

Домашнее задание №2

Теоретические задачи

1) Ответы в листьях регрессионного дерева.

В листах объекты X_1, X_2, \dots, X_n с target-ами y_1, y_2, \dots, y_n

а) средние значения

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

б) случайное значение

$$y^* = y_i \text{ с веро-ю } \frac{1}{n}$$

В общем случае $MSE(\hat{y}) = E \sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (E y_i^2 - 2 E y_i \hat{y} + \hat{y}^2)$

а) $MSE(\bar{y}) = \frac{1}{n} \left(\sum_{i=1}^n y_i^2 - 2 \bar{y} \sum_{i=1}^n y_i + n \bar{y}^2 \right)$, а т.к. $\bar{y}^2 = (E y^*)^2 <$

б) $MSE(y^*) = \frac{1}{n} \left(\sum_{i=1}^n y_i^2 - 2 \sum_{i=1}^n y_i E(y^*) + n E y^{*2} \right) < E y^{*2}$

→ ошибки при выборе среднего значения меньше.

2) Unsupervised decision tree.
нормальн. распрег.