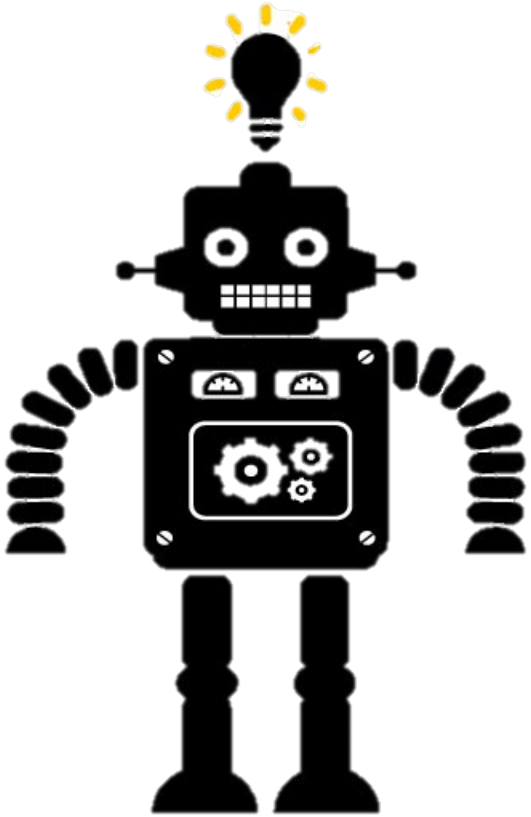
METIS

Review of Machine Learning

What is Machine Learning?

Machine learning allows computers to learn and infer from data.



Types of Machine Learning

Supervised — data points have known outcome

Unsupervised — data points have unknown outcome

Types of Supervised Learning

Regression

data points have continuous outcome

Classification

data points have categorical outcome

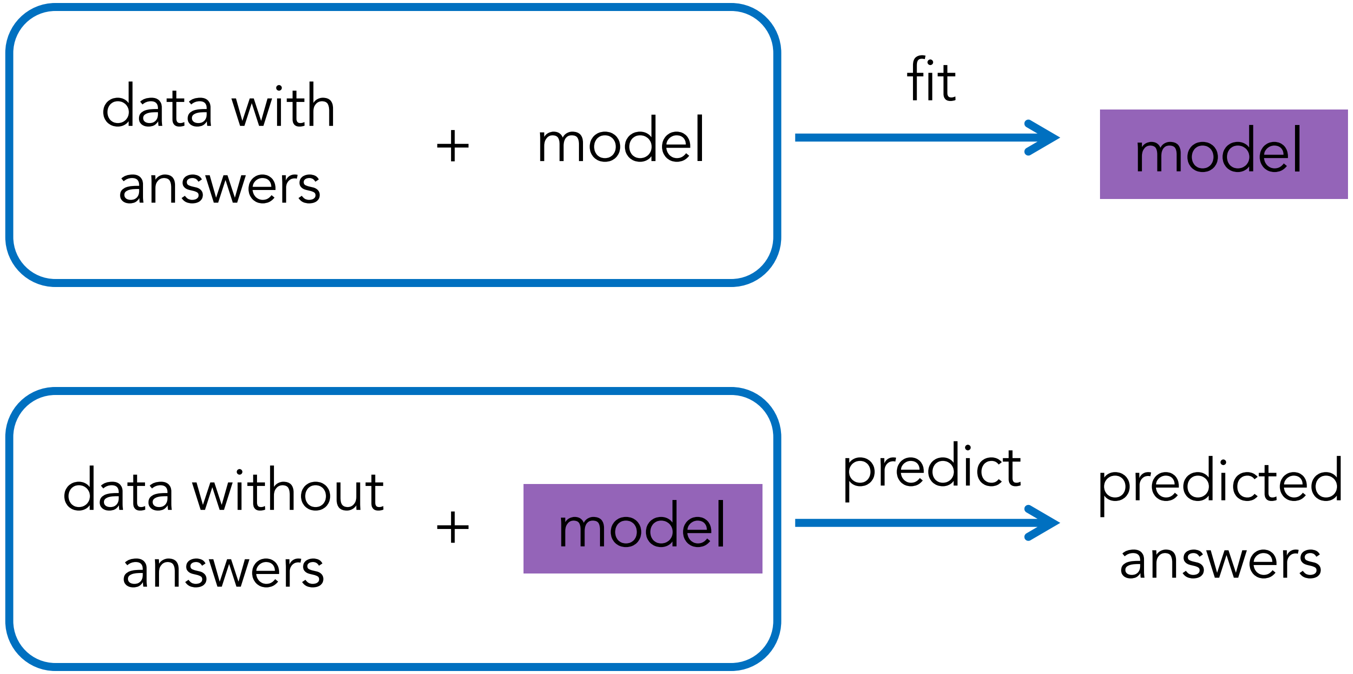
Machine Learning Vocabulary

- **Target:** predicted category or value of the data (column to predict)
- **Features**: properties of the data used for prediction (non-target columns)
- **Example:** a single data point within the data (one row)
- **Label:** the target value for a single data point

