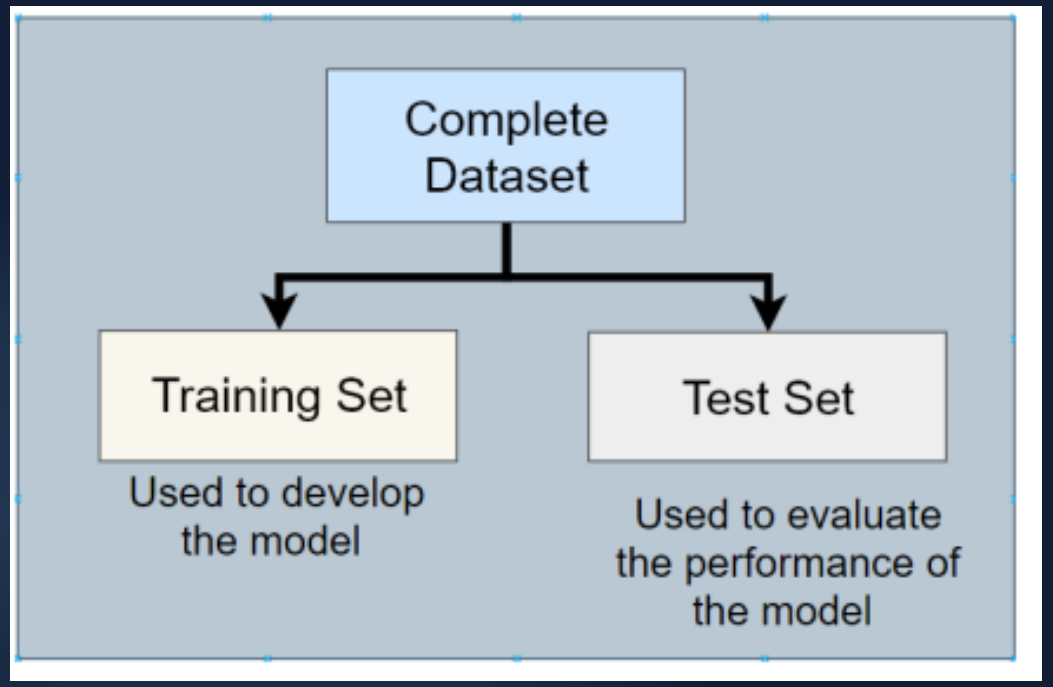
Week 3 Live Session

NEBA NFONSANG

Data Splitting

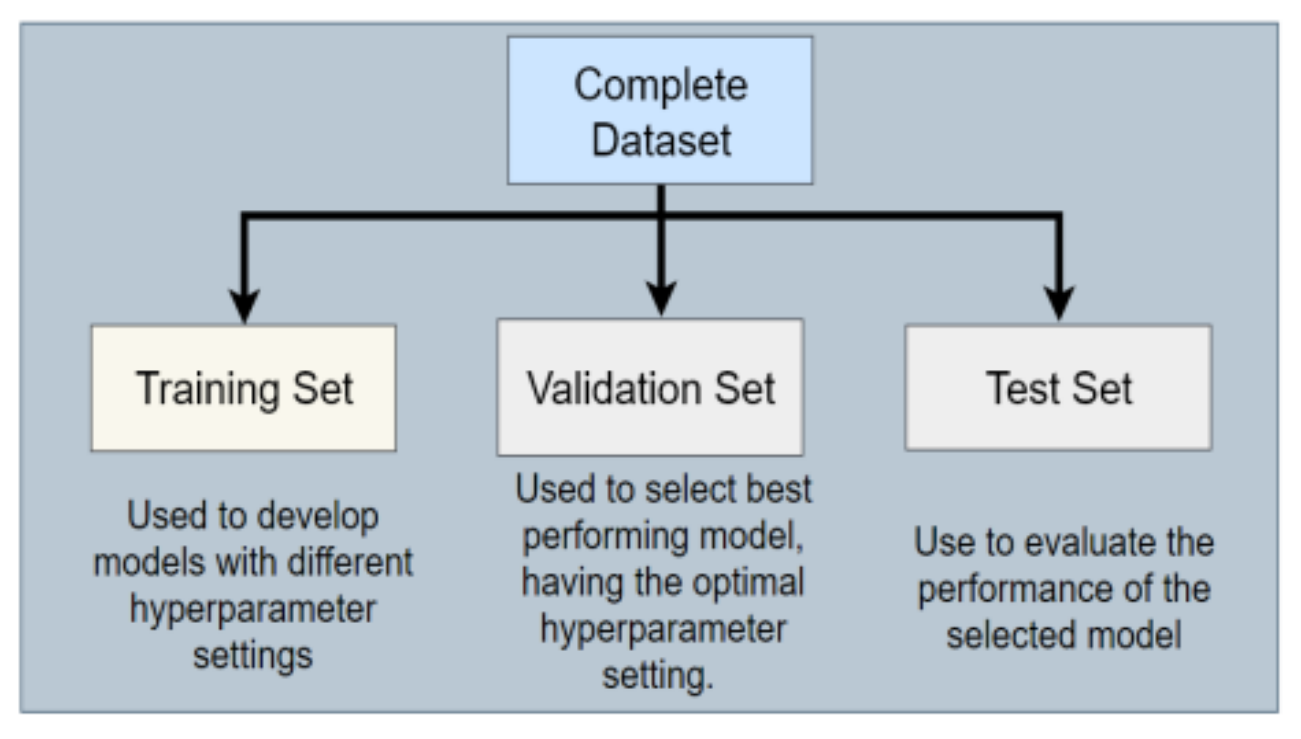
- Why do we need to split our data when building supervised learning models?
- What if we evaluate the model only on the training set?
- If you split your data into two sets, what ratios are commonly used?
- When should you use a higher proportion of the data for the training set?
- When is it okay to use a lower proportion of the data for training?

