Classification

-Definition:

- Given a training set where data is labeled with a special attribute called a class (a discrete value), we want to find a model for the class attribute as a function of the values of the other attributes

-Goal: use the model on unlabeled data to assign a class as accurately as possible

-Example: tumor size vs malignant

-Methods:

Instance-Based Classifiers:

e.g. Rote-learners, Nearest neighbor;

Decision Trees;

Naive Bayes;

Support Vector Machines;

Neural Networks

-K Nearest Neighbor Classifier

- Aggregation methods: Majority rule/ Weighted majority based on distance ($w = 1/d^2$)
- Scaling issues: Attributes should be scaled to prevent distance measures from being dominated by one attribute.
- Pros:

  -Simple to understand why a certain classification is made

  -Can adapt to new attributes
- Cons:

  -Expensive

  -not applicable to high dimensions