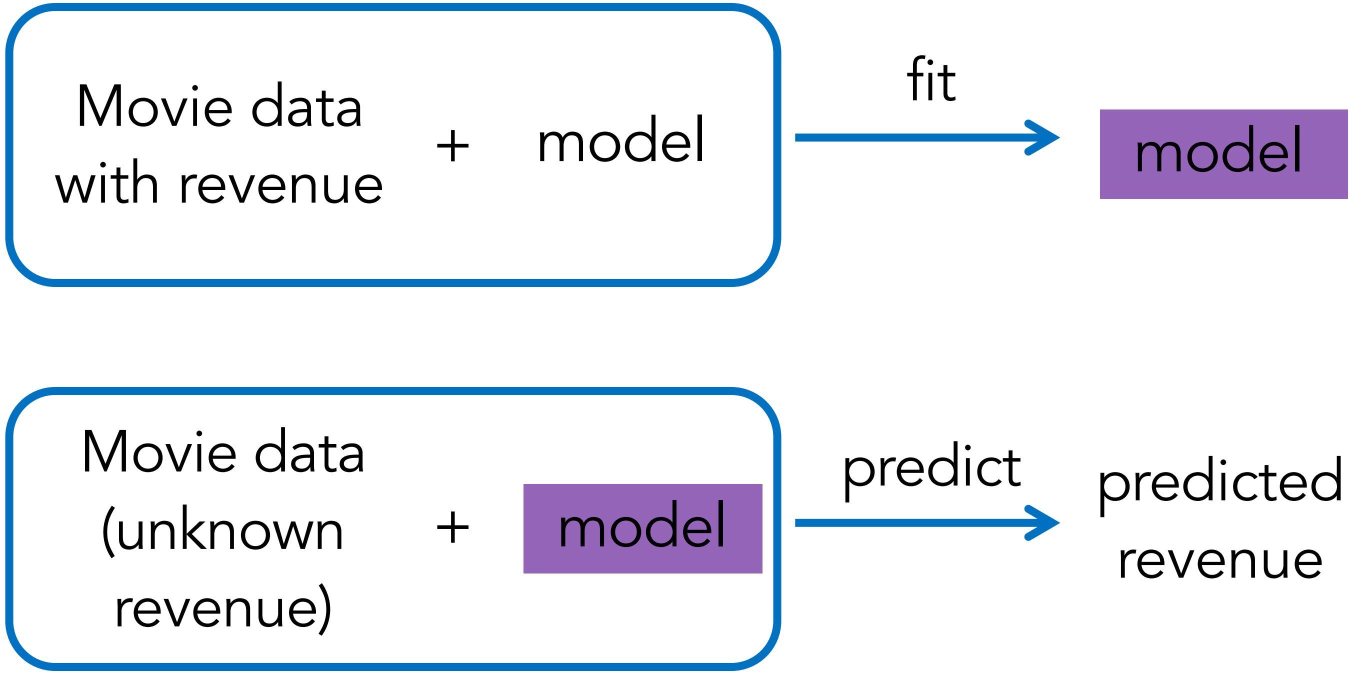
Machine Learning Vocabulary (Synonyms)

- **Target:** Response, Output, Dependent Variable, Labels
- **Features:** Predictors, Input, Independent Variables, Attributes
- **Example:** Observation, Record, Instance, Datapoint, Row
- **Label:** Answer, y-value, Category

