

# Algorytmiczna Analiza Danych

## Ćwiczenia 3

2025-10-17

Adrian Herda  
Politechnika Wrocławska

### 1. Zadanie 8 - $MSE$ jako suma kwadratu $bias$ 'u wariancji funkcji przybliżającej i wariancji szumu

$$MSE = \mathbb{E}[(y_0 - \hat{f}(x_0))^2] = \text{Bias}(\hat{f}(x_0))^2 + \text{Var}(\hat{f}(x_0)) + \sigma_\varepsilon^2 \quad (1)$$

#### 1.1. Rozwiązanie

1.  $\mathbb{E}[\varepsilon] = 0 \Rightarrow \mathbb{E}[\varepsilon]^2 = 0 \Rightarrow \sigma_\varepsilon^2 = \mathbb{E}[\varepsilon^2]$
2. Szum jest niezależny z żadną inną zmienną

*Proof.*

$$\begin{aligned} \mathbb{E}[(y_0 - \hat{f}(x_0))^2] &= \mathbb{E}[(f(x_0) + \varepsilon - \hat{f}(x_0))^2] \\ &= \mathbb{E}[(f(x_0) - \hat{f}(x_0))^2] + \underbrace{\mathbb{E}[\varepsilon^2]}_{\sigma_\varepsilon^2} + 2\mathbb{E}[(f(x_0) - \hat{f}(x_0))\varepsilon] \\ &= \mathbb{E}[(f(x_0) - \hat{f}(x_0))^2] + \sigma_\varepsilon^2 + \underbrace{2(f(x_0) - \mathbb{E}[\hat{f}(x_0)])\mathbb{E}[\varepsilon]}_0 \\ &= \mathbb{E}[f(x_0)^2 + \hat{f}(x_0)^2 - 2f(x_0)\hat{f}(x_0)] + \sigma_\varepsilon^2 \\ &= f(x_0)^2 + \mathbb{E}[\hat{f}(x_0)^2] - 2f(x_0)\mathbb{E}[\hat{f}(x_0)] + \sigma_\varepsilon^2 \\ &= f(x_0)^2 - 2f(x_0)\mathbb{E}[\hat{f}(x_0)] + \underbrace{\mathbb{E}[\hat{f}(x_0)^2] - \mathbb{E}[\hat{f}(x_0)]^2}_0 + \mathbb{E}[\hat{f}(x_0)^2] + \sigma_\varepsilon^2 \\ &= (f(x_0) - \mathbb{E}[\hat{f}(x_0)])^2 + (\mathbb{E}[\hat{f}(x_0)^2] - \mathbb{E}[\hat{f}(x_0)]^2) + \sigma_\varepsilon^2 \\ &= \text{Bias}(\hat{f}(x_0))^2 + \text{Var}(\hat{f}(x_0)) + \sigma_\varepsilon^2 \end{aligned} \quad (2)$$

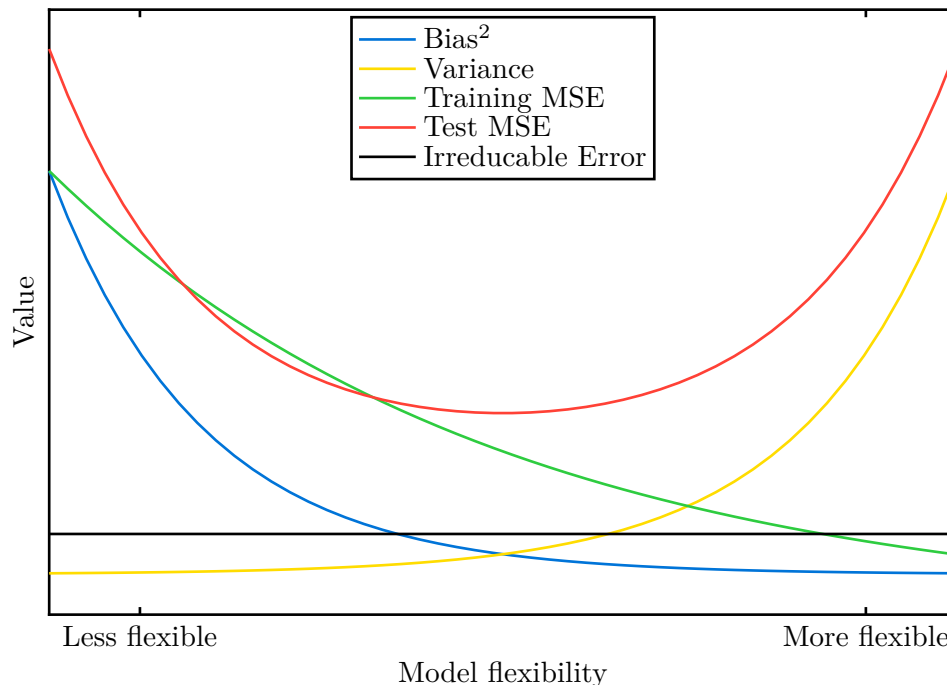
■

#### 1.2. Równowaga bias-variance

W miarę jak model staje się bardziej elastyczny, jego zdolność do uchwycenia złożoności danych rośnie, co prowadzi do zmniejszenia błędu bias. Jednakże, większa elastyczność może również prowadzić do większej wariancji, ponieważ model może zacząć „przeuczać” się na dane treningowe. W rezultacie, istnieje kompromis między bias a wariancją, który należy uwzględnić podczas projektowania modeli.

