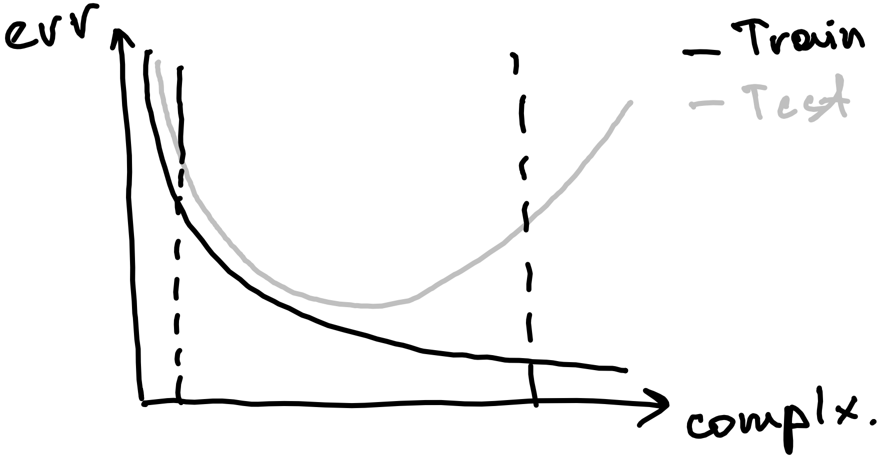
Bias-Variance Decomposition



$$\underbrace{b_1(x),\ b_2(x),\ \dots,\ b_N(x)}_{\text{agg} \to a(x)}$$

$$\text{agg} = \frac{1}{N}\sum(\ )$$

$$\boxed{b_j(x) - y(x) = \varepsilon_j(x)}$$

$y \in \mathbb{R}$, $p(x)$ $\quad b_j(x),\ j = 1, \dots, N$

$$\mathbb{E}_x\left[\left(y(x) - b_j(x)\right)^2\right] = \mathbb{E}_x\left(\varepsilon_j(x)\right)^2$$

$$\mathbb{E}_x\left[\left(y(x) - \frac{1}{N}\sum_{j=1}^{N} b_j(x)\right)^2\right] =$$

$$= \mathbb{E}_x \left(\frac{1}{N} \sum_{j=1}^{N} b_j(x) - \frac{1}{N} \sum_{j=1}^{N} y(x) \right)^2 =$$

$$= \mathbb{E}_x \left(\frac{1}{N} \sum_{j=1}^{N} \underbrace{\left(b_j(x) - y(x) \right)}_{\varepsilon_j(x)} \right)^2 =$$

$$= \mathbb{E}_x \left(\frac{1}{N} \sum_{j=1}^{N} \varepsilon_j(x) \right)^2 = \frac{1}{N^2} \mathbb{E}_x \left(\sum_j \varepsilon_j(x) \right)^2 =$$

$$= \frac{1}{N^2} \mathbb{E}_x \left(\sum_{j=1}^{N} \varepsilon_j^2(x) + \sum_{j \neq i} \varepsilon_j(x) \cdot \varepsilon_i(x) \right) =$$

$$\left\{ \begin{array}{l} \mathbb{E}_x \, \varepsilon_i(x) = 0 \\ \mathbb{E}_x \, \varepsilon_i(x) \cdot \varepsilon_j(x) = \mathbb{E} \, \varepsilon_j(x) \cdot \mathbb{E} \, \varepsilon_i(x) \end{array} \right\}$$

$$= \frac{1}{N^2} \sum_{j=1}^{N} \mathbb{E}\left(\varepsilon_j(x) \right)^2 = \frac{1}{N^2} \cdot N \cdot \mathbb{E}\left(\varepsilon_j(x) \right)^2 =$$

$$= \boxed{\frac{\mathbb{E}_x \left(\varepsilon_j(x) \right)^2}{N}}$$

Random Forest.

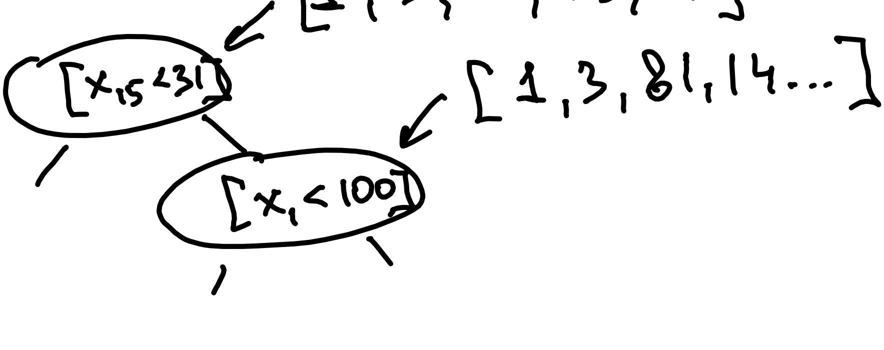
1. Bootstrap подвыборки.
2. Выбираем случ. фичи.

