4. Machine Learning Overview

4.3 Common Machine Learning Algorithms

Cristiano Bacelar de Oliveira

Agosto, 2020

Universidade Federal do Ceará

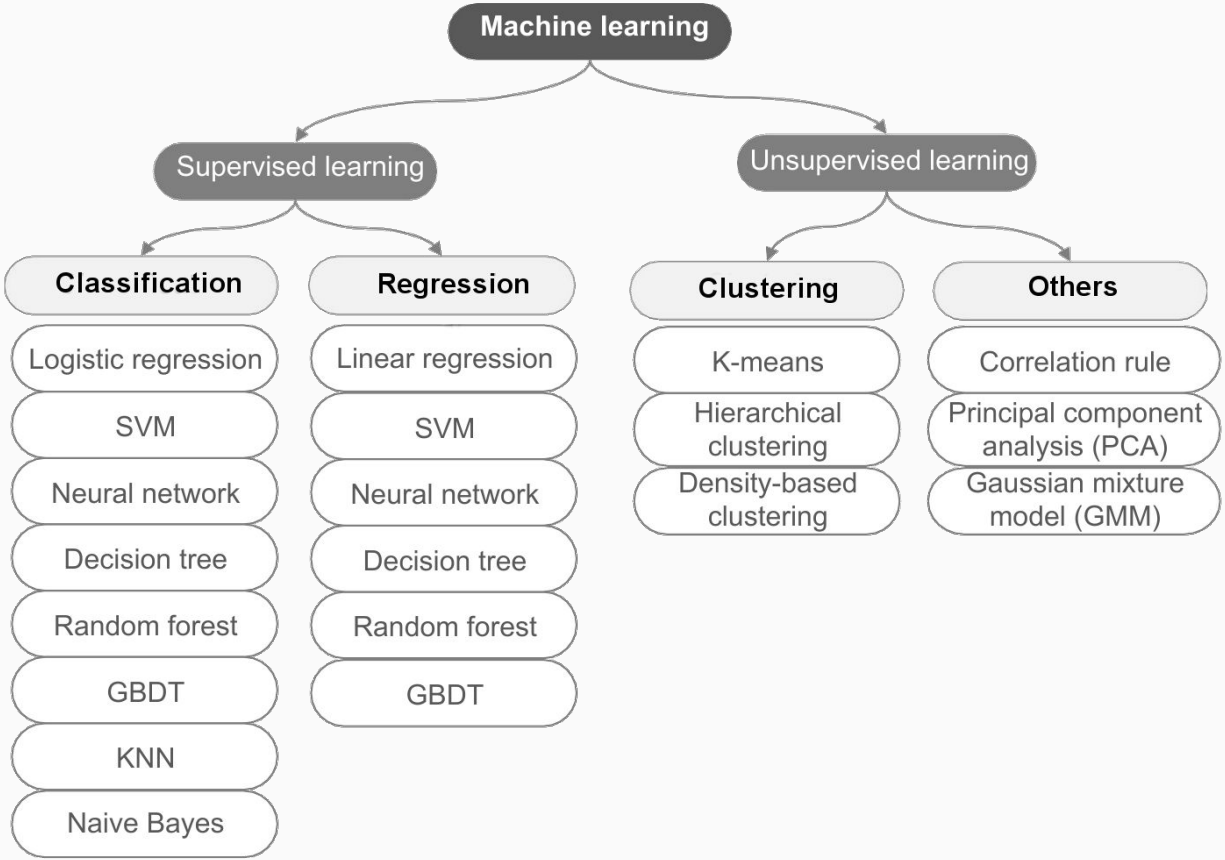
Index

- Machine Learning Algorithms Overview
- Unsupervised Learning Algorithms
- Supervised Learning Algorithms