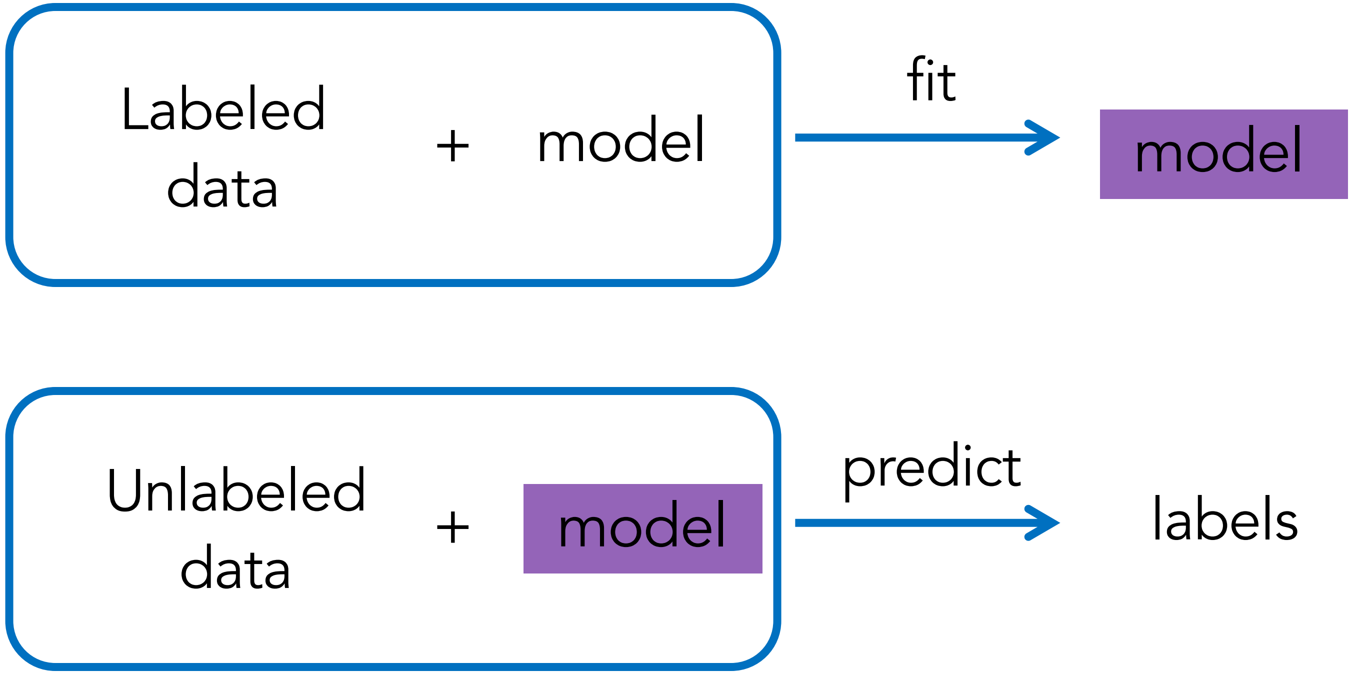
Supervised Learning Overview



Regression: Numeric Answers



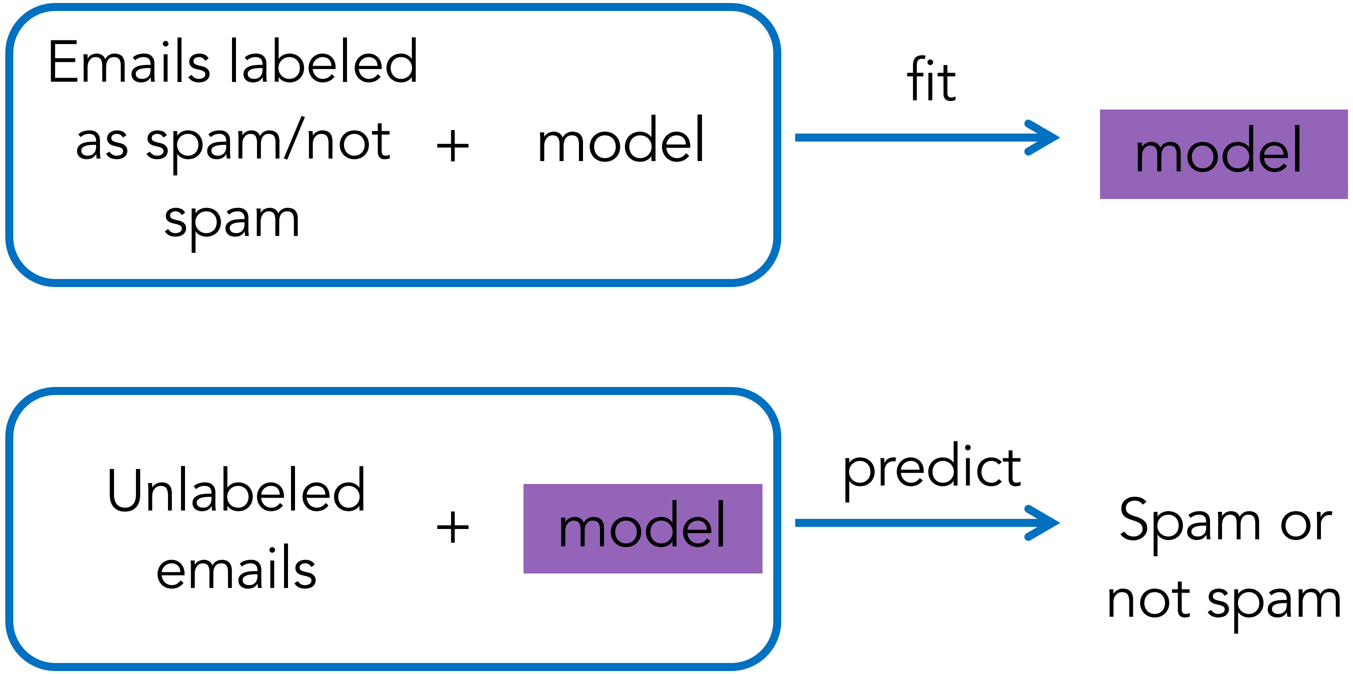
Classification: Categorical Answers



Labeled data + model $\xrightarrow{\text{fit}}$ model

Unlabeled data + model $\xrightarrow{\text{predict}}$ labels

Classification: Categorical Answers



Emails labeled as spam/not spam + model $\xrightarrow{\text{fit}}$ model

Unlabeled emails + model $\xrightarrow{\text{predict}}$ Spam or not spam

Two Types of Classification Predictions

- **Hard Prediction:** Predict a single category for each instance.
- **Probability Prediction:** Assign a probability distribution across the classes to each instance.

