Data Analysis(1)

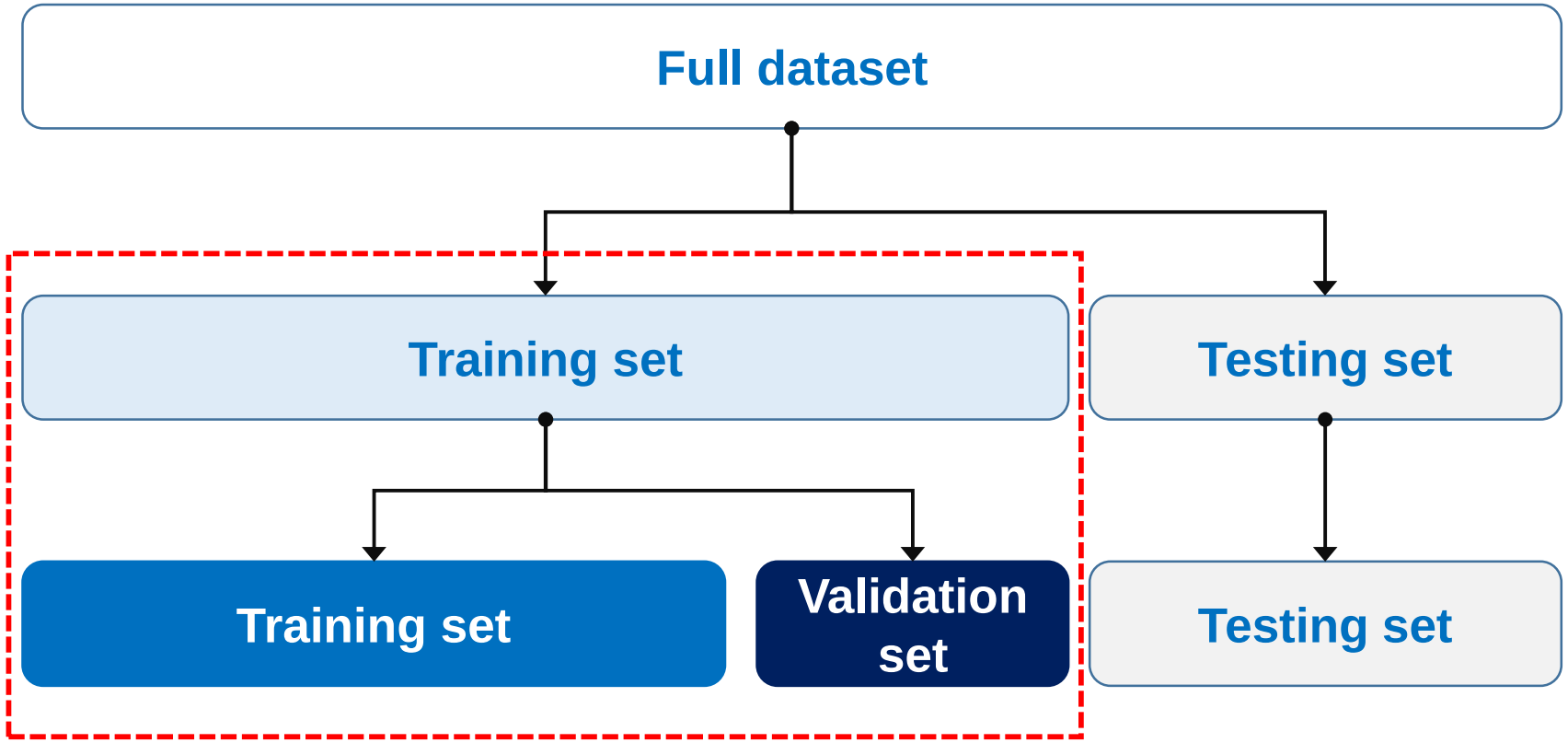
Dept. of Mechanical System Design Engineering,
Seoul National University of Science and Technology