Disclaimer

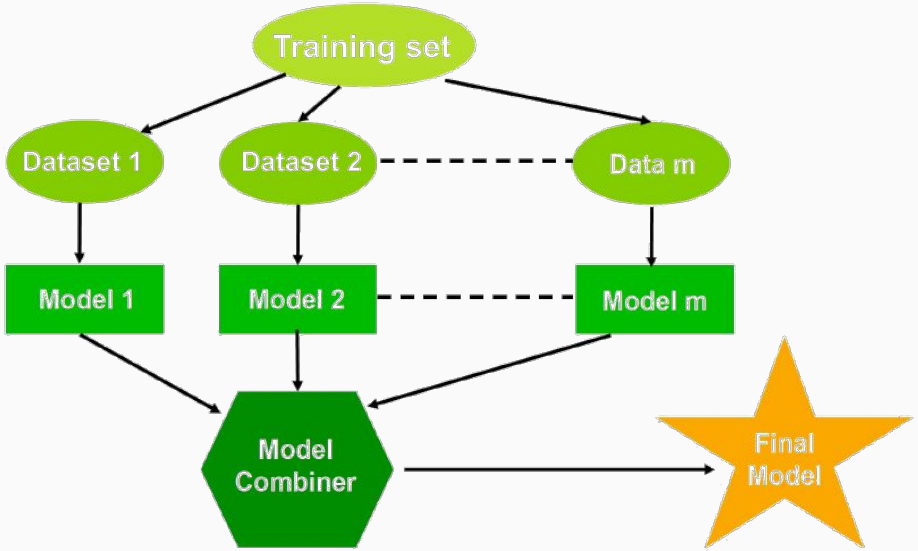
The following content is heavily based on HCIA-AI Course material by Huawei Technologies Co., Ltd., authored by Zhang Luxiao. Distribution is not allowed.

Machine Learning Algorithms Overview



Ensemble Learning

Ensemble learning is a Machine Learning paradigm in which multiple learners are trained and combined to solve the same problem. When multiple learners are used, the integrated generalization capability can be much stronger than that of a single learner.



Unsupervised Learning Algorithms

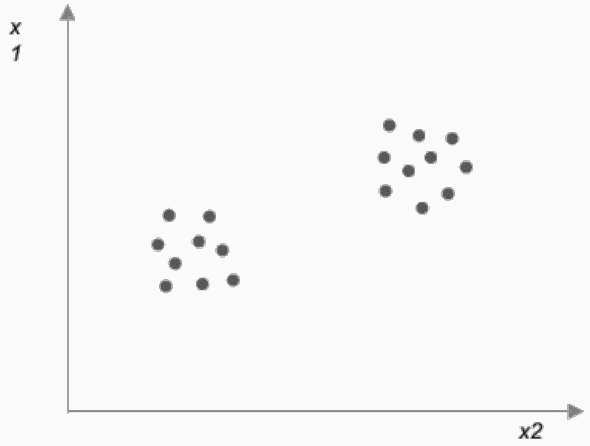
K-Means

K-means algorithms aims to partition the data into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

The k-means algorithm:

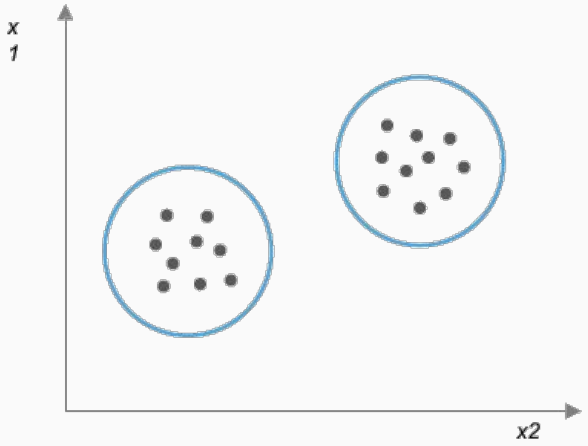
1. Estimate k initial means (centroids)
2. For each data sample:
    a. Compute its distance to all means
    b. Assign the sample to the group (cluster) that has the closest centroid.
3. Recalculate the means for each cluster
4. Repeat 2 and 3 until there are no changes in any mean

K-Means (cont.)



K-means clustering

The data is **not tagged**. K-means clustering can automatically classify datasets.

