INTRODUCTION TO MACHINE LEARNING
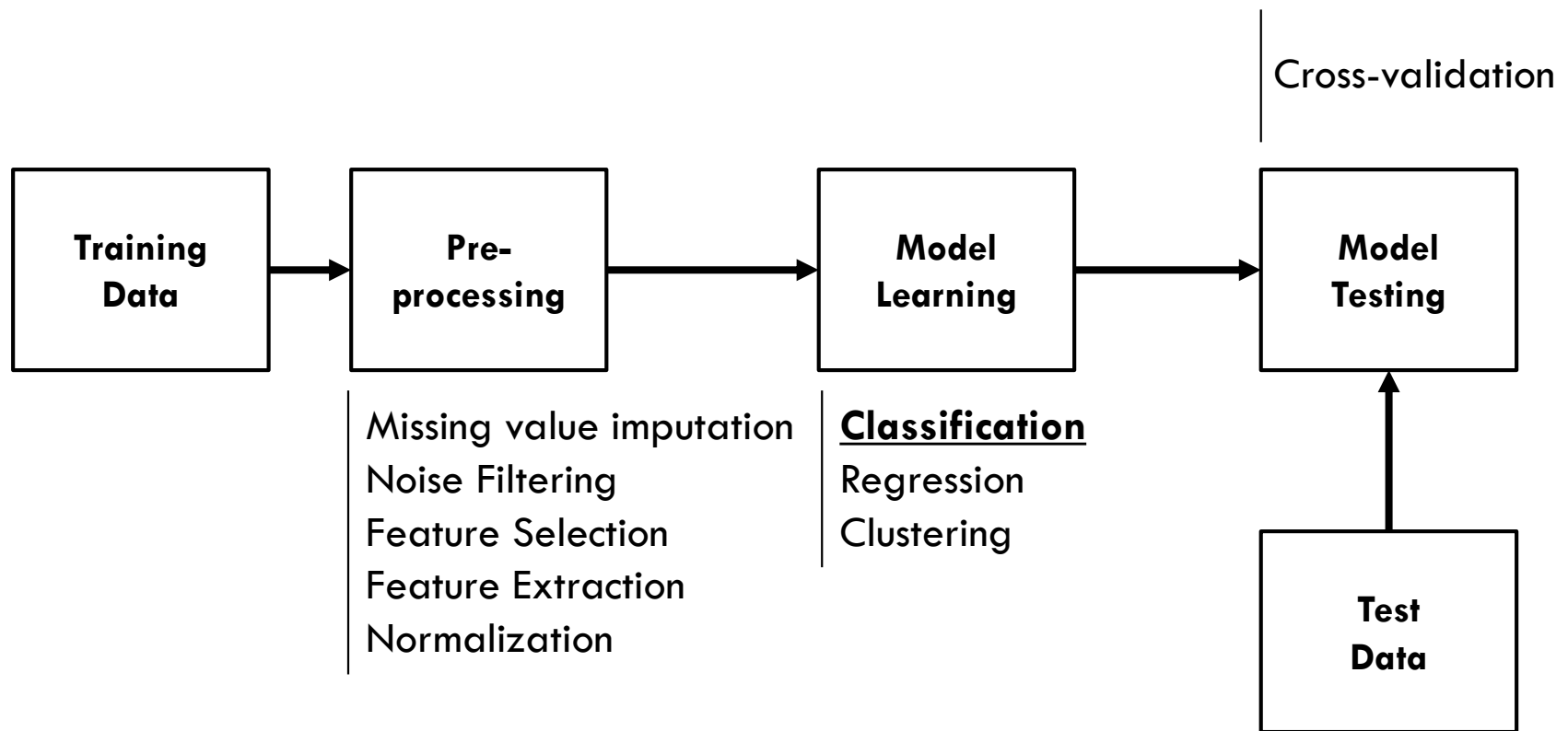
CLASSIFICATION PROBLEMS

Mingon Kang, PhD
Department of Computer Science @ UNLV

Learning Process in ML

Training Data	→	Pre-processing	→	Model Learning	→	Model Testing

Missing value imputation
Noise Filtering
Feature Selection
Feature Extraction
Normalization

Classification
Regression
Clustering

Cross-validation

Test Data

Ref: https://www.slideshare.net/liorrokach/introduction-to-machine-learning-13809045

Supervised learning

Data: $D = \{D_1, D_2, \ldots, D_n\}$ a set of *n* samples
   where $D_i = < \mathbf{X_i}, y_i >$
   $\mathbf{X_i}$ is a input matrix and y_i is a desired output