Data Partitioning



- What is the name of this data partitioning approach?
- What is the downside of using this type of data partitioning approach?

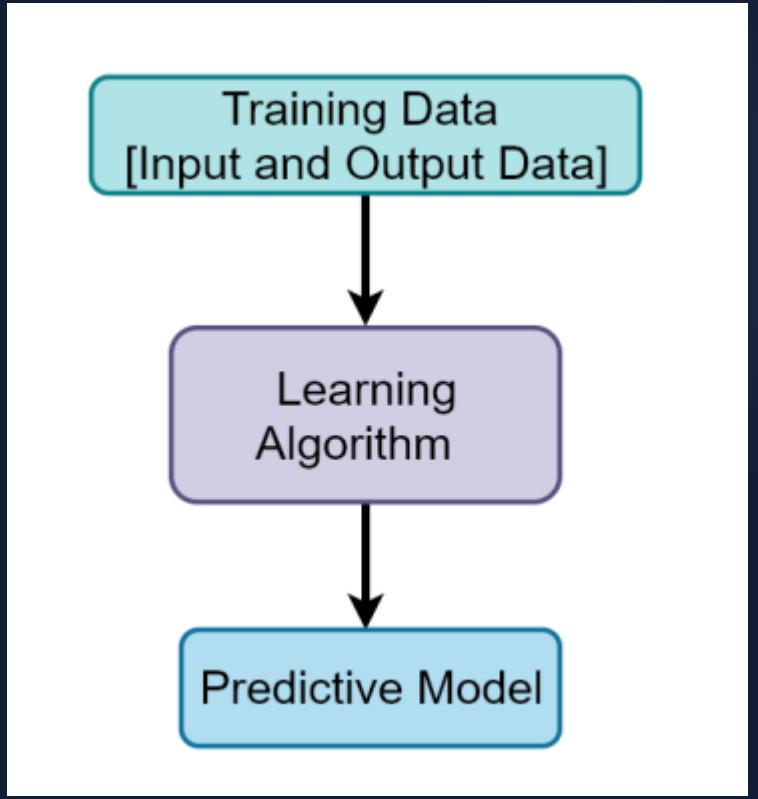
Data Partitioning



- What is the name of this data partitioning approach?
- What is the advantage of using this type of data partitioning approach?

Model Construction

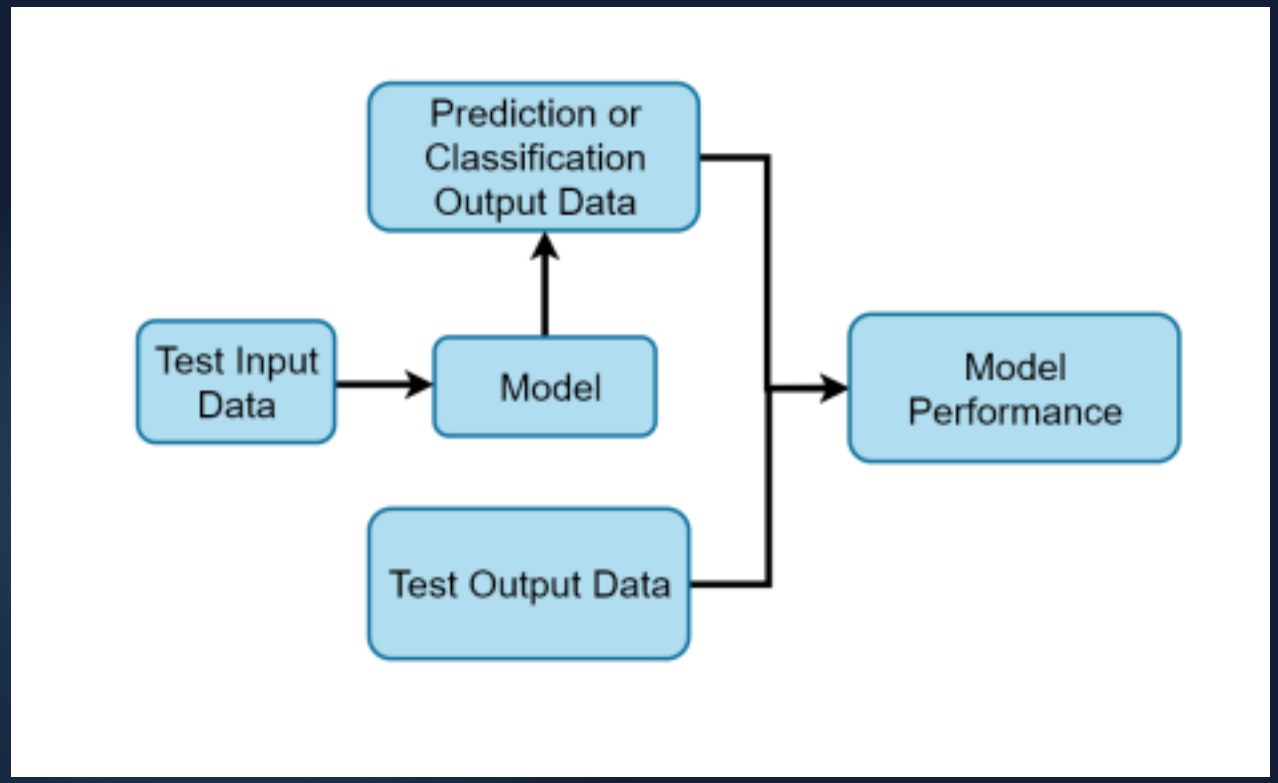
- Does this diagram represent model construction in a supervised learning or in an unsupervised learning?
- What is the difference between supervised learning and unsupervised learning?