Optymalny wybór złożoności/regularizacji minimalizuje sumę  $\text{Bias}^2 + \text{Var}$ .  
Składnik  $\varepsilon$  jest nieredukowalny.

## 2. Zadanie 9 - Wykresy względem elastyczności modelu



### 2.1. Wytlumaczenie każdej funkcji

**Niebieski** - W miarę jak modele stają się bardziej elastyczne i złożone, błąd treningowy maleje, ale z coraz mniejszym zyskiem.

**Czerwony** - Często obserwujemy kształt litery U w błędzie na zbiorze walidacyjnym. Jest to kombinacja błędu Bias (obserwowanego przy najmniej elastycznym modelu), błędu Variance (wynikającego z przeuczenia najbardziej elastycznych modeli) oraz składnika błędu nieusuwalnego.

**Niebieski** - Składnik błędu Bias, im bardziej elastyczny model tym lepiej się dopasowuje tym mniejszy jest Bias.

**Brązowy** - Składnik błędu Variance. Im bardziej elastyczny model tym większe różnice w wynikach na zbiorze treningowym i walidacyjnym.

**Czarny** - Składnik błędu nieusuwalnego. Niezmienna wartość szumu.

## 3. Zadanie 10 - Obliczanie wzorów na parametry modelu linowego

### 3.1. Dane

- $n$  obserwacji:  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ ,
- model linowy:  $y_i = \beta_0 + \beta_1 x_i + e_i$

### 3.2. Cel

Szacujemy  $\beta_0, \beta_1$  poprzez minimalizację błędu średniokwadratowego:

$$\text{MSE}(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^n \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)^2 \quad (3)$$

Udowodnij że:

1.  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
2.  $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

*Proof.* Różniczkujemy i przyrównujemy do zera:

$$\frac{\partial}{\partial \hat{\beta}_0} \text{MSE} = -\frac{2}{n} \sum_{i=1}^n \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right) = 0 \quad (4)$$

$$\frac{\partial}{\partial \hat{\beta}_1} \text{MSE} = -\frac{2}{n} \sum_{i=1}^n x_i \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right) = 0 \quad (5)$$

Z równania 4 mamy:

$$\begin{aligned} \sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i &= 0 \\ n\hat{\beta}_0 &= \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \quad (6)$$

Podstawiając do równania 5 mamy:

$$\begin{aligned}
0 &= \sum_{i=1}^n x_i y_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \\
0 &= \sum_{i=1}^n x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \\
0 &= \sum_{i=1}^n x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \cdot n\bar{x} - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \\
0 &= \sum_{i=1}^n x_i y_i - n\bar{y}\bar{x} + n\hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \\
\hat{\beta}_1 \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) &= \sum_{i=1}^n x_i y_i - n\bar{y}\bar{x} \\
\hat{\beta}_1 \left( \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \right) &= \sum_{i=1}^n x_i y_i - n\bar{y}\bar{x} + n\bar{y}\bar{x} - n\bar{y}\bar{x} \\
\hat{\beta}_1 \left( \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2\bar{x}x_i + \sum_{i=1}^n \bar{x}^2 \right) &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \bar{y}x_i + \sum_{i=1}^n \bar{y}\bar{x} - \sum_{i=1}^n \bar{x}y_i \\
\hat{\beta}_1 \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) &= \sum_{i=1}^n (x_i y_i - \bar{y}x_i) + (\bar{y}\bar{x} - \bar{x}y_i) \\
\hat{\beta}_1 \sum_{i=1}^n (x_i^2 - \bar{x}^2) &= \sum_{i=1}^n x_i(y_i - \bar{y}) - \bar{x}(y_i - \bar{y}) \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned} \tag{7}$$

■

### 3.3. Udowodnij że funkcja ta zawsze przejdzie przez punkt $(\bar{x}, \bar{y})$

*Proof.* Podstawiając  $x = \bar{x}$  do równania 4 prostej:

$$\begin{aligned}
y &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
&= (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \bar{x} \\
&= \bar{y}
\end{aligned} \tag{8}$$

■

## 4. Zadanie 11 - Błąd systematyczny, wariancja i błąd standardowy dla estymatorów w modelu liniowym

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{9}$$

Estymaotry jak w modelu z zadania w sekcji 3.