Metrics for Classification

- **Hard Prediction:** Accuracy, Precision, Recall (Sensitivity), Specificity, F1 Score
- **Probability Prediction:** Log-loss (aka Cross-Entropy), Brier Score, AUC (ROC), Precision-Recall Curves

Metrics for Regression

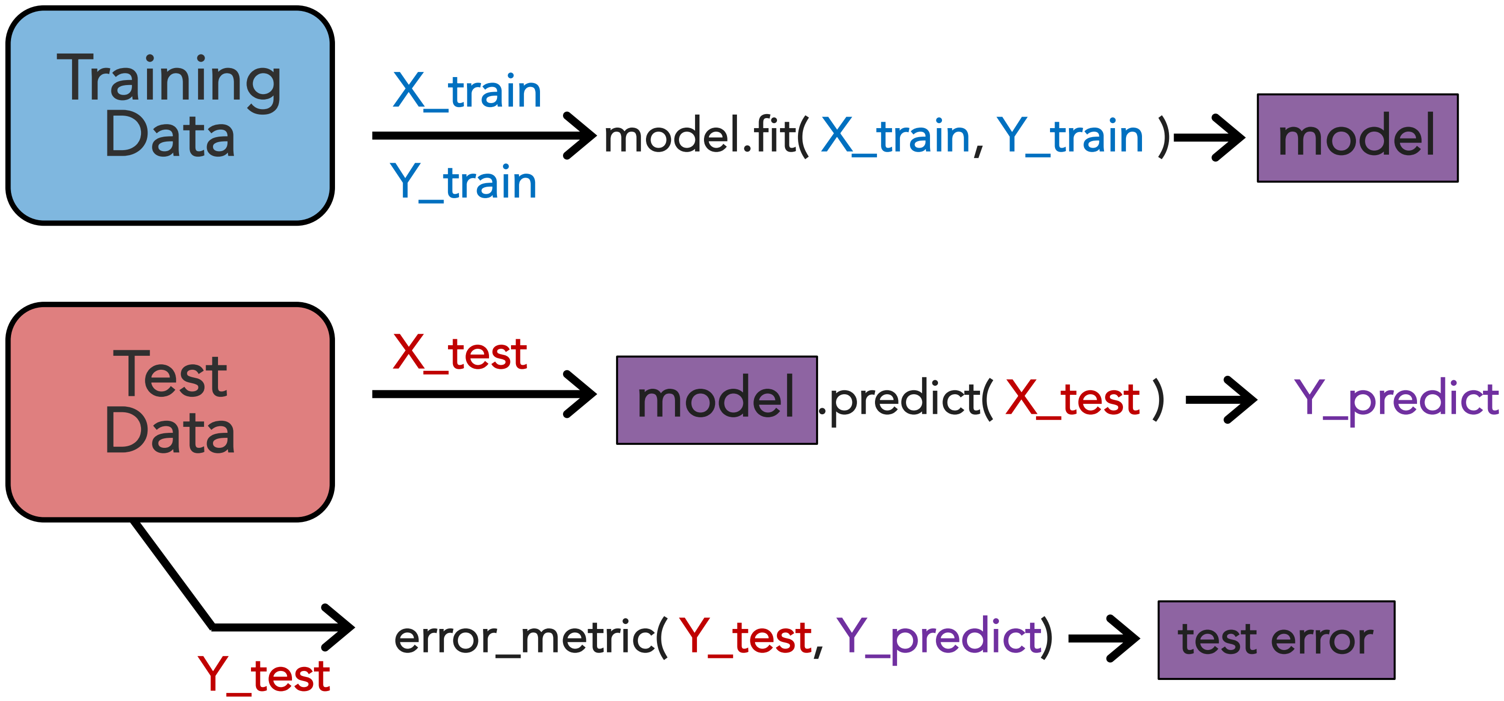
- Root Mean Square Error (RMSE)

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

- Mean Absolute Deviation

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Fitting Training and Test Data