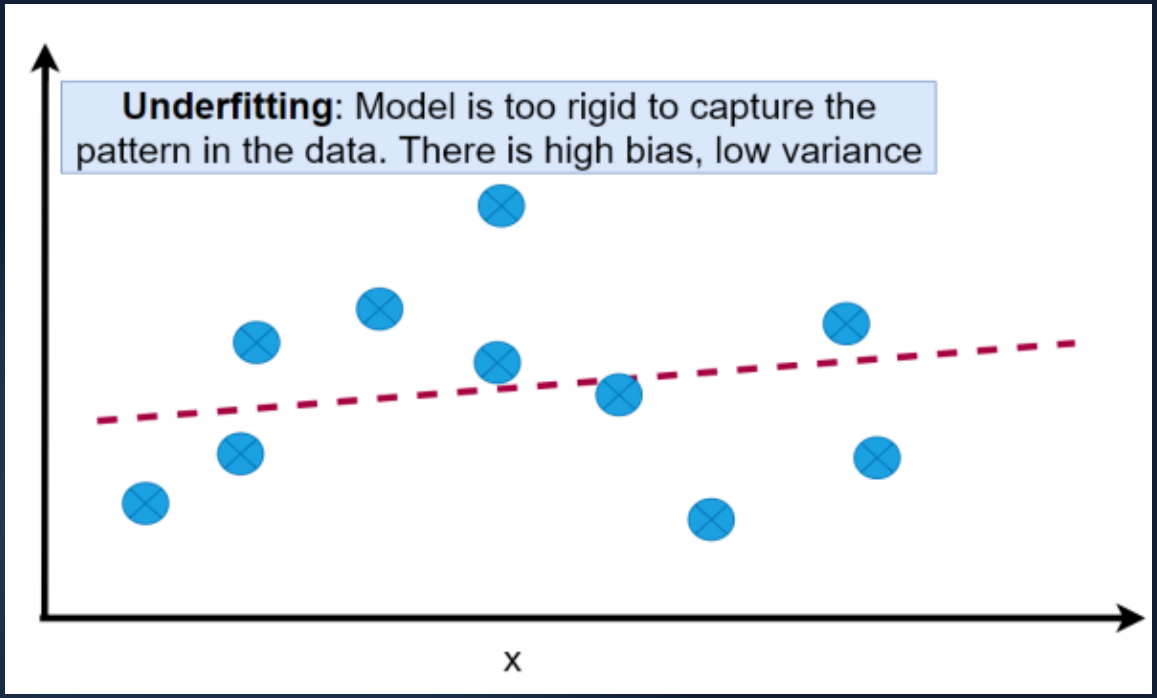
Supervised learning

- What is the difference between regression supervised learning and classification supervised learning?