Prof. Ju Yeon Lee
(jylee@seoultech.ac.kr)

Review

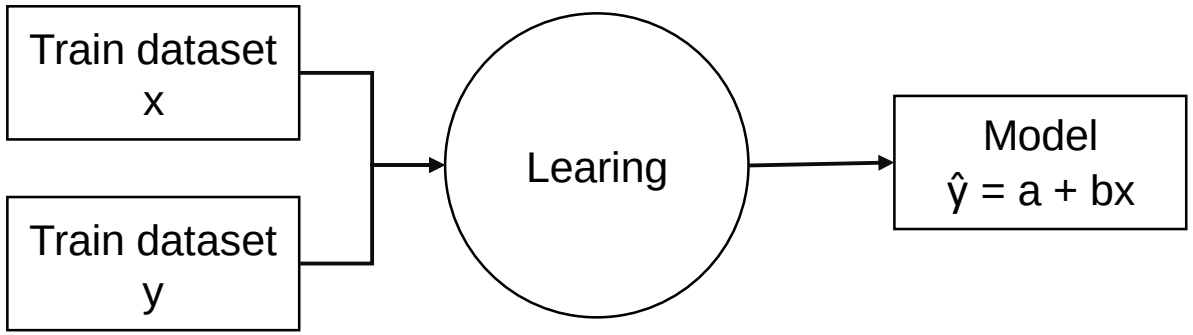
Applying Machine Learning Algorithms for Missing Data

Datasets

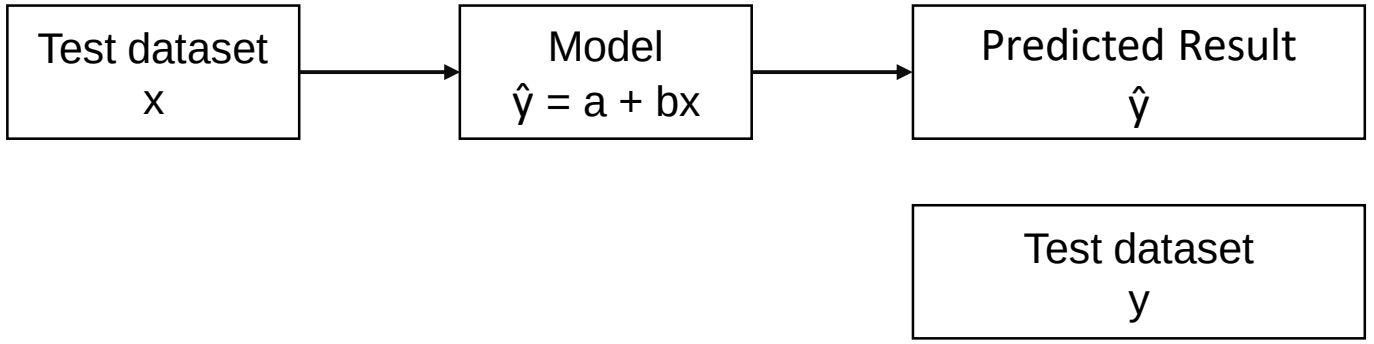

Full dataset
Training set
Testing set
Training set
Validation set
Testing set

Datasets (Supervised Learning)

