Что прогнозируем?

и

$X \quad y \quad a(x)$ $\qquad \underbrace{\frac{1}{n}\sum_{i=1}^{n} L(y_i, a(x_i))}_{\text{среднее}}$

ЗБЧ $\quad \bar{X} \to \mathbb{E}(X)$

$$\mathbb{E}\left[L(y, a(x)) \mid x\right] \longrightarrow \min_{w}$$

$\boxed{№1}$ MSE $\qquad p(y \mid x)$

$$L(y, a(x)) = (y - a)^2$$

$$\mathbb{E}\left[L(y, a) \mid x\right] = \int_{-\infty}^{+\infty} (y - \underline{a})^2 \cdot p(y \mid x)\, dy \longrightarrow \min_{a}$$

$$\frac{\partial}{\partial a}\left(\int_{-\infty}^{+\infty} (y - \underline{a})^2 \cdot p(y \mid x)\, dy\right) =$$

$$\int_{-\infty}^{+\infty} \cancel{-2} \cdot (y - a) \cdot p(y \mid x)\, dy = 0$$

$$\underbrace{\int y \cdot p(y \mid x)\, dy}_{\mathbb{E}(y \mid x)} - \int a\, p(y \mid x)\, dy = 0$$

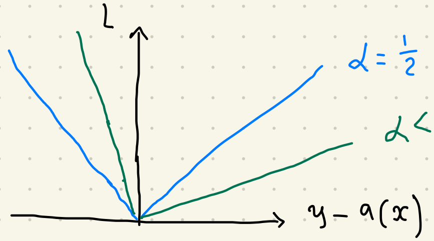
$$\mathbb{E}(y \mid x) - a = 0$$

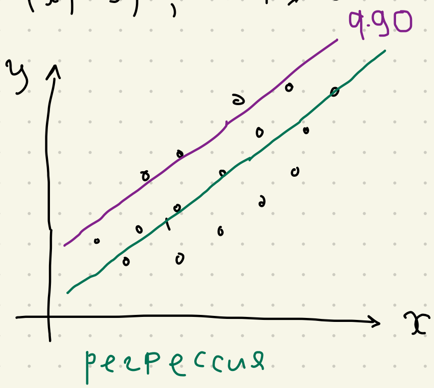
$$a = \mathbb{E}[y \mid x]$$

$\boxed{№2}$ MAE $\quad \mathrm{Med}(y|x)$

$\boxed{№3}$

$$L(y, a(x)) = \begin{cases} \alpha \cdot (y - a(x)), & a(x) \leq y \\ (1-\alpha)\cdot(a(x) - y), & a(x) \geq y \end{cases}$$



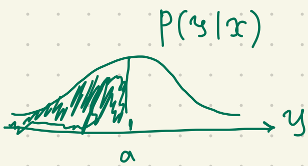


$$\frac{d}{d\alpha}\left[\int_{a(\alpha)}^{b(\alpha)} f(t,\alpha)\,dt\right] = \int_{a(\alpha)}^{b(\alpha)} \frac{d f(t,\alpha)}{d\alpha}\,dt + f(b(\alpha),\alpha)\cdot\frac{d b(\alpha)}{d\alpha} -$$

$$- f(a(\alpha),\alpha)\cdot\frac{d a(\alpha)}{d\alpha}$$

$$\mathbb{E}\left[L(y,a)\,|\,x\right] = \int_{\substack{a>y\\ -\infty}}^{a} (1-\alpha)\cdot(a-y)\cdot p(y|x)\,dy \; +$$

$$+ \int_{a<y}^{+\infty} \alpha\cdot(y-a)\cdot p(y|x)\,dy \longrightarrow \min_a$$

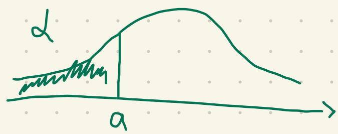


$$\frac{\partial}{\partial a}\left(\ldots\right) = (1-\alpha)\cdot\underbrace{\int_{-\infty}^{a} p(y|x)\,dy}_{\mathbb{P}[y<a|x] \;=\; P} - \alpha\cdot\underbrace{\int_{a}^{+\infty} p(y|x)\,dy}_{1-\mathbb{P}(y<a|x) \;=\; 1-P} = 0$$

$$(1-\alpha)p - \alpha(1-p) = 0$$

$$p - \alpha p - \alpha + \alpha p = 0$$

$$\mathbb{P}[y < a \mid x] = \alpha \quad \longleftrightarrow \quad a - q_\alpha$$



$\boxed{№4}$

$$L(y, \hat{p}(x)) = -\left[y \cdot \ln \hat{p}(x) + (1-y)\cdot \ln(1-\hat{p}(x))\right]$$

$p(x)$ — реальность

$\hat{p}(x)$ — прогноз

y_i	0	1
	$1-p(x_i)$	$p(x_i) = \mathbb{P}(y_i = 1 \mid x_i)$

$L(y_i, \hat{p}(x_i))$	$-\ln(1-\hat{p}(x_i))$	$-\ln \hat{p}(x_i)$
	$1-p(x_i)$	$p(x_i)$

$$\mathbb{E}\left[L(y_i, \hat{p}(x_i)) \,\|\, x\right] =$$

$$-p(x)\ln\hat{p}(x) - (1-p(x))\ln(1-\hat{p}(x)) \longrightarrow \min_{\hat{p}}$$

$$\frac{\partial}{\partial \hat{p}}[\ldots] = -\frac{p}{\hat{p}} + \frac{1-p}{1-\hat{p}} = 0 \qquad \hat{p} = p = \mathbb{P}(y=1 \mid x)$$