Model Evaluation

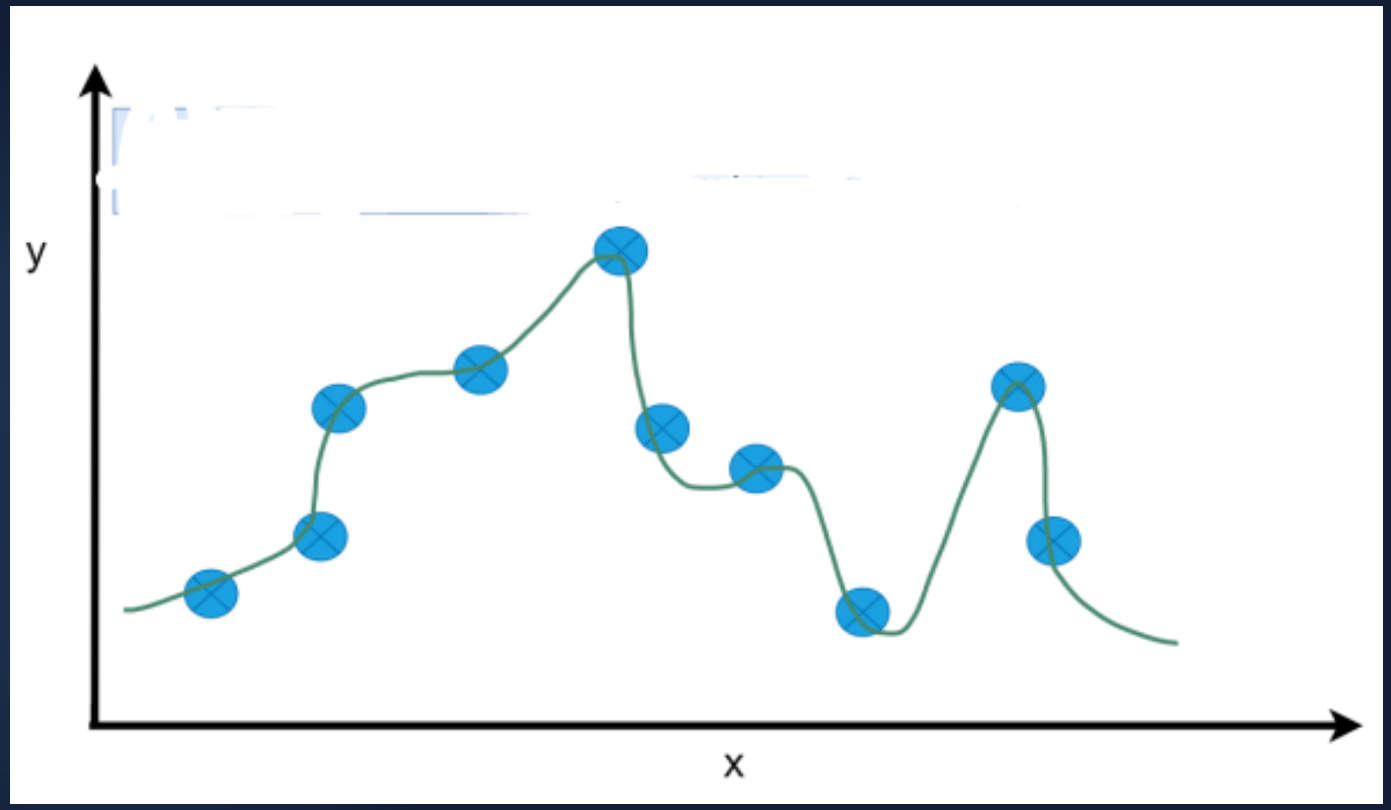


Underfitting



- What is the difference between bias and variance?
- When is a model said to underfit the data?