Objective: learning the mapping $f: X \rightarrow y$
   subject to $y_i \approx f(\mathbf{X_i})$ for all *i* = 1,...,n

<u>Classification</u>: y is discrete

<u>Regression</u>: y is continuous

Classification

- "General process related to categorization, the process in which ideas and objects are recognized, differentiated, and understood." Classification in Wikipedia

- "The problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known." Statistical classification in Wikipedia

Types of classification problems

- Binary classification
  - Only two classes, but one sample has one label
- Multi-class classification
  - Multiple classes, but one sample has one label
- Multi-label classification
  - One sample can have multiple class labels
- Image segmentation
  - Traditionally, clustering problem
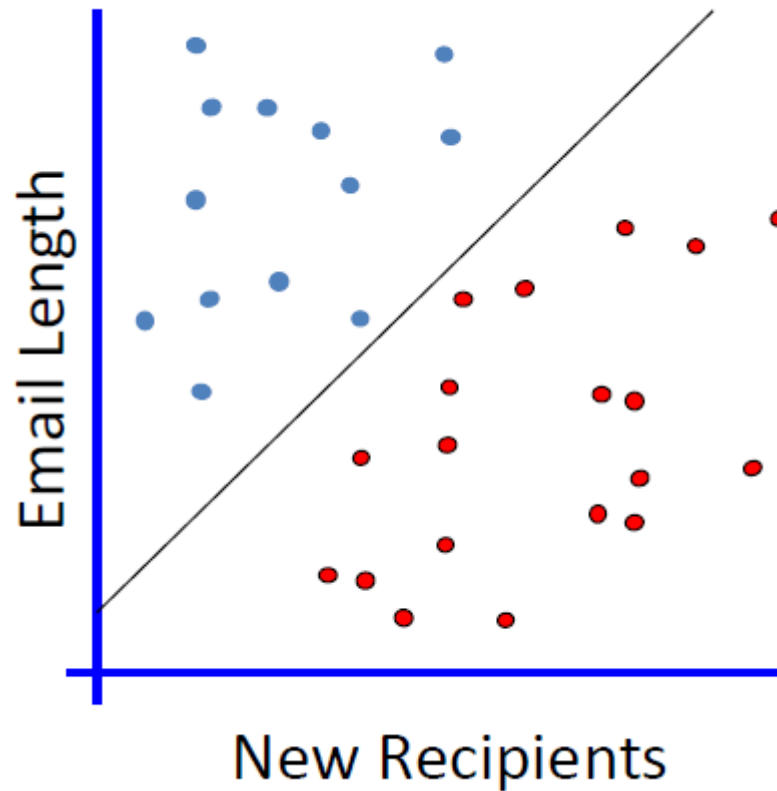 - Recently, pixel-based classification problem

Other types?

- Classification with sequential data
  - Spatial/Temporal data
- Weakly supervised data
  - One image may have multiple labels, but we don't know where the objects are

Discussion

- Examples of Machine Learning problems for the other types?
  - What applications?