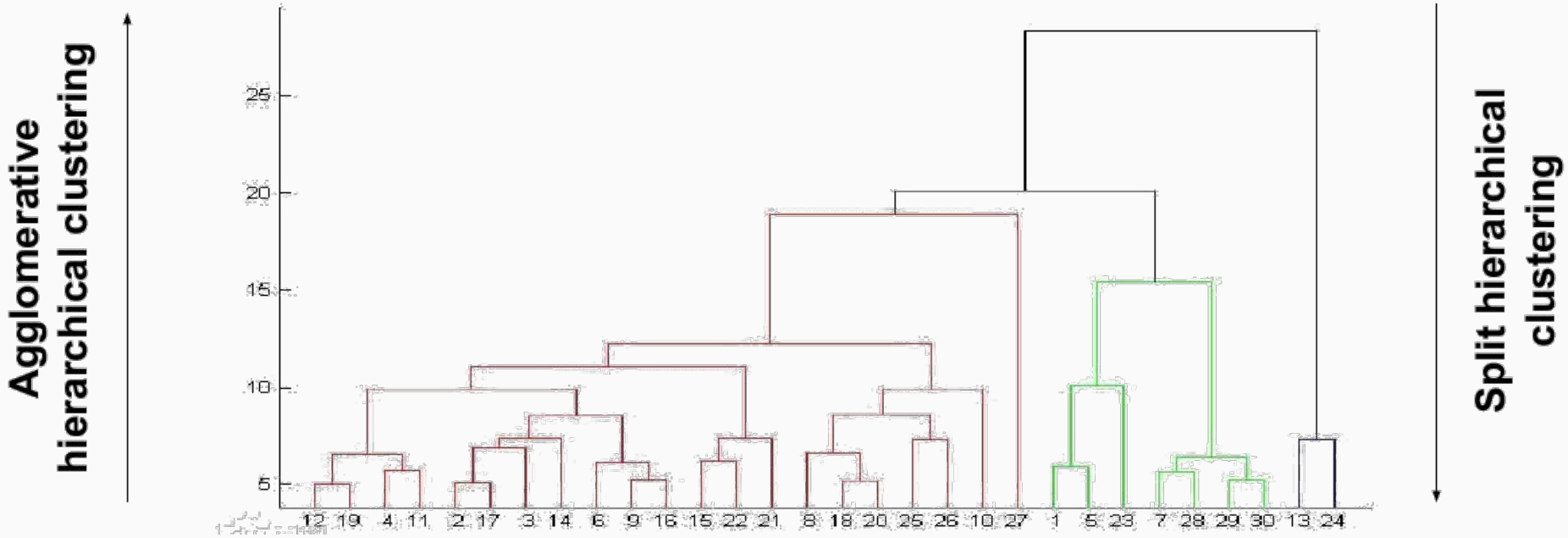


Hierarchical Clustering

Hierarchical clustering divides a dataset at different layers and forms a tree-like clustering structure. The dataset division may use a "bottom-up" aggregation policy, or a "top-down" splitting policy.

The hierarchy of clustering is represented in a tree graph. The root is the unique cluster of all samples, and the leaves are the cluster of only a sample.

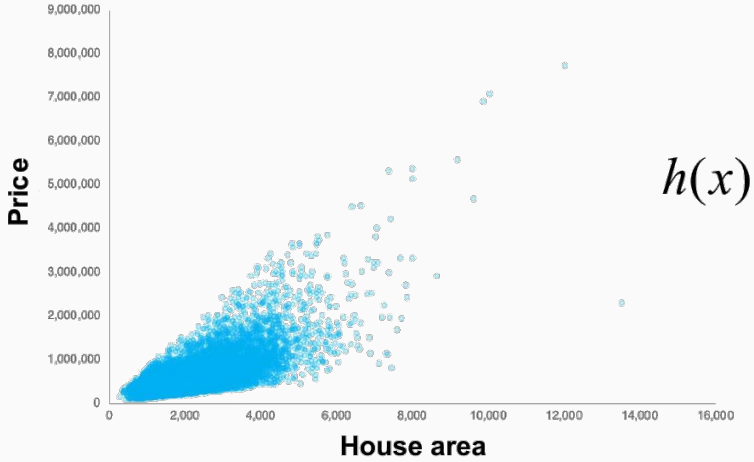
Hierarchical Clustering (cont.)



Supervised Learning Algorithms

Linear Regression

Linear regression is a statistical analysis method to determine the quantitative relationships between two or more variables through regression analysis in mathematical statistics.