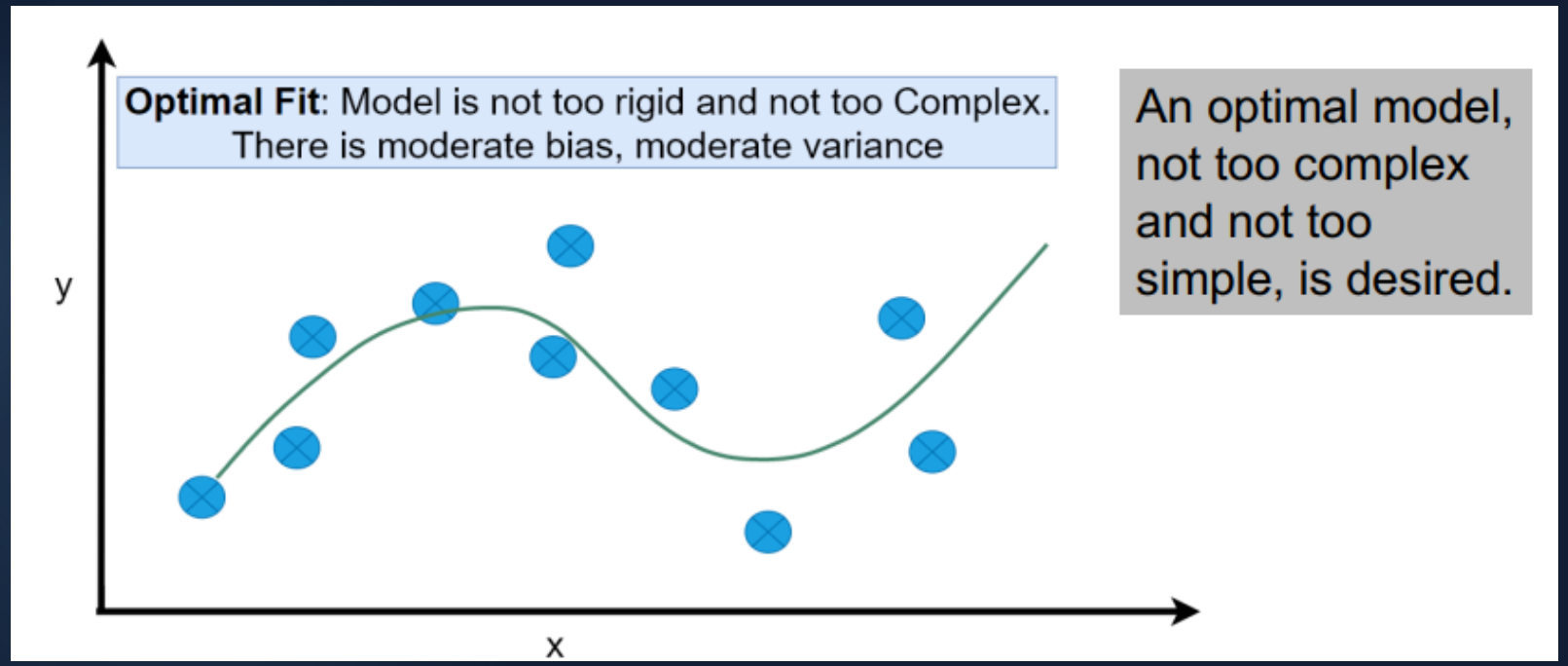
Overfitting



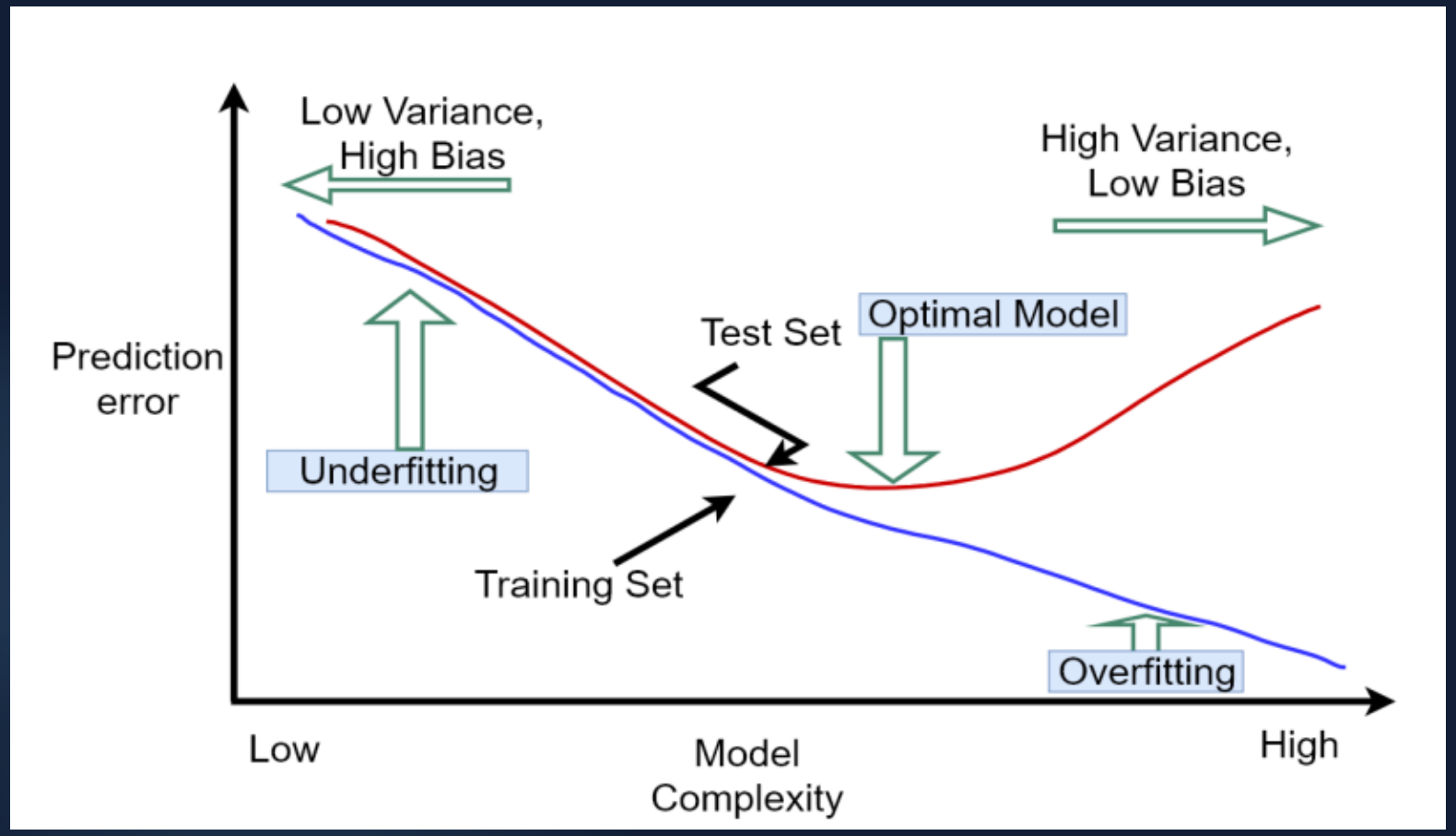
- What are some of the characteristics of overfitting?
- When is a model said to overfit the data?