Oraz niech:  $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$

## 4.1. Błąd systematyczny

$$E[\hat{\beta}_0] = \beta_0 \Rightarrow \text{Bias}(\hat{\beta}_0) = 0 \quad (10)$$

$$E[\hat{\beta}_1] = \beta_1 \Rightarrow \text{Bias}(\hat{\beta}_1) = 0 \quad (11)$$

## 4.2. Wariancja

Zmienne  $e_i$  dla  $i \in \{1, \dots, n\}$  są niezależne

### 4.2.1. Reprezentacja estymatorów za pomocą szumów

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})e_i}{S_{xx}} \quad (12)$$

$$\hat{\beta}_0 = \beta_0 + \frac{1}{n} \sum_{i=1}^n e_i - \bar{x} \frac{\sum_{i=1}^n (x_i - \bar{x})e_i}{S_{xx}} \quad (13)$$

### 4.2.2. Wariancja estymatorów

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{Var}\left(\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})e_i}{S_{xx}}\right) \\ &= \text{Var}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})e_i}{S_{xx}}\right) \\ &= \frac{1}{(S_{xx})^2} \text{Var}\left(\sum_{i=1}^n (x_i - \bar{x})e_i\right) \\ &= \frac{1}{(S_{xx})^2} \sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(e_i) \\ &= \frac{\sigma^2}{S_{xx}} \end{aligned} \quad (14)$$

$$\begin{aligned}
\text{Var}(\hat{\beta}_0) &= \text{Var}\left(\beta_0 + \frac{1}{n} \sum_{i=1}^n e_i - \bar{x} \frac{\sum_{i=1}^n (x_i - \bar{x}) e_i}{S_{xx}}\right) \\
&= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n e_i - \bar{x} \frac{\sum_{i=1}^n (x_i - \bar{x}) e_i}{S_{xx}}\right) \\
&= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n e_i\right) + \text{Var}\left(\bar{x} \frac{\sum_{i=1}^n (x_i - \bar{x}) e_i}{S_{xx}}\right) - 2 \text{Cov}\left(\frac{1}{n} \sum_{i=1}^n e_i, \bar{x} \frac{\sum_{i=1}^n (x_i - \bar{x}) e_i}{S_{xx}}\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(e_i) + \frac{\bar{x}^2}{(S_{xx})^2} \sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(e_i) - 2 \frac{\bar{x}}{n S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \text{Var}(e_i) \\
&= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{S_{xx}} - 2 \bar{x} \frac{\sigma^2}{n S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \\
&= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{S_{xx}} - 2 \bar{x} \frac{\sigma^2}{n S_{xx}} \cdot 0 \\
&= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{S_{xx}} \\
&= \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)
\end{aligned} \tag{15}$$

#### 4.3. Błąd standardowy estymatorów

$$\text{SE}(\hat{\beta}_0) = \sqrt{\text{Var}(\hat{\beta}_0)} = \sqrt{\sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} = \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \tag{16}$$

$$\text{SE}(\hat{\beta}_1) = \sqrt{\text{Var}(\hat{\beta}_1)} = \sqrt{\frac{\sigma^2}{S_{xx}}} = \frac{\sigma}{\sqrt{S_{xx}}} \tag{17}$$

### 5. Zadanie 12 - Suma zmiennych losowych o rozkładzie normalnym ma rozkład normalny

1.  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$
2.  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$
3.  $Z = X + Y$

Udowodnij że  $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$

*Proof.* Funkcje charakterystyczne zmiennych losowych:

$$\varphi_X(t) = \exp\left(i\mu_X t - \sigma_X^2 \frac{t^2}{2}\right), \quad \varphi_Y(t) = \exp\left(i\mu_Y t - \sigma_Y^2 \frac{t^2}{2}\right) \tag{18}$$

Zmienna losowa  $Z$  jest sumą niezależnych zmiennych losowych  $X$  i  $Y$ , więc jej funkcja charakterystyczna jest iloczynem funkcji charakterystycznych  $X$  i  $Y$ :

$$\begin{aligned}\varphi_Z(t) &= \varphi_X(t) \cdot \varphi_Y(t) \\&= \exp\left(i\mu_X t - \sigma_X^2 \frac{t^2}{2}\right) \cdot \exp\left(i\mu_Y t - \sigma_Y^2 \frac{t^2}{2}\right) \\&= \exp\left(i\mu_X t - \sigma_X^2 \frac{t^2}{2} + i\mu_Y t - \sigma_Y^2 \frac{t^2}{2}\right) \\&= \exp\left(i(\mu_X + \mu_Y)t - (\sigma_X^2 + \sigma_Y^2) \frac{t^2}{2}\right) \\&= \exp\left(i\mu_Z t - \sigma_Z^2 \frac{t^2}{2}\right)\end{aligned}\tag{19}$$

■