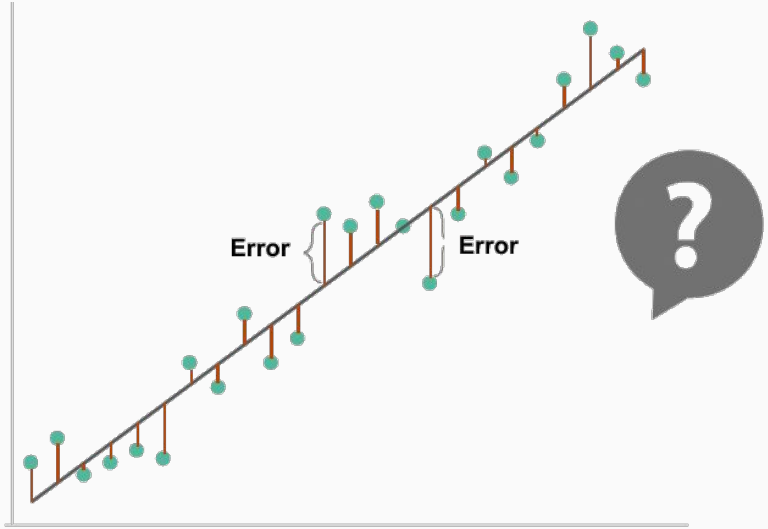
$$h(x) = w_o + w_1 x$$

Which line is the best parameter?

HUAWEI

Linear Regression - Loss Function

To find the optimal parameter, construct a loss function and find the parameter values when the loss function becomes the minimum.



Loss function of linear regression: $J(w)=\frac{1}{2m}\sum\left(h(x)-y\right)^2$

Goal:

$$\arg\min_{w} J(w)=\frac{1}{2m}\sum\left(h(x)-y\right)^2$$

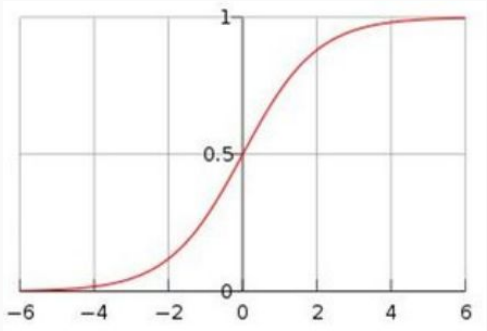
- where, *m* indicates the number of samples,
- *h(x)* indicates the predicted value, and *y* indicates the actual value.

Logistic Regression

The logistic regression model is used to solve classification problems. The model is defined as follows:

$$P(Y = 1|x) = \frac{e^{wx+b}}{1 + e^{wx+b}}$$

$$P(Y = 0|x) = \frac{1}{1 + e^{wx+b}}$$



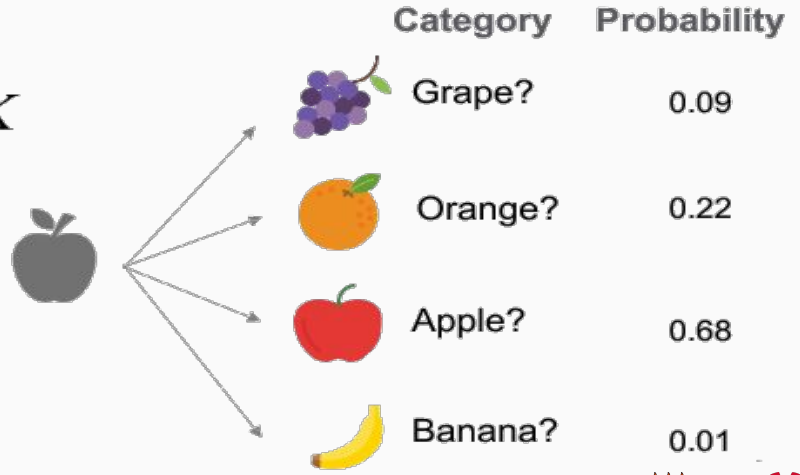
where *w* indicates the weight, *b* indicate the bias, and *wx* + *b* is regarded the linear function of *x*. Compare the preceding two probability values. The class with the higher probability value is the class of *x*.