Optimal fit

- When is a model said to be optimal?
- What is the Occam's Razor principle?



Bias-Variance Trade off



- When is the model said to be optimal?

Hyper parameters

- What are hyperparameters?
- What are the hyperparameters for these models?

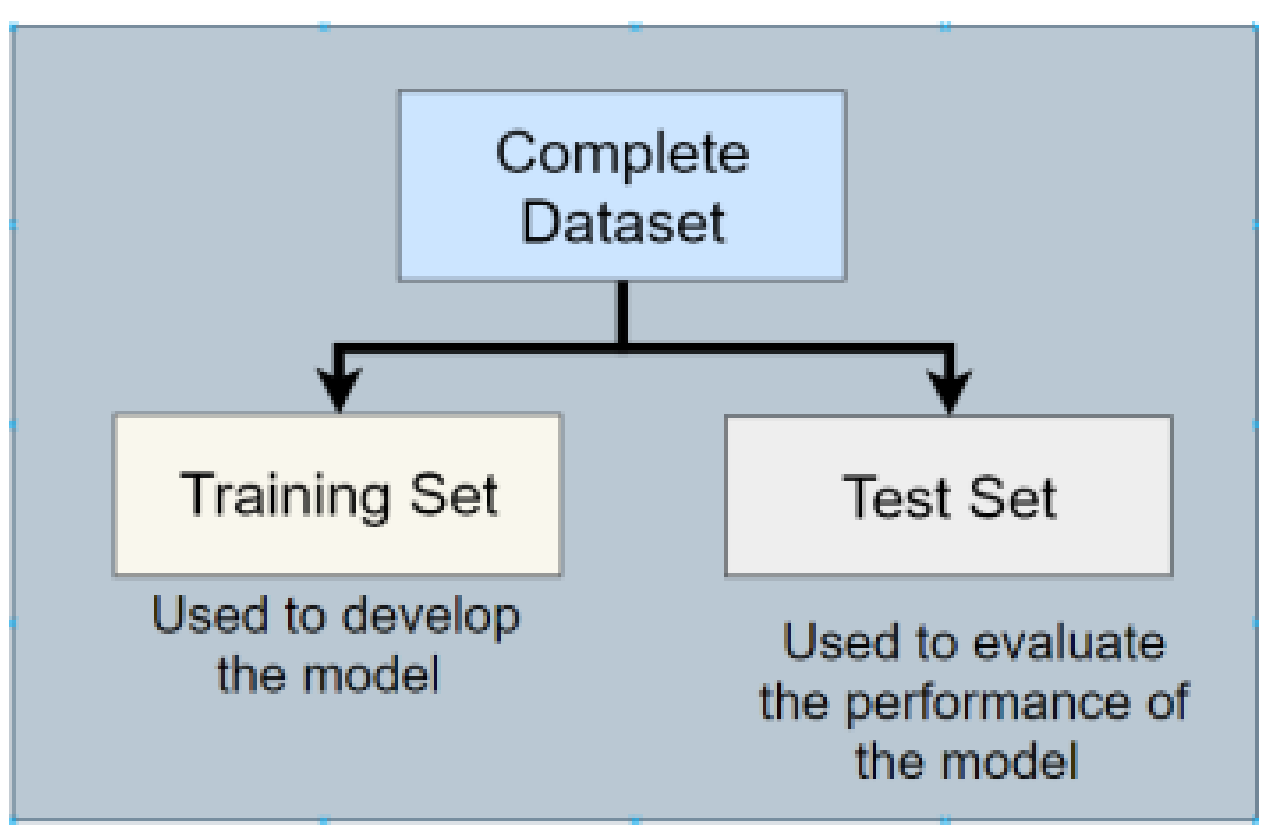
Model
Polynomial regression
Naïve Bayes
Decision Tree
K-Nearest Neighbor

Model Selection and Comparison

- How is model selection different from model comparison?
- What is the purpose of model selection?
- What is the purpose of model comparison?