Training Data

X_train

Y_train

model.fit(X_train, Y_train) → model

Test Data

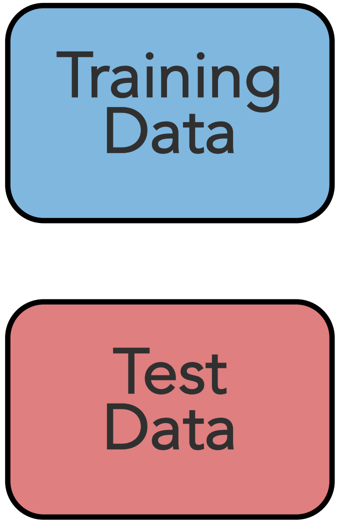
X_test

model.predict(X_test) → Y_predict

Y_test

error_metric(Y_test, Y_predict) → test error

Using Training and Test Data



Training Data

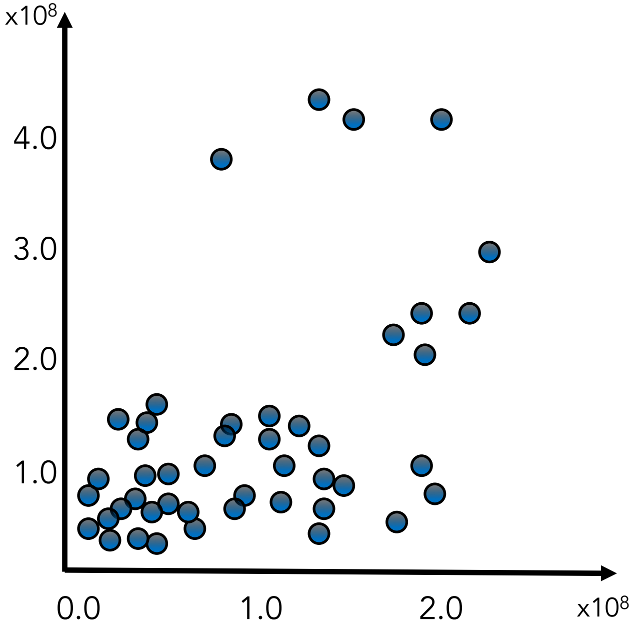
