1.1. Ответы в листьях регрессионного дерева

Δ MSE выбора среднего элемента:

1) $$E\frac{1}{n}\sum_1^n (y_i-\bar y)^2 = \frac1n\left(\sum_1^n Ey_i^2 - 2\bar y\sum_1^n Ey_i + n\bar y^2\right) =$$

$$\overset{*}{=} \frac1n\sum_1^n Ey_i^2 - 2\bar y^2 + \bar y^2 = \frac1n\sum_1^n Ey_i^2 - \bar y^2 \quad (1)$$

$(*)$: $Ey_i = \sum_1^n \frac{y_i}{n} = \bar y$, т.к. y равновероятны

2) MSE выбора случайного элемента:

$$E\frac1n\sum_1^n (y_i-y_k)^2 = \frac1n\left(\sum_1^n Ey_i^2 - 2\sum_1^n E(y_iy_k) + nEy_k^2\right) =$$

$$= \frac1n\sum_1^n Ey_i^2 + \frac1n\left(-2\sum_{1,i\ne k}^n Ey_iEy_k - 2Ey_k^2 + nEy_k^2\right) =$$

$$= \frac1n\sum_1^n Ey_i^2 - 2\frac{n-1}{n}\bar y^2 + \frac{n-2}{n}Ey_k^2 \;\; (=)$$

Выразим $Dy_k = Ey_k^2 - (Ey_k)^2$

$$\frac{n-2}{n}Dy_k = \frac{n-2}{n}Ey_k^2 - \frac{n-2}{n}(Ey_k)^2 = \frac{n-2}{n}Ey_k^2 - \frac{n-2}{n}\bar y^2$$

$$(=)\; \frac1n\sum_1^n Ey_i^2 + \frac{n-2}{n}Dy_k + \frac{-2n+2+n-2}{n}\bar y^2 =$$

$$= \frac1n\sum_1^n Ey_i^2 - \bar y^2 + \frac{n-2}{n}Dy_k \quad (2)$$

Сравним (1) и (2). Видим, что $(2)-(1) = \frac{n-2}{n}Dy_k$

Дисперсия случайной величины — ~~положительна~~ неотрицательна.

При выборе случайного элемента ошибка увеличивается на $\frac{n-2}{n}Dy_k$.