① Modeling

Have a variable that we want to predict

Have some data about the variable

Ⓐ Propose a model $\boxed{\hat{y} = 5}$

Ⓑ Choose a loss function

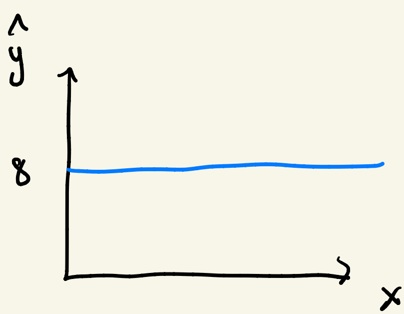
Ⓒ Choose the model with lowest value for loss function

A Better Model

$\hat{y} = \theta$, θ is a scalar variable

$\theta = 8$ is the best choice of θ given our data

$\hat{y} = 8$



Loss Functions

Helps us choose θ value by assigning a numerical score that represents how good/bad that choice of θ is.

We choose θ that minimizes loss function

$$\text{model: } \hat{y} = \theta$$

$$\hat{\theta} = \arg\min_{\theta} L(\theta)$$

Minimizing $L(\theta)$ is the same as minimizing any other function!

$$\begin{cases} f'(x) = 0 & \text{then solve for } x \\ f''(x) \geq 0 & \leftarrow x \text{ is a minimum} \end{cases}$$

$$L'(\theta) = 0, \quad \text{then solve for } \theta$$

$$\boxed{L''(\theta) \geq 0 \quad \Rightarrow \quad \theta \text{ is a minimum}}$$

$L(\theta)$ is convex!

$$\boxed{\text{Model: } \hat{y} = \theta}$$

Loss function: $L(\theta) = \boxed{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

$L'(\theta) = 0$ solve for θ

$$L'(\theta) = 0 \qquad \boxed{L''(\hat{\theta}) > 0}$$

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)^2$$

$$L'(\theta) = \frac{1}{n}\sum_{i=1}^{n} 2(y_i - \theta) \cdot -1 = 0$$

$$-\frac{1}{n}\sum_{i=1}^{n} 2(y_i - \theta) = 0$$

$$\sum_{i=1}^{n} 2(y_i - \theta) = 0$$

$$\sum_{i=1}^{n}(y_i - \theta) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\theta = 0$$

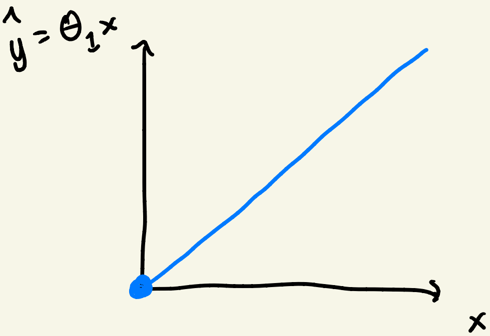
$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}\theta$$

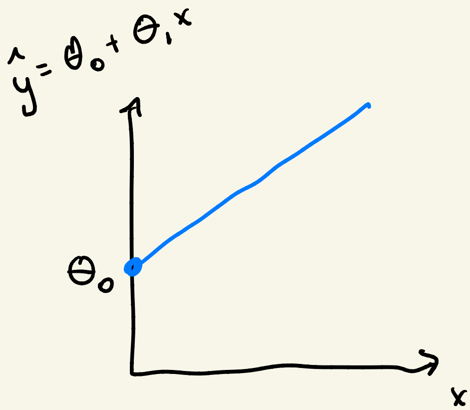
$$\sum_{i=1}^{n} y_i = n\theta \Rightarrow \boxed{\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}}$$

Better Model (SLR)

$\boxed{\hat{y} = \theta_0 + \theta_1 x}$, θ_0 & θ_1 are scalars

$\hat{y} = \theta_1 x$





$\ell(\theta) = (y - \hat{y})^2$ Squared Error (1 point)

$L(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2$ MSE

$$\begin{cases} \frac{\partial}{\partial \theta_0} L(\theta_0, \theta_1) = 0 \\ \frac{\partial}{\partial \theta_1} L(\theta_0, \theta_1) = 0 \end{cases}$$

solving for θ_0, θ_1 gives us $\hat{\theta}_0, \hat{\theta}_1$

then the model can be updated to be

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$$