Linear Classification
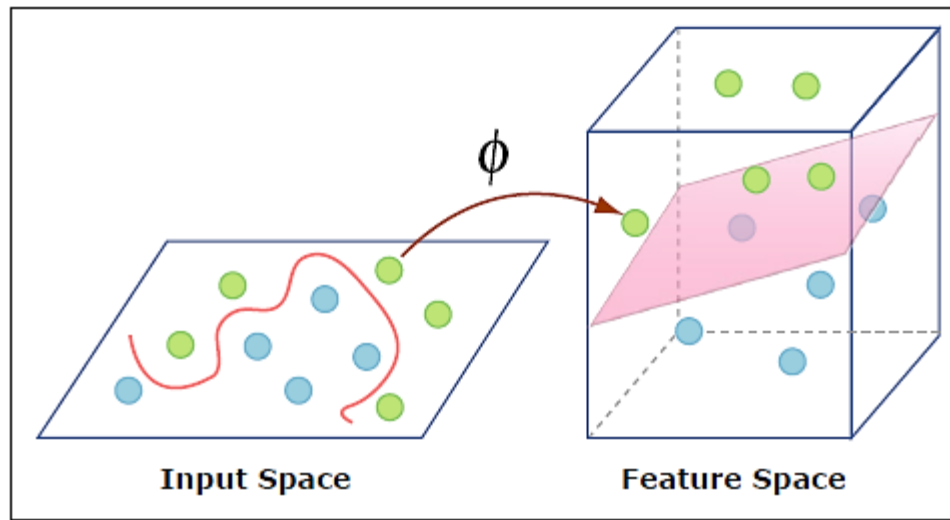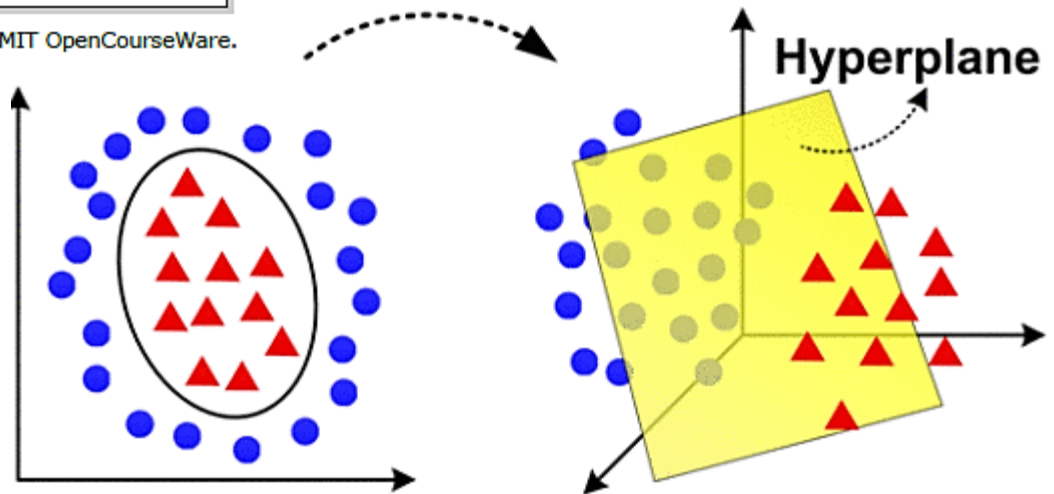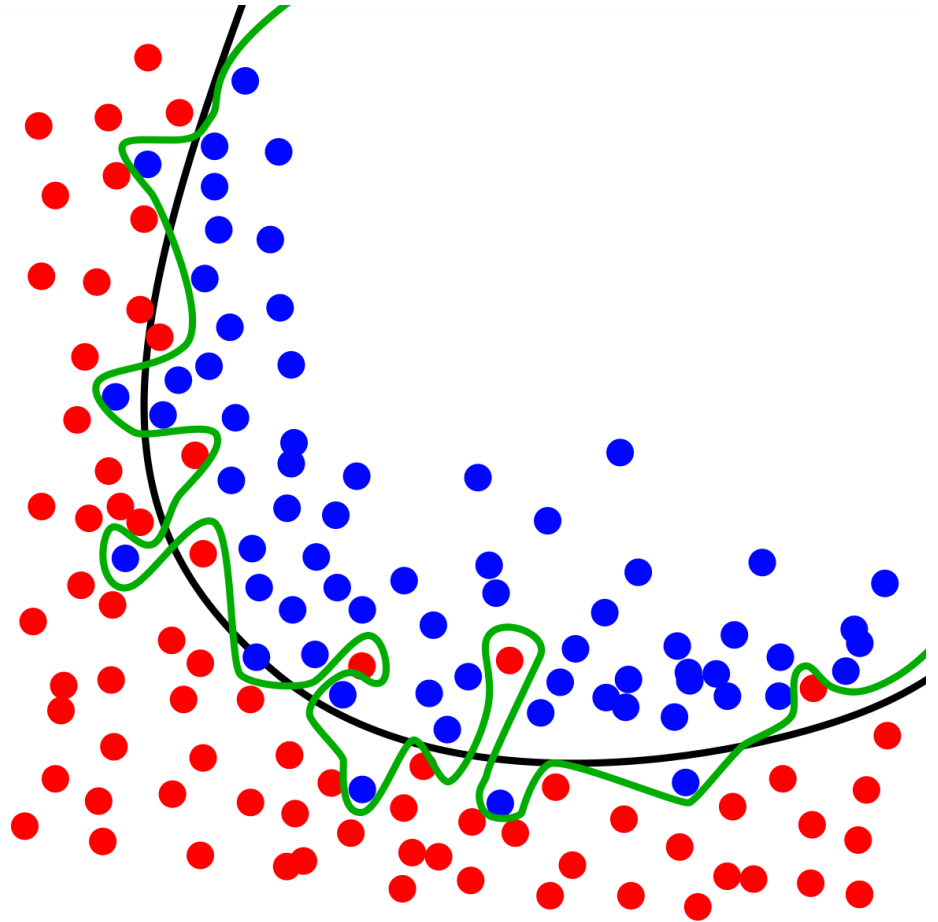
Non-Linear Classification
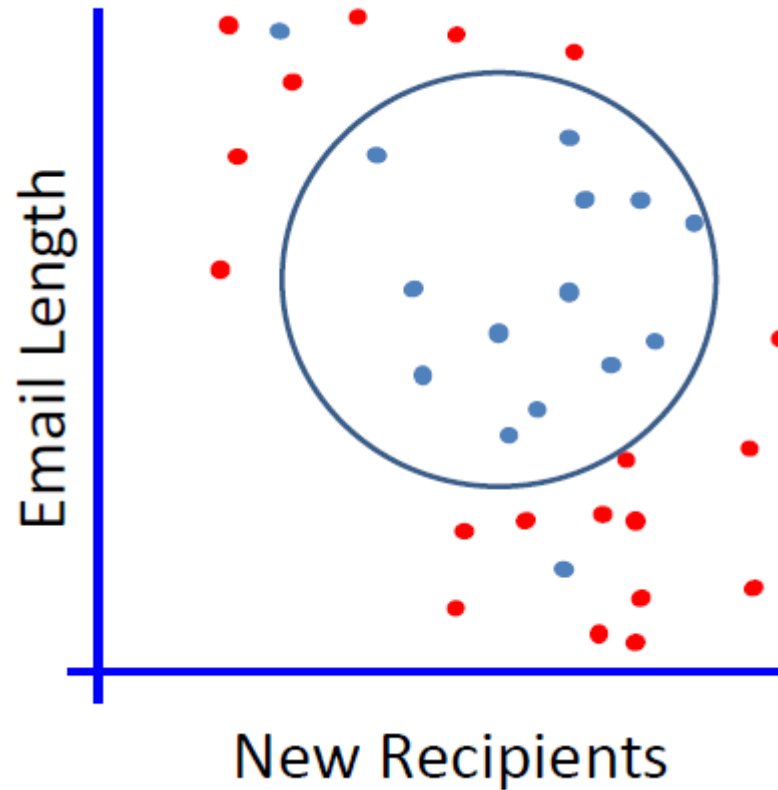
Kernel Trick