- **Train dataset : dataset to find the model**

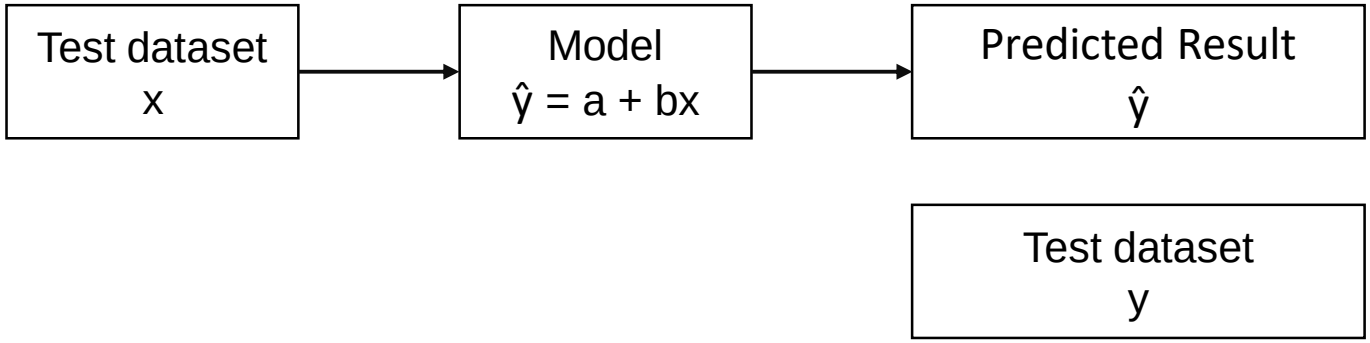


- **Test dataset : dataset to validate the model**



Model Evaluation (Regression)

- **Test dataset : dataset to validate the model**



Error (Data analysis) or residual (Statistics) : y – ŷ

SSE (Sum of Squared Error)

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Test set, Predicted value, Actual value

MSE (Mean Squared Errors)

$$MSE = \frac{SSE}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Test set, Predicted value, Actual value

RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{MSE} = \sqrt{SSE/n}$$

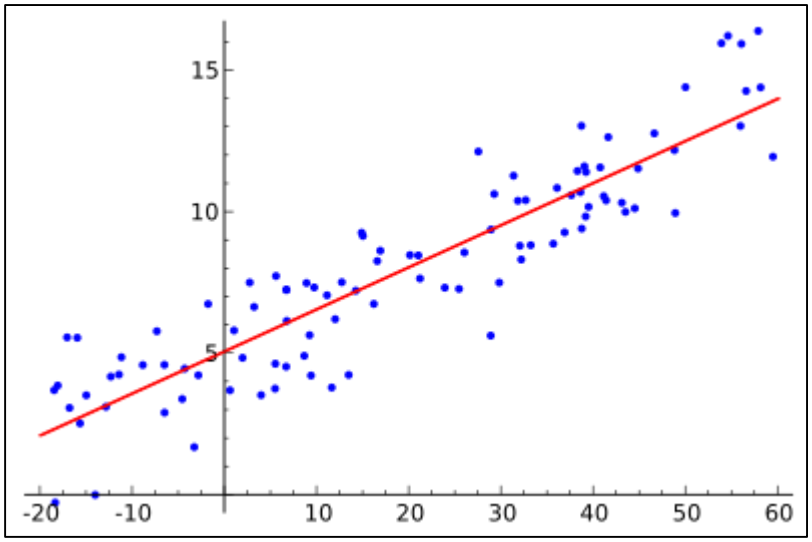
$$= \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$

Test set, Predicted value, Actual value

Linear Regression

- **Simple linear regression :**
  The very simplest case of a single scalar predictor variable x and a single scalar response variable y

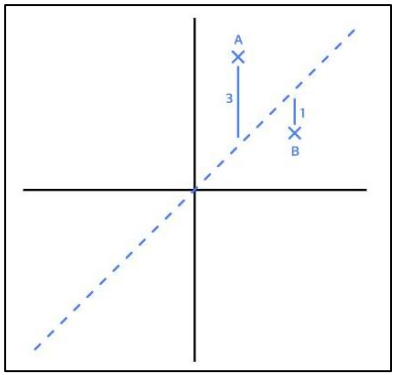
$$\hat{y} = b_0 + b_1 x_1$$



- **Multiple linear regression :**
  generalization of simple linear regression to the case of more than one independent variable; a special case of general linear models, restricted to one dependent variable

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4$$

$$MSE = \frac{SSE}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$




Thank you
Q & A