Logistic Regression Extension - Softmax

Logistic regression applies only to binary classification problems.

Softmax regression is a generalization of the logistic regression for multiple classes.

$$p(y=k \mid x;w)=\frac{e^{w_k^T x}}{\sum_{l=1}^{K} e^{w_l^T x}}, k=1,2\cdots,K$$



Decision Tree

A decision tree is a tree structure (a binary tree or a non-binary tree) where each non-leaf node represents a test on a feature attribute.

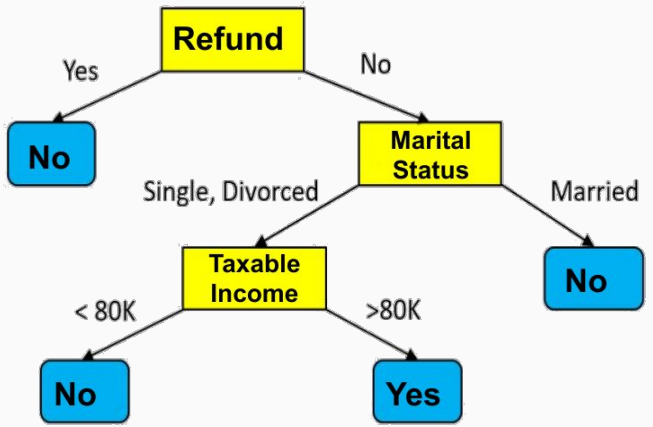
Each branch represents the output of a feature attribute in a certain value range, and each leaf node stores a category.

To use the decision tree, start from the root node, test the feature attributes of the items to be classified, select the output branches, and use the category stored on the leaf node as the final result.

Decision Tree Example

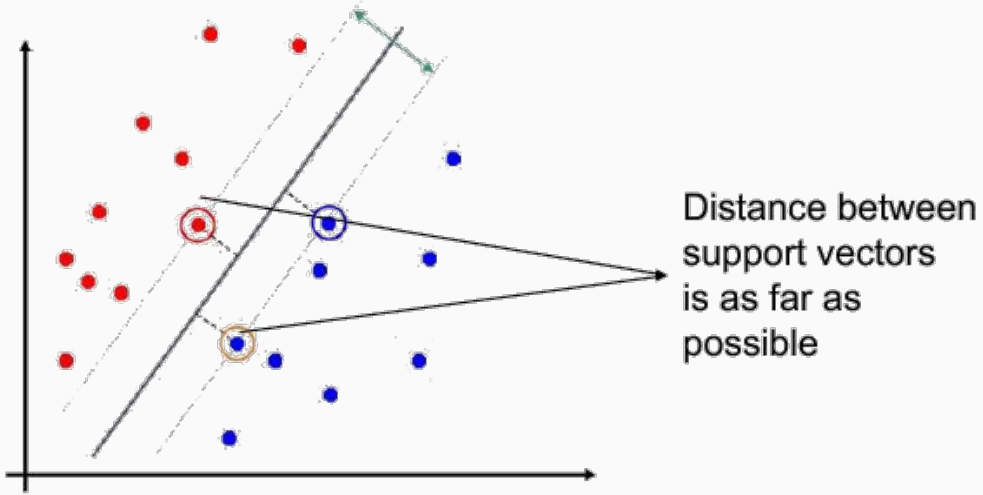
The following figure shows a classification when a decision tree is used. The classification result is impacted by three attributes: Refund, Marital Status, and Taxable Income.

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125,000	No
2	No	Married	100,000	No
3	No	Single	70,000	No
4	Yes	Married	120,000	No
5	No	Divorced	95,000	Yes
6	No	Married	60,000	No
7	Yes	Divorced	220,000	No
8	No	Single	85,000	Yes
9	No	Married	75,000	No
10	No	Single	90,000	Yes

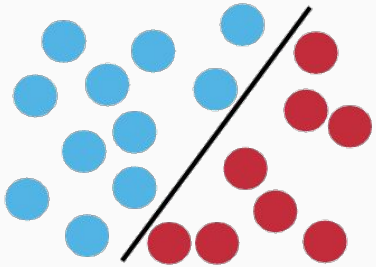


Support Vector Machine

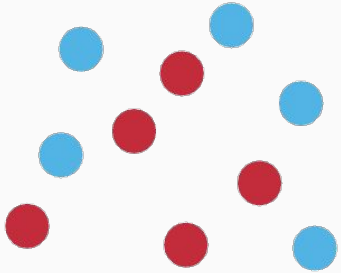
Support Vector Machines (SVMs) are binary classification models whose the core idea is to find a set of point (support vectors) for describe a set of functions that will act as boundaries for the different classes.



