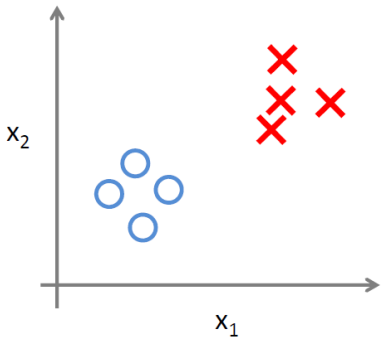
Learning

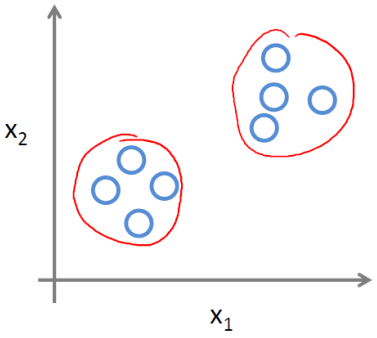
Classification

Supervised Learning



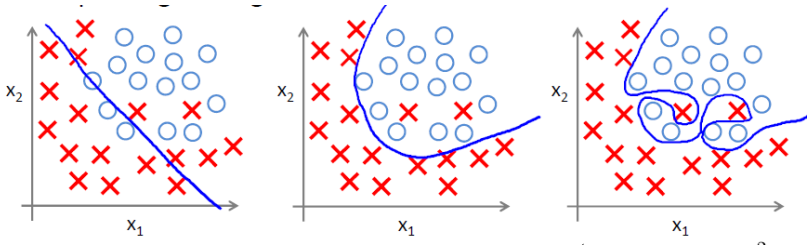
Clustering

Unsupervised Learning



Data Models

Fitting the data with a model



Under fitting

Fitting

Over Fitting

Over fitting model

Over fittintg: If you have too many parameters (features) , your model may be learned to fit the training set very well, with almost 0 error. But fail to generalize to new examples.

Preventing over fitting

More ways:

1- Reduce the number of features (attributes)

- Manually, check which feature to keep
- Model selection (Rough Set –Attribute reduction algorithms)

2- Regularization

- Keep all features but reduce their effect (different methods)

Rough Set Theory

- Can be used for feature selection, feature extraction, data reduction, and others
- Identifies partial or total dependencies in data, eliminates redundant data
- Simple Idea:

  If the number of equivalent types(element sets) derived from attribute set A is the same as that derived from A – ai, then attribute ai is regarded as redundant.

Rough Set Theory

- The reduct: the set of attributes after removing unwanted ones
- heuristic search is used, Find the best reduct as search problem
- Common Algorithms of Attribute Reduct
  - Discrenibility Matrix
  - Quick Reduct
  - Entropy-Based algorithm