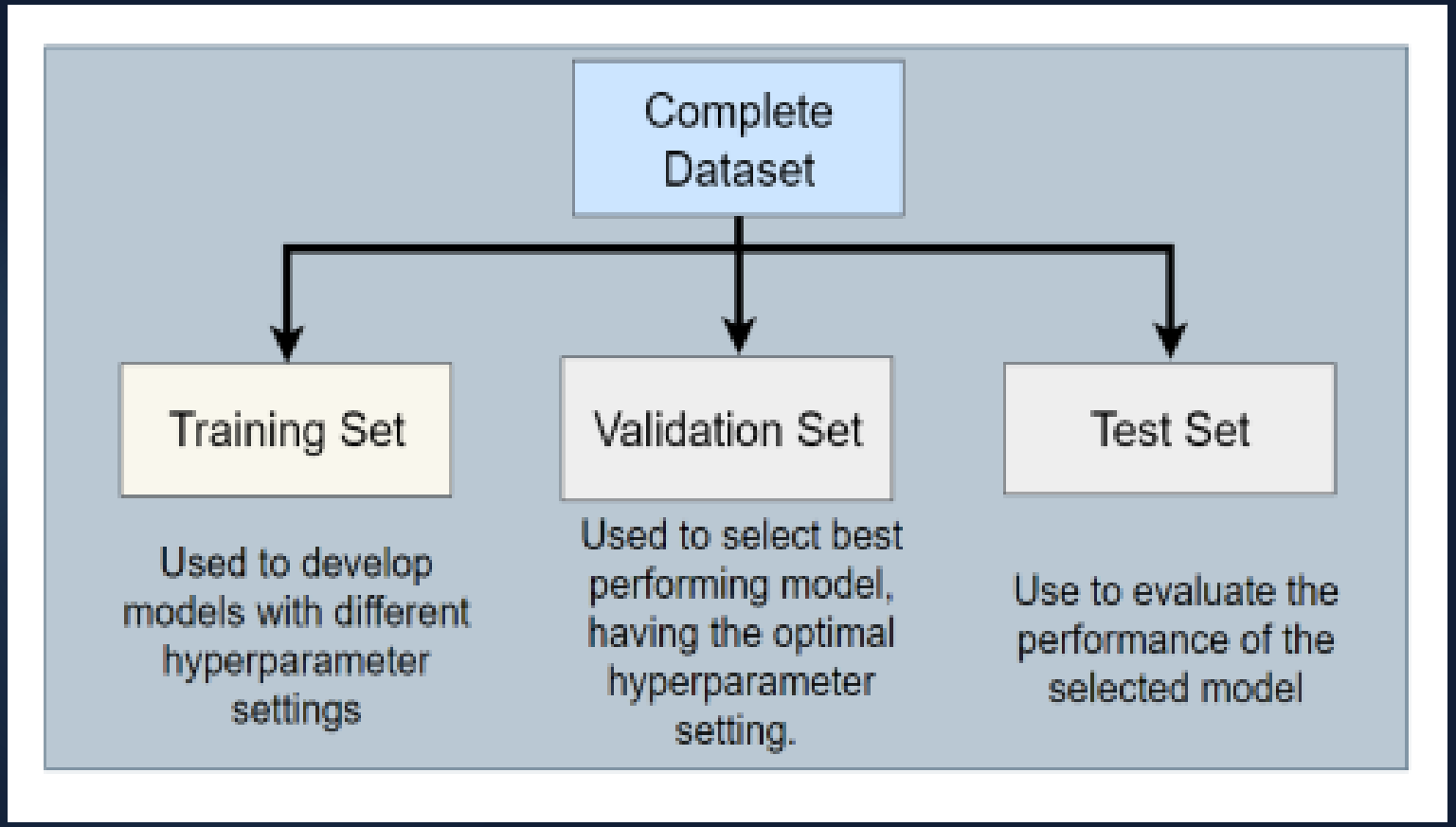
Model Evaluation

Two-way holdout



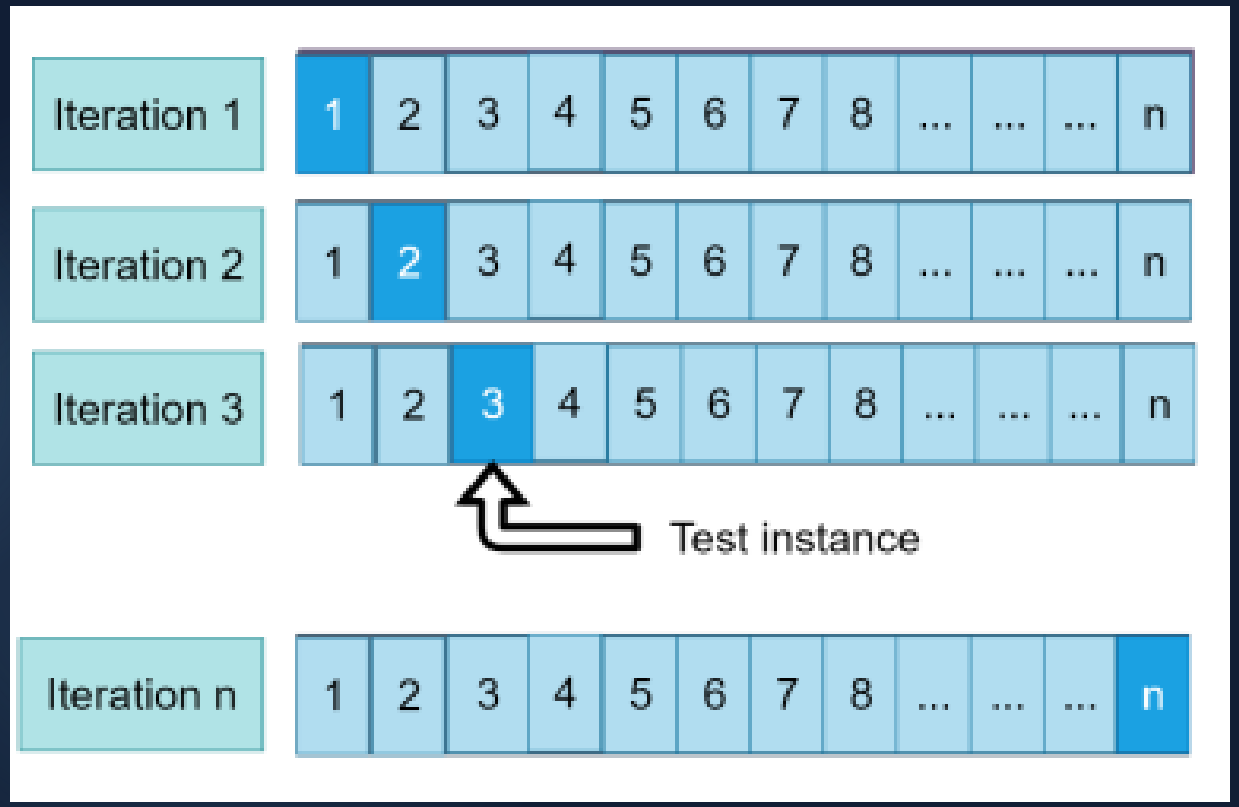
Model Evaluation

Three-way holdout validation



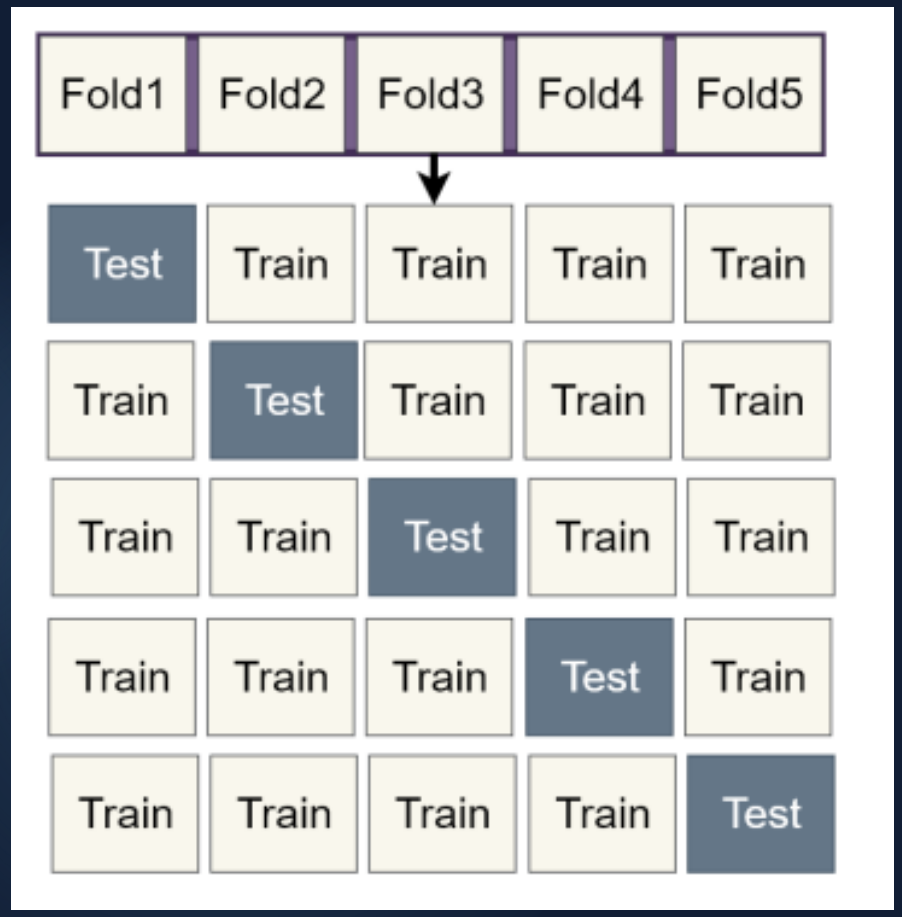
Model Evaluation

Leave-one-out cross validation



Model Evaluation

K-fold cross validation



Group Activity

- Discuss at least 5 evaluation metric or performance measures used to evaluate the performance of a regression model model.
- Discuss at least 5 evaluation metrics or performance measures used to evaluate the performance of a classification model.

Define these measures whether by describing them or mathematically (instead of just listing them).
How are these measures used?