27 Задача классификации, выборка размером N

$x_j^1 \leq x_j^2 \dots \leq x_j^N$.

Есть два объекта, x^c и x^{c+1}, разбиение по которым дает ~~максимальное~~ минимальное значение загрязнености.

Под разбиением по объектам я понимаю что все объекты $x^i \leq x^c$ идут в R_{m1}, остальные (т.е. те, которые $x^i \geq x^{c+1}$) идут в R_{m2}.

Доказать $y^c \neq y^{c+1}$.

Докажем от противного. Пусть $y^c = y^{c+1}$, тогда можно подобрать пример, где разбиение x^c, x^{c+1} не будет давать минимального значения загрязнености.

Например:

	x^1_1	x^2_2	x^3	x^4	x^5
y^i:	0	0	1	1	1

разбиваем по x^3, x^4

При разбиении $\hat{Q}_m = \frac{3}{5} \cdot \left(1 - \left(\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right)\right) + \frac{2}{5} \cdot (1-1) \approx 0{,}26$

При разбиении x^2, x^3 $\hat{Q}_m = 0$.

$\boxed{2}$ Как использовать для нахождения разбиения?

Можно разбивать по „границе классов".
т.е искать такие x^i и x^j, стоящие рядом
и разбивать по ним.

$\boxed{29}$. $\exists N$ размер выборки

Каждый объект может попасть в выборку равновероятно. $\Rightarrow$ $\exists A_i$ - объект i попал в выборку

$$P(A_i) = \frac{1}{N} \qquad P(\overline{A}_i) = 1 - \frac{1}{N}$$

Каждый объект выбирается N раз. $\Rightarrow$

$$\overline{A} = \bigcap_i \overline{A}_i$$ - объект не попал в выборку N раз.

$$P(\overline{A}) = \left(1 - \frac{1}{N}\right)^N \xrightarrow[N \to \infty]{} \frac{1}{e} \approx 0,37$$

$\Rightarrow$ 37 % объектов останутся вне bootstrap выборки.

$\Rightarrow$ out of bag ошибка - это ошибка на тех 37% данных, на которой классификаторы не обучались

$\boxed{22}$

$$j(\beta_0, \beta) = \sum_{i=1}^{N} (1-y_i)\overbrace{\ln\left(1+e^{\beta_0+\beta^T x^i}\right)}^{\to 0 \text{ при } \beta \to \infty} + y_i \cdot \underbrace{\ln\left(1+e^{-\beta_0-\beta^T x^i}\right)}_{\to 0 \text{ при } \beta \to -\infty}$$

в данном случае разделяющая плоскость будет являться минимумом функции:

$$\omega = \beta.$$

$\boxed{21}$

$$J(\beta) = C\sum_{i=1}^{N} (1-y_i)\ln\left(1+e^{\beta^T x^i}\right) + y^{(i)} \cdot \ln\left(1+e^{-\beta^T x^i}\right) + \frac{1-\beta}{2}\|\beta\|_2^2 + \rho\|\beta\|$$

2 3 4

1. Константы опускаем, тк они положительны и на выпуклость не влияют.

2. Для выпуклости для суммы достаточно рассмотреть элемент суммы, тк сумма выпуклых ф-й = вып сумма

$f = \ln\left(1+e^{\beta^T x^i}\right)$; $\frac{\partial f}{\partial \beta} = \frac{x^i \cdot e^{\beta^T x^i}}{1+e^{\beta^T x^i}}$

$$\frac{\partial f}{\partial^2 \beta} = \frac{x^i\left(x_i e^{\beta^T x^i} \cdot \left(1+e^{\beta^T x^i}\right) - e^{\beta^T x^i} \cdot x_i\right)}{\left(1+e^{\beta^T x^i}\right)^2} = \frac{(x^i)^2 \cdot e^{\beta^T x^i}}{\left(1+e^{\beta^T x^i}\right)^2} \geq 0 \quad \forall \beta, x_i$$

③ $\|\beta\|_2^2 = \sum_i \beta_i^2$

$t = \sum t_i \Rightarrow t_i = \beta_i^2$

$\frac{d t_i}{d^2 \beta} = 2 > 0$ Выпукло!

④ $\|\beta\|_1 = \sum_i |\beta_i|$; $t_i = |\beta_i|$, $t = \sum t_i$

Доказательство по определению:

$f(\alpha x_i + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$

$\alpha \in [0, 1]$ $\forall x_1, x_2$

$|\alpha x_1 + (1-\alpha) x_2| \leq \alpha f(x_1) + (1-\alpha) f(x_2)$

$|\underbrace{(\alpha x_1)}_{A} + \underbrace{((1-\alpha) x_2)}_{B}| \leq |\underbrace{\alpha x_1}_{A}| + |\underbrace{(1-\alpha) x_2}_{B}|$

$|A + B| \leq |A| + |B|$

истинно по неравенству треугольника.

⑤ $J(\beta)$ выпукла как сумма выпуклых функций.