$$X \to \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_N$$

Вероятность, что конкр. объект попадёт в $\tilde{X}_k$?

$$1-\left(1-\frac{1}{n}\right)^n \xrightarrow[n\to\infty]{} 1-\frac{1}{e} \approx 63\%$$

$$b_1, \quad b_2, \quad \dots \quad b_N$$



$\boxed{BVD}$ $y \in \mathbb{R}$, $y, x \sim p(x, y)$

$a(x)$

$$R(a) = \mathbb{E}_{x,y} (y - a(x))^2 =$$

$$= \mathbb{E}_{x,y} (\underbrace{y - \mathbb{E}(y|x)} + \underbrace{\mathbb{E}(y|x) - a(x)})^2$$

$$= \mathbb{E}_{x,y} ((y - \mathbb{E}(y|x))^2 + (\mathbb{E}(y|x) - a(x))^2$$

$$+ 2 \underbrace{(y - \mathbb{E}(y|x)) \cdot (\mathbb{E}(y|x) - a(x))}_{*}) \to 0$$

$$* \; \mathbb{E}_{x,y} [(y - \mathbb{E}(y|x))(\cdot \mathbb{E}(y|x) - a(x))] = \\ = \triangle$$

$$\mathbb{E}_{x,y} f(x,y) = \int \int f(x,y) \cdot \underbrace{p(x,y)}_{p(y|x) \cdot p(x)} dx \cdot dy$$

$$= \int \Big(\underbrace{\int f(x,y) \cdot p(y|x) dy}_{\mathbb{E}_y [f(x,y) | x]} \Big) \cdot p(x) dx =$$

$$= \mathbb{E}_x\left[\mathbb{E}_y\left[f(x,y)\,|\,x\right]\right] \qquad \mathbb{E}_y\left[\mathbb{E}(y|x)\,|\,x\right]$$

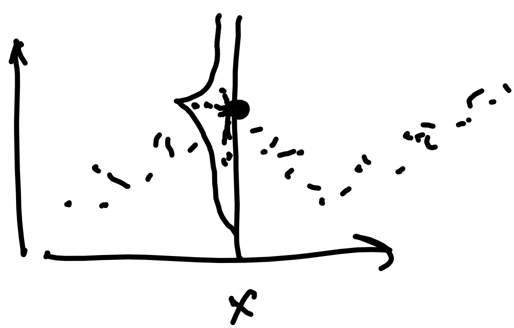
$$\Delta = \mathbb{E}_x\left[\mathbb{E}_y\left[(y - \mathbb{E}(y|x))\,|\,x\right] \cdot (\mathbb{E}(y|x) - a(x))\right] =$$

$$= \mathbb{E}_x\left[\underbrace{(\mathbb{E}(y|x) - \mathbb{E}(y|x))}_{0} \cdot (\mathbb{E}(y|x) - a(x))\right]$$

$$= 0$$

$$\boxed{R(a) = \mathbb{E}_{x,y}(y - a(x))^2 = \mathbb{E}_{x,y}(y - \mathbb{E}(y|x))^2 + \mathbb{E}_{x,y}(\mathbb{E}(y|x) - a(x))^2}$$

$$\boxed{a^*(x) = \mathbb{E}(y|x)}$$



$\mu(X^n)(x)$

$\mu(X^n)$ – модель, обученная на выборке X^n

$y = a \cdot x + \varepsilon, \quad \varepsilon \sim N(0, \sigma_2^2)$

$x \sim N(0, \sigma_1^2)$

$y | x \sim N(ax, \sigma_2^2)$

$$\sum_{i=1}^n (y_i - kx_i)^2$$

$\mu(X^n)(x) = k \cdot x$

$$\underbrace{\mu}\ (X^n) = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$R(a) = \mathbb{E}_{X^n}\left[\mathbb{E}_{x,y}\left[(y - \mu(X^n)(x))^2\right]\right] \ominus$$

$$p(X^n) = \prod_{i=1}^n p(x_i, y_i)$$

$$\ominus \int \left[\int (y - \mu(X^n)(x))^2 \cdot p(x,y)\,dx\,dy\right] \cdot$$