$\boxed{N^o 5}$

$$L(y, \hat{p}(x)) = |y - \hat{p}(x)| \qquad y \in \{0,1\}$$

$$\mathbb{E}[L(y,\hat{p})\,|\,x] = p(x)\cdot[1-\hat{p}] + (1-p(x))\cdot[\cancel{+}\hat{p}] \xrightarrow[\hat{p}]{} \min$$

$$p - 2p\hat{p} + \hat{p} = p - \hat{p}[2p-1]$$

$$\frac{\partial}{\partial \hat{p}}(\ldots) = 2p - 1 = 0$$

$$\varnothing$$

$$p - \hat{p}\underbrace{[2p-1]}_{} \longrightarrow \min_{\hat{p}\in[0;1]}$$

$>0 \Rightarrow \hat{p} = 1$

$<0 \Rightarrow \hat{p} = 0$

$=0 \quad p = \frac{1}{2} \quad \hat{p}$ – поэтому

не вер.-ть

logloss — $\mathbb{P}(y = 1 \,|\, x)$

SVM — х.з. но не вер.-ть

7

Энтропия и дивергенция

Опр. A — событие

$\mathbb{P}(A)$ — вероятность

$S(A)$ — „удивление" от A

$$S(A) = \log_{\frac{1}{2}} \mathbb{P}(A)$$

$\mathbb{P}(A) = 1/2 \quad S(A) = 1$

$\mathbb{P}(A) = (1/2)^{10} \quad S(A) = 10$

X — дискр. сл. вел. „Акинатор" „Данетки"

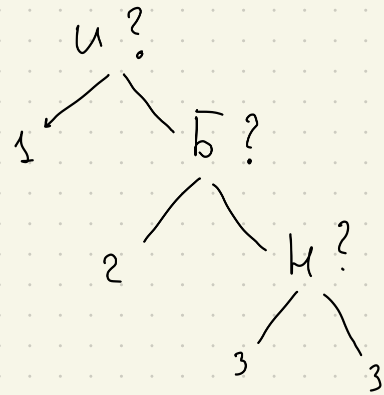
x	путин	навальный	бэтман	искандер
$\mathbb{P}(X=x)$	$1/8$	$1/8$	$1/4$	$1/2$

N — число вопросов

→ угадать за наим кол-во шагов, но гарантир.

min стратегия $\quad \max N$

→ угадать за наим. число шагов в среднем



$$\mathbb{E}(N) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 = \frac{7}{4} < 2$$

Опр. Энтропия $H(X)$

это минимально возможное $\mathbb{E}(N)$

мера непредсказуемости сл. вел. X

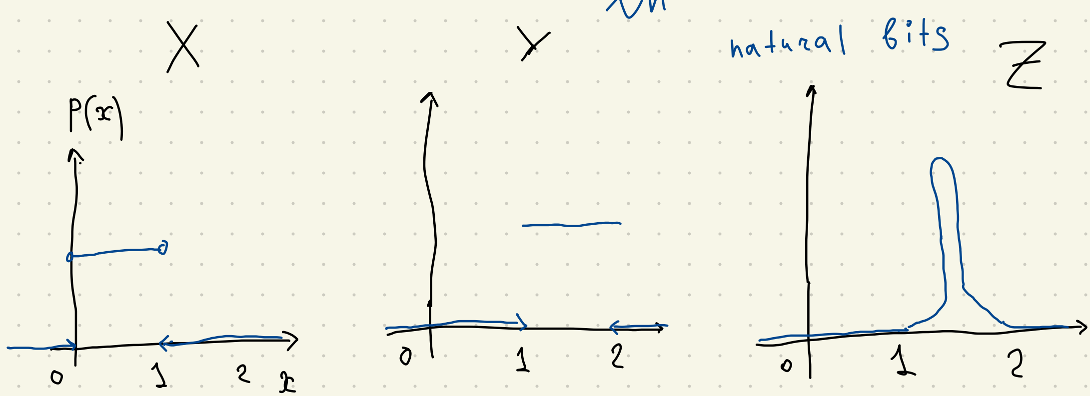
$$H(X) = \mathbb{E}\left[-\log \mathbb{P}(X = x)\right]$$

$P(x)$

$\log_2$ биты

$\ln$ наты

natural bits



$H(X) = H(Y)$ $\quad$ $H(Y) > H(Z)$

$\log_{\frac{1}{2}}$ → ?

биты → байты

Рг.

Manhat.

0 1

Хранитель

$ О Р

X - число подбр. до орла

$X_1 = 5$

$X_2 = 8$

....

1

бит

байт = 8 бит

$a = 0\,0\,0\,0\,0\,0\,0\,0$

$б = 1\,0\,0\,0\,0\,0\,0\,0$

$в = 0\,0\,1\,0\,0\,0\,0\,0$

$2^8 = 256$

$$H(X) = \mathbb{E}\left[-\log_2 \mathbb{P}(X=k)\right] = -\sum_{i=1}^{\infty} P_i \cdot \log P_i$$

$+\log_{\frac{1}{2}}$

$$= \frac{1}{2}\cdot\log_{\frac{1}{2}}\frac{1}{2} + \frac{1}{4}\cdot\log_{\frac{1}{2}}\frac{1}{4} + \ldots =$$

$$= \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \ldots = \sum_{i=1}^{\infty}\frac{i}{2^i} \overset{?}{=} 2$$