-out cross validation

1) Взять лучшую модель
2) Взять p лучших моделей

$$R^2$$

Коэфф. детерминации

$$y_i = w_0 + w_1 x_i \qquad \hat{y}_i = \hat{w}_0 + \hat{w}_1 x_i$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \underbrace{\frac{1}{n}\sum(\hat{y}_i - \bar{\hat{y}}_i)^2}_{} + \frac{1}{n}\sum(y_i - \hat{y}_i)^2$$

$$TSS = ESS + RSS \to \min$$

Total Estimated Error Residual Regression)

$$R^2 = \frac{ESS}{TSS} \to \max$$

$$R^2 = \mathrm{corr}^2(X, y)$$

1) В регрессии без константы ($w_0 = 0$)
R^2 неприменим. $R^2_{extended}$

г) R^2 не убывает при добавлении новых признаков

$$R^2_{adj} = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} =$$

$$= 1 - (1 - R^2) \cdot \frac{(n-1)}{n-k} \qquad R^2_{adj} \le R^2$$

Квантильная регрессия

$a(x)$ – лин. регрессия

$$Q(a, X) = \sum_{i=1}^{n} \rho_\tau (y_i - a(x_i))$$

$$\rho_\tau(z) = \underbrace{(\tau - 1)}_{} \underbrace{[z<0]}_{1} \underbrace{z}_{z} + \underbrace{\tau}_{} \underbrace{[z \geq 0]}_{0} \underbrace{z}_{z}$$

$$I\{cond\} = [cond] = \begin{cases} 1, & cond \\ 0, & !cond \end{cases}$$

$$a(x_i) = w_0 + w_1 x_i + \varepsilon_i \qquad E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$$

ε_i – iid

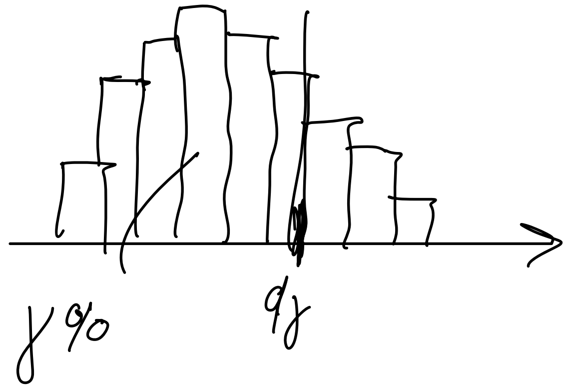
MSE: $a(x) \approx E(y|x)$

MAE: $a(x) \approx median[P(y|x)]$

$a(x) \approx$ квантиль порядка τ

q_γ – квантиль порядка γ

$p(y|x \le q_\gamma) = \gamma$



$\gamma = 0{,}5$

Классификация

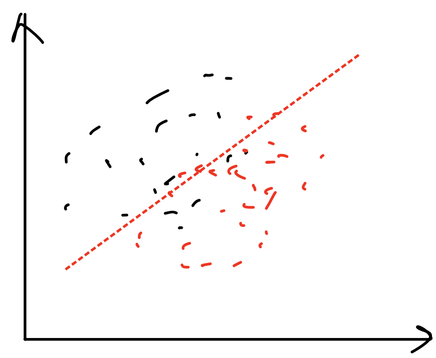
X - такие же

$y \in \{-1, 1\}$ / $y \in \{0, 1\}$

$P(y_i = 1)$

$\langle w, x_i \rangle = 0$

$\text{sign}\langle w, x_i \rangle$



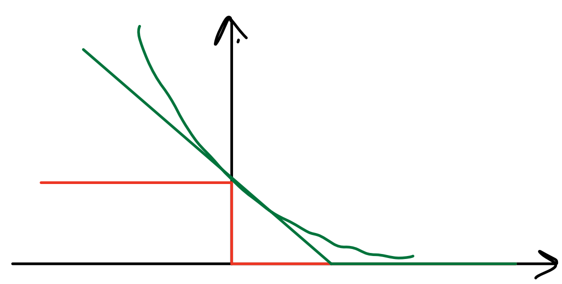
$\sum_{i=1}^{n} I\{y_i \ne \text{sign}\langle w, x_i \rangle\}$

misclassification loss

$\sum_{i=1}^{n} I\{\underbrace{y_i \langle w, x_i \rangle}_{} < 0\} \to \min$

Отступ

margin M



Отступ положителен, если класс верный

$$F(M) = I\{M < 0\} = \begin{cases} 1, & M < 0 \\ 0, & M \geq 0 \end{cases}$$

$$L(M) = \sum_{i=1}^{n} I\{M_i < 0\}$$

$$L(M) \leq \tilde{L}(M)$$

$$P(y|x) = \frac{1}{2}$$

1) $\tilde{L}(M) = \log(1 + e^{-M})$ – логистическая

2) $\tilde{L}(M) = \max(0, 1 - M)$ – (В SVM)

3) $\tilde{L}(M) = \max(0, -M)$ персептрон

4) $\tilde{L}(M) = e^{-M}$ – экс

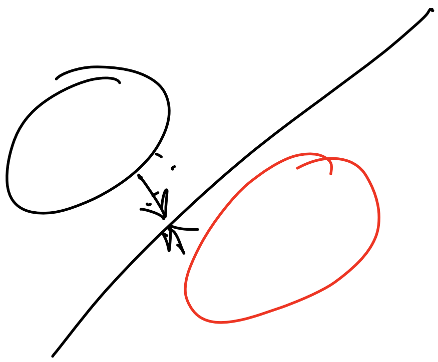
5) $\tilde{L}(M) = 2/(1 + e^{M})$ – сигмоида

