

Ансамблирование

Рассматриваем ошибку ансамбля $\mathcal{A}(x) = \frac{1}{k} \sum_i^k A_i(x)$

$$\begin{aligned} L(\mathcal{A}) &= \underbrace{\mathbb{E}_{x,y} \left[(y - \mathbb{E}[y|x])^2 \right]}_{\text{Шум}} + \underbrace{\mathbb{E}_{x,y} \left[(\mathbb{E}_X [A_i(X)(x)] - \mathbb{E}[y|x])^2 \right]}_{\text{Смещение}} \\ &+ \underbrace{\frac{1}{k} \mathbb{E}_{x,y} \left[\mathbb{E}_X \left[\left(A_i(X)(x) - \mathbb{E}_X [A_i(X)(x)] \right)^2 \right] \right]}_{\text{Дисперсия } A_i} \\ &+ \underbrace{\frac{k(k-1)}{k^2} \mathbb{E}_{x,y} \left[\mathbb{E}_X \left[\left(A_i(X)(x) - \mathbb{E}_X [A_i(X)(x)] \right) \left(A_j(X)(x) - \mathbb{E}_X [A_j(X)(x)] \right) \right] \right]}_{\text{Ковариация } A_i, A_j} \end{aligned}$$

Оптимизируем критерий на выборке $(x_i, y_i)_{i=1}^m$:

$$(1 - \mu) \underbrace{\frac{1}{mk} \sum_{i=1}^k \sum_{j=1}^m (y_j - A_i(x_j))^2}_{\text{MSE}} + \mu \underbrace{\frac{1}{mk} \sum_{i=1}^k \sum_{j=1}^m (\mathcal{A}(x_j) - A_i(x_j))^2}_{\text{Дисперсия прогнозов}}$$