Задача 2 Разложение ошибки

$y_i = 3x_i^2 + u_i$ – истинная зависимость

$$x_i = \begin{cases} 0, \text{ с вероят. } 1/3 \\ 1, \text{ с вероят. } 1/3 \\ 2, \text{ с вероят. } 1/3 \end{cases}$$

$$u_i = \begin{cases} 1, \text{ с вероят. } 1/2 \\ -1, \text{ с вероят } 1/2 \end{cases}$$

$\hat{y}_i = wx_i$ – модель, $X = (x_i, y_i)_{i=1}^N$

$$Q = \frac{1}{N}\sum_{i=1}^{N}(y_i - wx_i)^2 \longrightarrow \min_w$$

$$\frac{\partial Q}{\partial w} = \frac{2}{N}\sum_{i=1}^{N}(y_i - wx_i)(-x_i) = 0.$$

$$\sum_{i=1}^{N}(wx_i - y_i)x_i = 0.$$

$$w = \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2}$$

$$\hat{y}_i = x_i \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2}$$

$$E[(y-\hat{y})^2] = \underbrace{\mathcal{D}[y]}_{\text{шум}} + \underbrace{\left(E[\hat{y}] - E[y]\right)^2}_{\text{смещение}} + \underbrace{\mathcal{D}[\hat{y}]}_{\text{разброс}}$$

(фикс. X).

1) Шум: $\mathcal{D}[y] = E[y - E[y]]^2 = E[3x_i^2 + u_i - E[3x_i^2 + u_i]]^2 =$

$= E[3x_i^2 + u_i - 3x_i^2]^2 = E[u_i^2] = 1$

2) Смещение: $E[y_i] = E[3x_i^2 + u_i] = 3E[x_i^2] + E[u_i] = 3 \cdot \frac{5}{3} = 5$

$$E[\hat{y}_i] = E\left[x_i \cdot \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2}\right] = E[x_i] E\left[\frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2}\right] =$$

$$= 1 \cdot E\left[\sum_{i=1}^{N} y_i x_i\right] \cdot E\left[\sum_{i=1}^{N} \frac{1}{x_i}\right] = E[y_i] \cdot E[x_i] \cdot E\left[\left(\frac{1}{x_i}\right)^2\right] =$$

$$= 5 \cdot \frac{1}{12} \cdot 5 = \frac{25}{12}$$

$$\text{Bias} = \left(5 - \frac{25}{12}\right)^2 = \left(\frac{60-25}{12}\right)^2 = \left(\frac{35}{12}\right)^2 = \frac{1225}{144}$$

3) Разброс: $\mathbb{D}[\hat{y}] = E(\hat{y}^2) - (E(\hat{y}))^2$

$E[\hat{y}^2] = E[x_i^2] \cdot E\left[\frac{\sum y_i^2 x_i^2}{\sum x_i^4}\right] = E[x_i^2] \cdot E[y_i^2] \cdot E[x_i^2] \cdot$

$\cdot E\left[\left(\frac{1}{x_i}\right)^4\right] = E[x_i^2]\left(E[9x_i^4] + E[6x_i^2 u_i] + E[u_i^2]\right) \cdot E[x_i^2] \cdot$

$\cdot E\left[\left(\frac{1}{x_i}\right)^4\right] = \frac{5}{3} \cdot 62 \cdot \frac{5}{3} \cdot \frac{17}{48} = \frac{5 \cdot 5 \cdot 62 \cdot 17}{3 \cdot 3 \cdot 48} = \frac{5525}{108}$

$\mathbb{D}[\hat{y}] = \frac{5525}{108} - \frac{625}{144} = \frac{20225}{432}$

Ответ: $\text{Шум} = 1$

$\text{Bias} = \frac{1225}{144}$

$\text{Variance} = \frac{20225}{432}$