Image by MIT OpenCourseWare.

Ref:https://ieeexplore.ieee.org/abstract/document/7341753
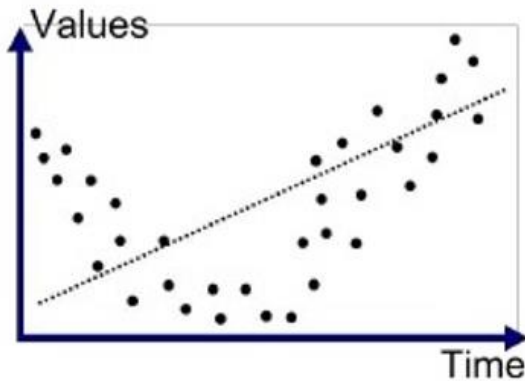
Overfitting

- □ Production of an analysis that corresponds too closely or exactly to a particular set of data
- □ Fail to fit additional data or predict future observations reliably

Ref: https://en.wikipedia.org/wiki/Overfitting

Underfitting

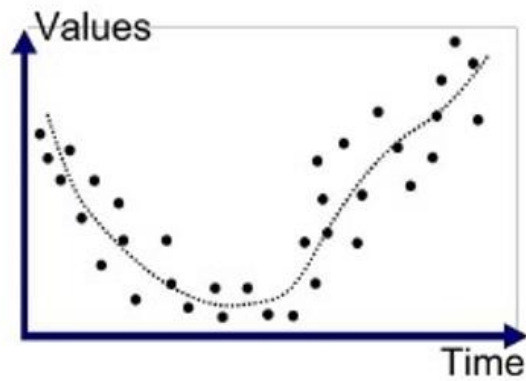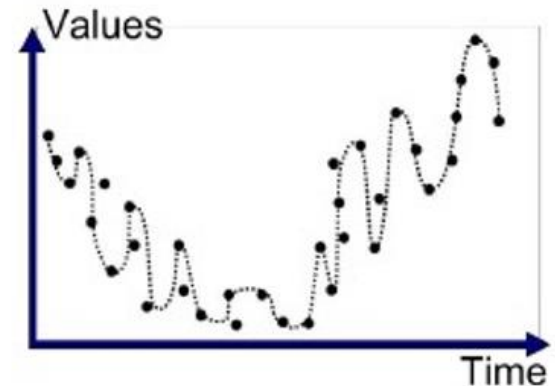□ Underfitting occurs when a model is too simple

Underfitted   Good Fit/Robust   Overfitted