Conditional Expectation and Quadratic Loss

Your Name

Decomposition of the Conditional Quadratic Loss

- Given a response variable Y and a predictor X, consider the **conditional expected squared error (CMSE)**:

$$\mathbb{E}\left((Y-y)^2 \mid X=x\right).$$

- We aim at finding a lower bound for the MSE.
- For this we rely on the identity:

$$Y-y=(Y-\mathbb{E}[Y \mid X])+(\mathbb{E}[Y \mid X]-y),$$

Expanding the Expectation

- Expand the square of the previous identity as:

$$\begin{gathered}(Y-y)^2=(Y-\mathbb{E}[Y\mid X])^2+(\mathbb{E}[Y\mid X]-y)^2\\ +2(Y-\mathbb{E}[Y\mid X])(\mathbb{E}[Y\mid X]-y).\end{gathered}$$

- Taking conditional expectations on both sides and noticing that the product vanishes by the properties of the conditional expectation:

$$\mathbb{E}\left((Y-y)^2\mid X\right)=\underbrace{\mathbb{E}\left((Y-\mathbb{E}[Y\mid X])^2\mid X\right)}_{\text{Conditional Variance}}+\underbrace{(\mathbb{E}[Y\mid X]-y)^2}_{\text{Conditional Bias}}.$$

Key Inequality

Since the last term is always **non-negative**, we obtain:

$$\mathbb{E}\left((Y-y)^2 \mid X\right) \geq \mathbb{E}\left((Y-\mathbb{E}[Y \mid X])^2 \mid X\right).$$

- The **minimum** occurs when:

$$y = \mathbb{E}[Y \mid X].$$

- This shows that the **best prediction under quadratic loss** is the **conditional expectation**.

Interpretation

- The **expected squared error** decomposes into two terms:

$$\text{Total Error} = \text{Irreducible Variance} + \text{Squared Bias}.$$

- The best predictor minimizes the **bias term**.
- In regression, the optimal predictor (in MSE sense) is:

$$h^*(x) = \mathbb{E}[Y \mid X = x].$$