fit the model

Test Data

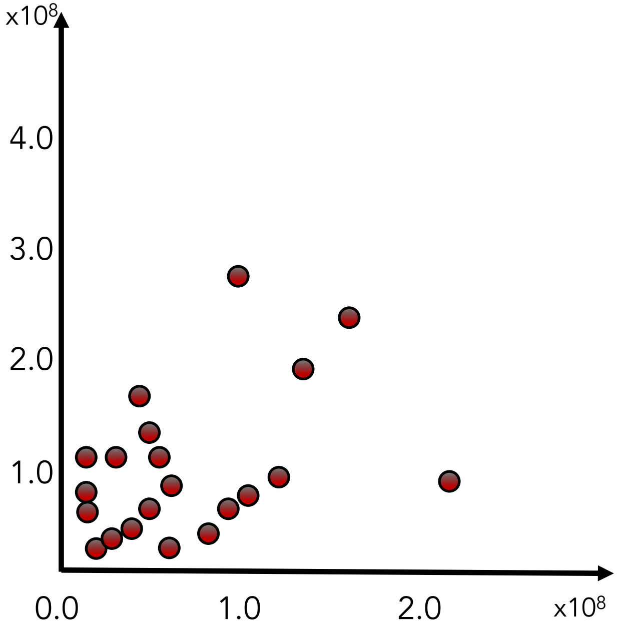
measure performance

- predict label with model
- compare with actual value
- measure error

Using Training and Test Data

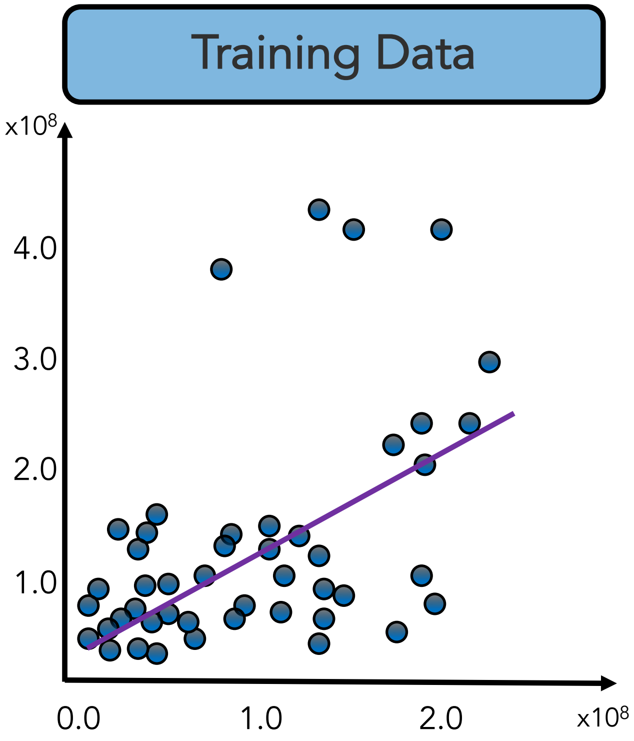


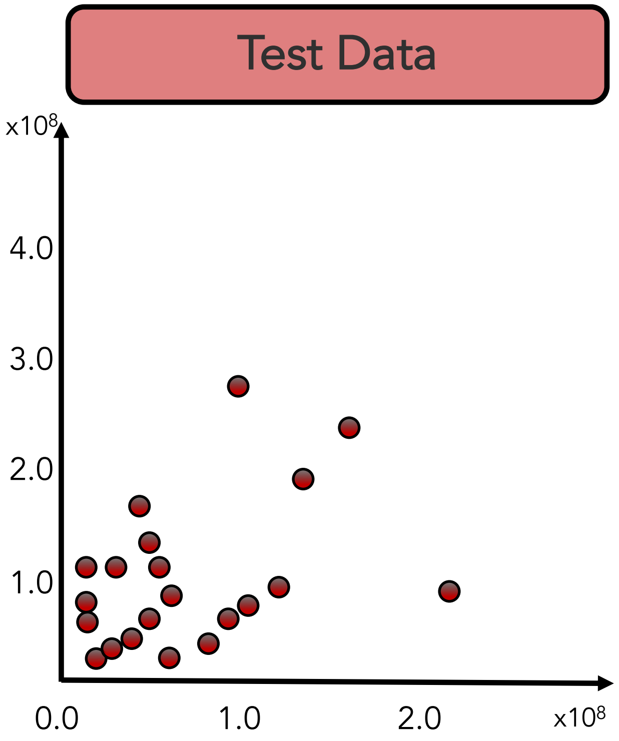
Training Data