$$\cdot \prod_{i=1}^n p(x_i, y_i)\,dx_1 \ldots dx_n \cdot dy_1 \ldots dy_n$$

$$R(a) =$$

$$= \mathbb{E}_{x^n}\left[\mathbb{E}_{x,y}\left(y - \mathbb{E}(y|x)\right)^2 + \mathbb{E}_{x,y}\left(\mathbb{E}(y|x) - \mu(x^n)(x)\right)^2 \right]$$

$$= \mathbb{E}_{x,y}\left(y - \mathbb{E}(y|x)\right)^2 + \underbrace{\mathbb{E}_{x,y}\left[\mathbb{E}_{x^n}\left(\mathbb{E}(y|x) - \mu(x^n)(x)\right)^2\right]}_{*}$$

$$* = \mathbb{E}_{x^n}\Big(\underbrace{\mathbb{E}(y|x) - \mathbb{E}_{x^n}\mu(x^n)(x)} + \underline{\mathbb{E}_{x^n}\mu(x^n)(x) - \mu(x^n)(x)}\Big)^2 = \mathbb{E}_{x^n}\Big[\left(\mathbb{E}(y|x) - \mathbb{E}_{x^n}(\mu(x^n)(x))\right)^2$$

$$+ \left(\mathbb{E}_{x^n}(\mu(x^n)(x)) - \mu(x^n)(x)\right)^2 + \underbrace{2\cdot\left(\mathbb{E}(y|x) - \mathbb{E}_{x^n}\mu(x^n)(x)\right)\cdot\left(\mathbb{E}_{x^n}\mu(x^n)(x) - \mu(x^n)(x)\right)}_{0}\Big]$$

$$\boxed{\begin{aligned} R(a) = {} & \underbrace{\mathbb{E}_{x,y}\left[\left(y - \mathbb{E}(y|x)\right)^2\right]}_{\text{noise}} \\ + {} & \underbrace{\mathbb{E}_{x,y}\left[\left(\mathbb{E}(y|x) - \mathbb{E}_{x^n}\mu(x^n)(x)\right)^2\right]}_{\text{bias}} \\ + {} & \underbrace{\mathbb{E}_{x,y}\left[\mathbb{E}_{x^n}\left[\left(\mu(x^n)(x) - \mathbb{E}_{x^n}\mu(x^n)(x)\right)^2\right]\right]}_{\text{variance}} \end{aligned}}$$

Семинар

1. $x_{t+1} = x_t - \alpha \cdot \nabla f(x_t)$

$$x_{t+1} = x_t - \alpha \cdot H^{-1} \cdot \nabla f(x_t)$$

$$f(x) = x^T A x \quad , \quad A = A^T > 0$$

$$\nabla f(x) = 2 A x$$

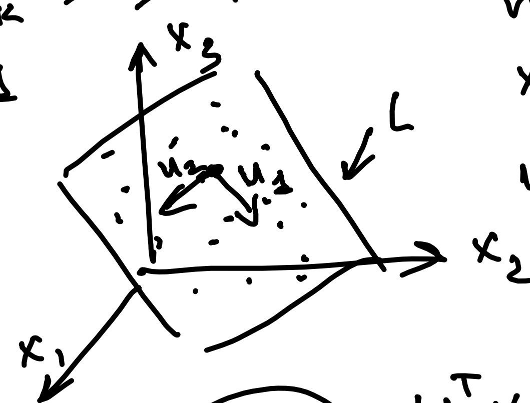
$$x_{t+1} = x_t - \alpha \cdot A x_t = (I - \alpha A) x_t .$$

$$x_{t+1} = x_t - \alpha \cdot \cancel{2 A^{-1}} \cdot \cancel{A} x_t = (1-\alpha) x_t$$

2. $\begin{cases} u_k^T S u_k \to \cancel{\min} \max\limits_{u_k} \\ \|u_k\| = 1 \end{cases}$ $\quad S = \dfrac{X^T X}{n}$

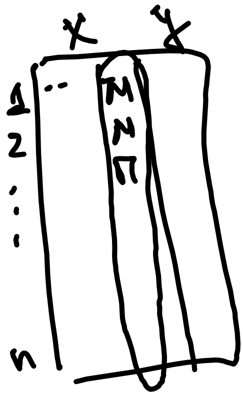
$x_i \in \mathbb{R}^3$

$u_k \in \mathbb{R}^3$



$$x_{proj} = \underbrace{\underset{d \times D}{U^T} \cdot x}_{d \times 1}$$

3




```
x = np.array(["M", "M", ...])
y = np.array([...]).
x_enc = [...]
map_dict = { }

unq_vals = np.unique(x)
for val in unq_vals:
    map_dict[val] = \
        y[x == val].mean()
x_enc = []
for val in x:
    x_enc.append(map_dict[val]).
```