Персептрон

$$\tilde{L}(M) = \max(0, -M)$$

$$\tilde{L}(M) = \lambda \|w\|_2^2 + \sum_{i=1}^{n} \max(0, \underbrace{-y_i \langle w, x_i \rangle}_{M})$$

$$\nabla_w L(w, x, y) = 2\lambda w + \sum_{i=1}^{n} \begin{cases} 0, & M > 0 \\ -y_i x_i & M \leq 0 \end{cases}$$



SVM $\tilde{L} = \max(0, 1 - M)$

$a(x) = \text{sign}(\langle w, x\rangle + b)$ $\quad w \in R^d, b \in R$

1) Линейно разделимый случай

1) Умножим w и b на одинаковую константу. c

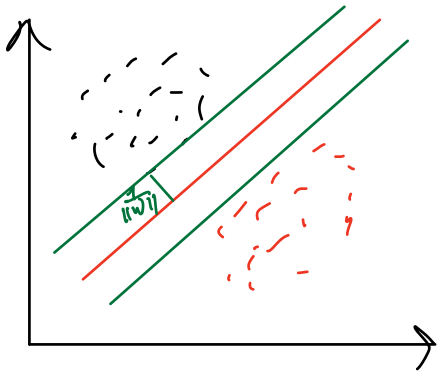
$a(x) = \text{sign}(c(\langle w, x\rangle + b))$

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

Расстояние до ближайшего объекта

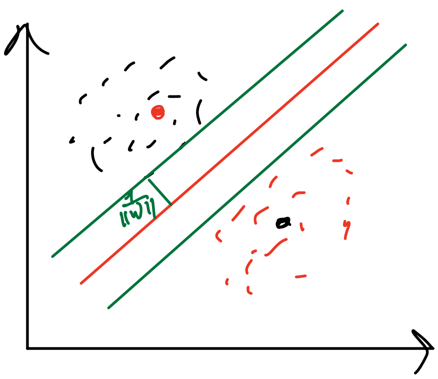
$$\min_{x \in X} \frac{|\langle w, x\rangle + b|}{\|w\|} = \frac{1}{\|w\|} \underbrace{\min_{x \in X} |\langle w, x\rangle + b|}_{1} = \frac{1}{\|w\|}$$

Отступ (margin)



$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min \\ y_i\left(\langle w, x\rangle + b\right) \geq 1 \end{cases}$$

Неразделимый случай



$\exists x_i \in X: \; y_i(\langle w, x_i\rangle + b) < 1$

$\exists \xi \geq 0:$

$y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i$

$\frac{1}{\|w\|}$ – мягкий отступ (soft margin)

$$\left.\begin{cases} \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{n}\xi_i \to \min\limits_{w,b,\xi} \\ y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}\right\} \max$$

Сведение к безусловной

$$\frac{1}{2}\|w\|^2 + C\boxed{\sum \max\left(0, 1 - y_i(\langle w, x_i\rangle + b)\right)} \to \min_{w,b}$$

Вероятности.

$\forall x \in X : \quad p(y=1 \mid x)$

$b(x) : X \to [0,1] \qquad b(x_i) \approx p(y_i = 1 \mid x_i)$

$X, y \qquad y = (0, 0, 1, 0, 1, 1, 0, \ldots)$

$$\underbrace{\prod_{i=1}^{n} \underbrace{b(x_i)^{[y_i=1]}} \underbrace{(1-b(x_i))^{[y_i=0]}}} \to \max \quad // \log$$

$$-\sum_{i=1}^{n} [y_i = 1] \ln b(x_i) + [y_i = 0] \ln(1 - b(x_i)) \to \min$$

log-loss

$$E\big[-[y_i=1] \ln b(x_i) - [y_i=0] \ln(1-b(x_i))\big]$$

$$y_i = \begin{cases} 1 & p(y_i \mid x_i) \\ 0 & 1 - p(y_i \mid x_i) \end{cases}$$

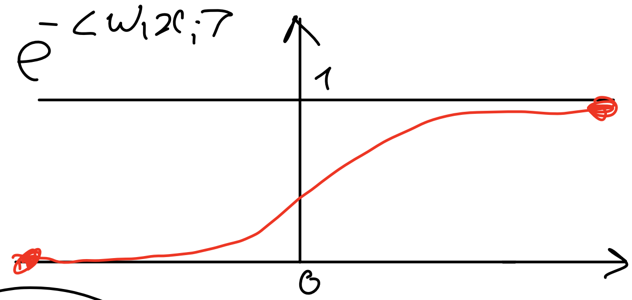
$E(y_i) = p(y_i \mid x_i)$

$$= \underline{-p \ln b(x_i) - (1-p) \ln(1 - b(x_i))}$$

$b^*(x_i) = p$

Логистическая регрессия

$$\sigma(\langle w, x_i \rangle) = \frac{1}{1+e^{-\langle w, x_i \rangle}}$$



$$p(y=1|x) = \frac{1}{1+e^{-\langle w, x \rangle}} = b(x).$$

$$\langle w, x \rangle = \log \frac{p(y=1|x)}{p(y=0|x)}$$

$$= \sum_{i=1}^{n} \log\left(1+e^{-y_i \langle w, x_i \rangle}\right)$$

Метрики

$$\text{Accuracy}(a(x_i), x) = \frac{1}{n} \sum I\{a(x_i) = y_i\}$$

	$y_i=1$	$y_i=-1$
$a(x_i)=1$	TP	FP
$a(x_i)=-1$	FN	TN

Multi-label классификация

Микро-макро усреднение

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Recall} = \frac{TP}{TP+FN}$$

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

ROC-AUC / PR-AUC