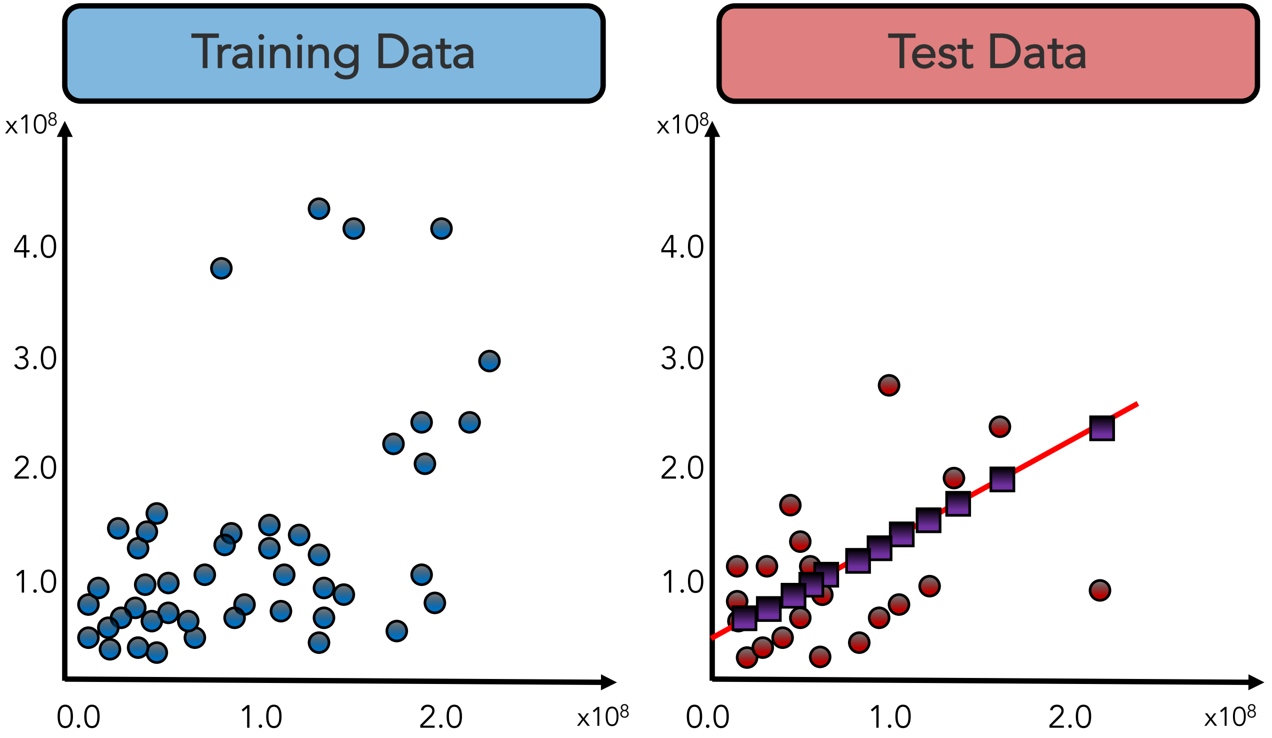
Test Data

Using Training and Test Data



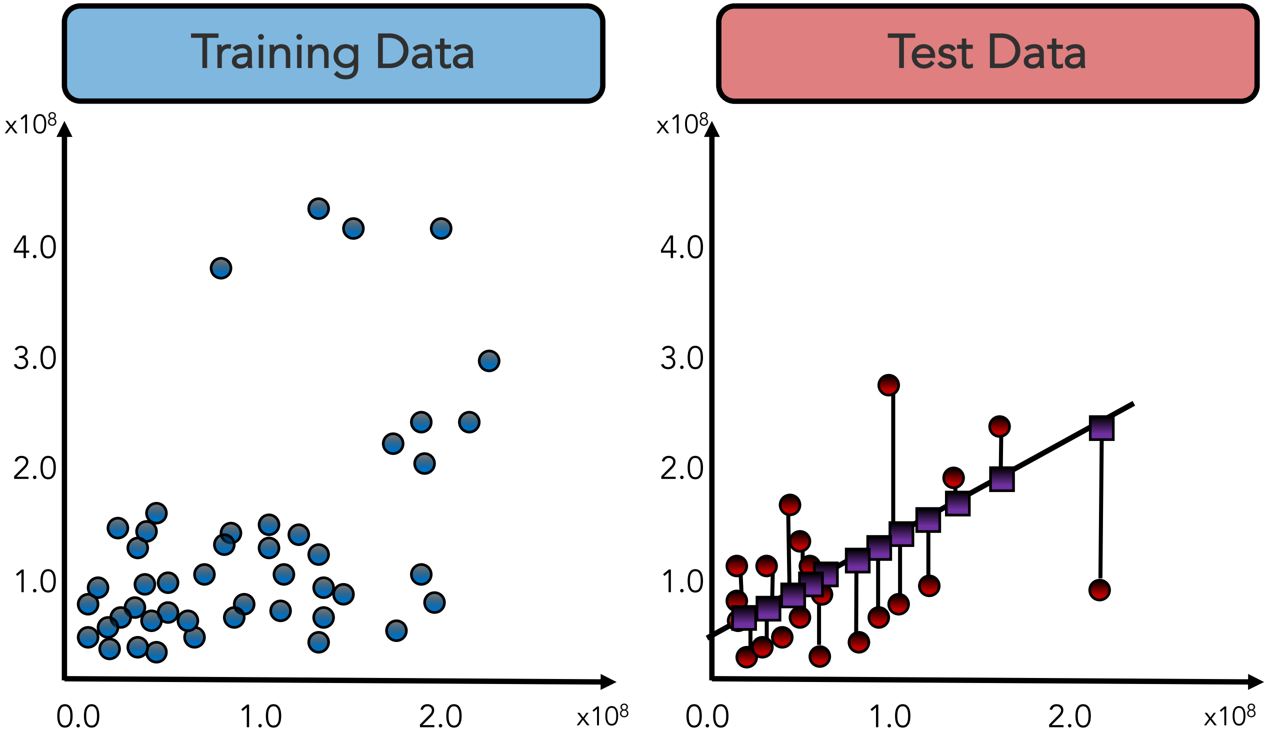


Using Training and Test Data



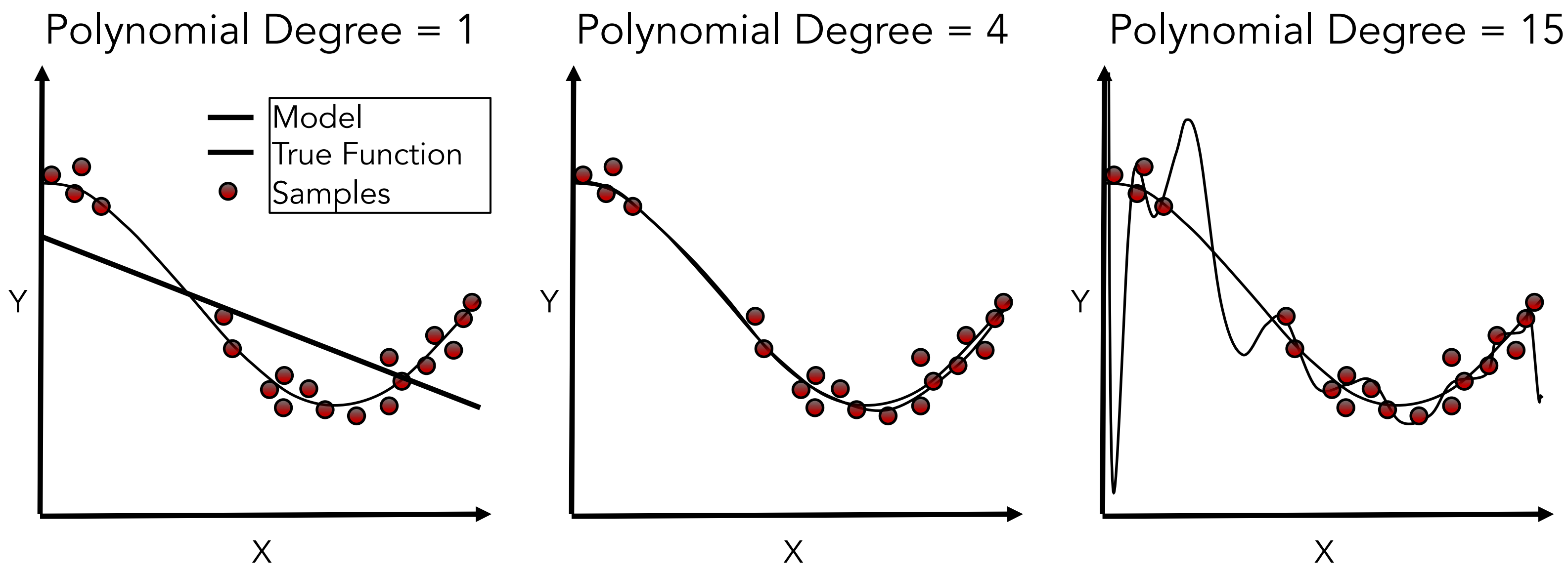
Make predictions

Using Training and Test Data



Measure error

How Well Does the Model Generalize?


Polynomial Degree = 1

Polynomial Degree = 4

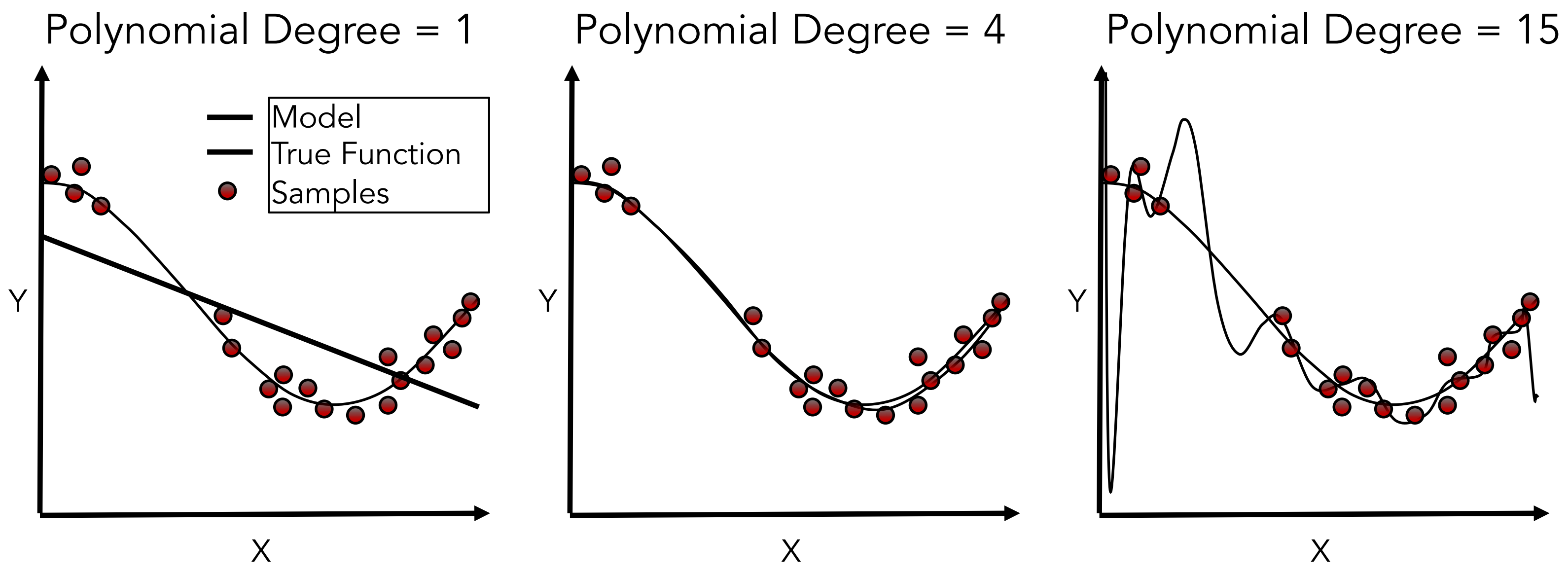
Polynomial Degree = 15

Poor on Training Set
Poor at Predicting

Just Right

Very Good on Training Set
Poor at Predicting

Underfitting vs Overfitting

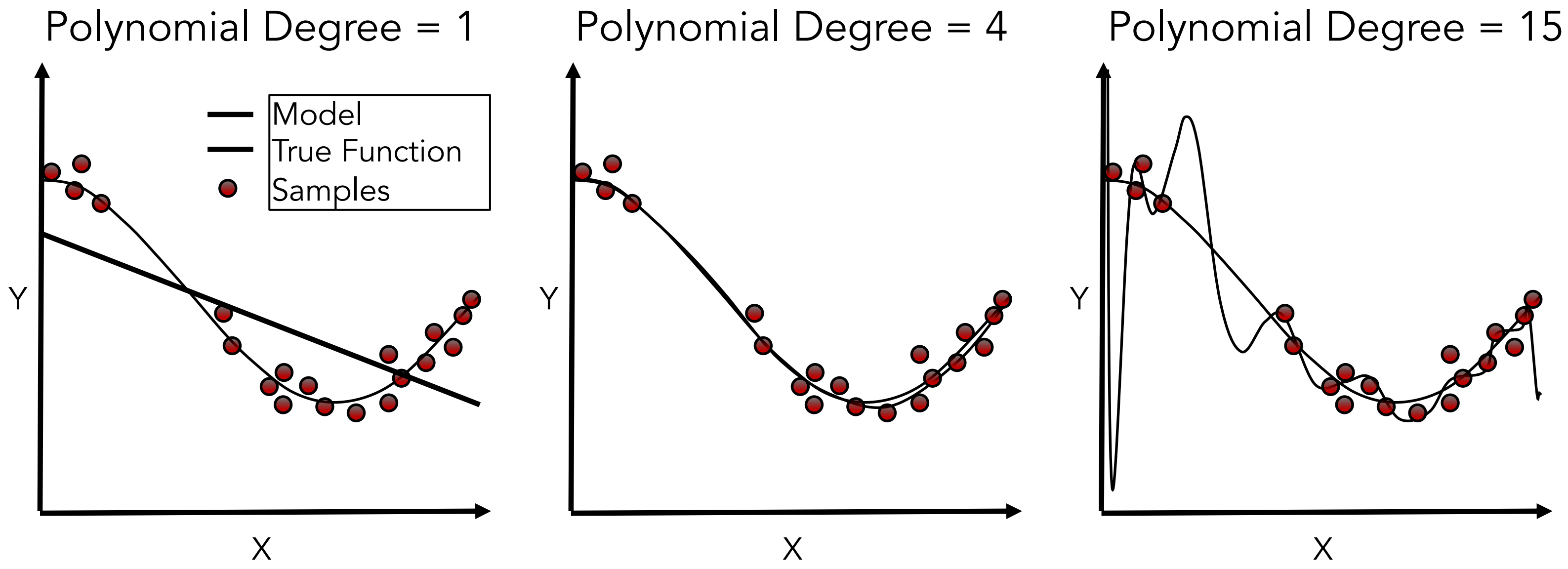


Underfitting

Just Right

Overfitting

Bias/Variance Tradeoff



High Bias
Low Variance

Just Right

Low Bias
High Variance

Questions?

METIS