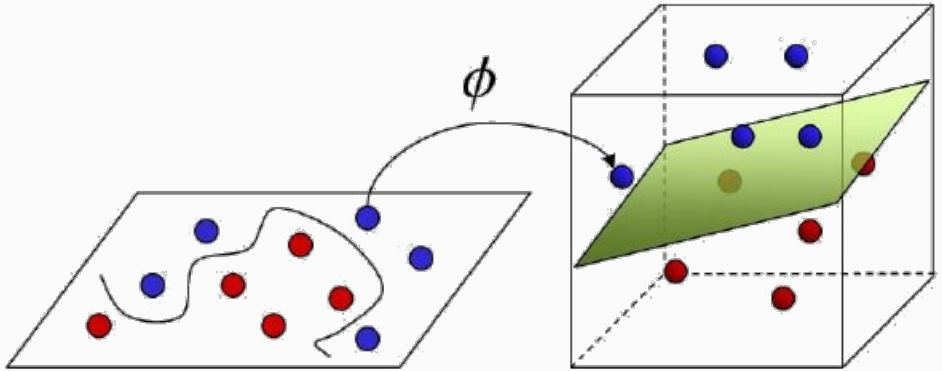
Support Vector Machine (Cont.)



Linear SVM can function well for linear separable datasets.



Nonlinear datasets cannot be split with straight lines.



Naive Bayes

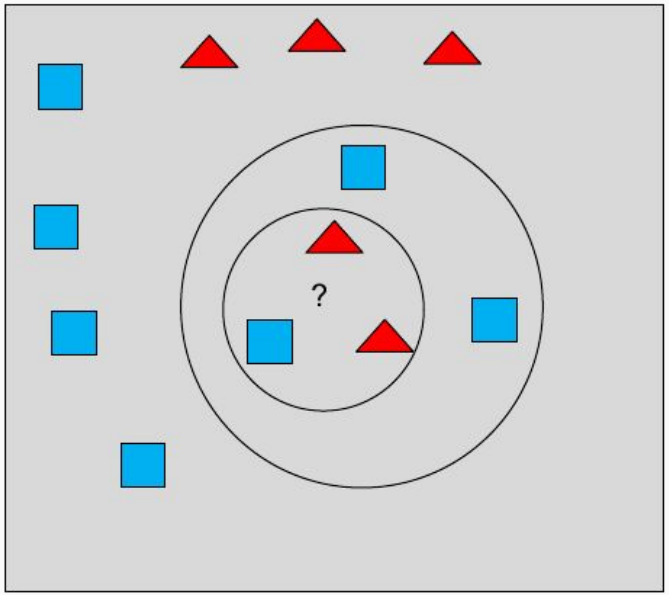
Naive Bayes algorithm: a simple multi-class classification algorithm based on the Bayes theorem. It assumes that features are independent of each other. For a given sample feature X, the probability that a sample belongs to a category H is:

$$P(C_k \mid X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n \mid C_k)P(C_k)}{P(X_1,\ldots,X_n)}$$

- 1,2....n are data features, which are usually described by measurement values of m attribute sets.
  - For example, the color feature may have three attributes: red, yellow, and blue.
- C_k indicates that the data belongs to a specific category .
- $(C_k \mid 1,2....n)$ is a posterior probability, or a posterior probability of under condition C_k.
- (C_k) is a prior probability that is independent of 1,2....n.
- $(1,2....n)$ is the priori probability of .

K-Nearest Neighbors

The KNN classification algorithm is a one of the simplest machine learning algorithms. According to this method, if the majority of k samples most similar to one sample (nearest neighbors in the eigenspace) belong to a specific category, this sample also belongs to this category.



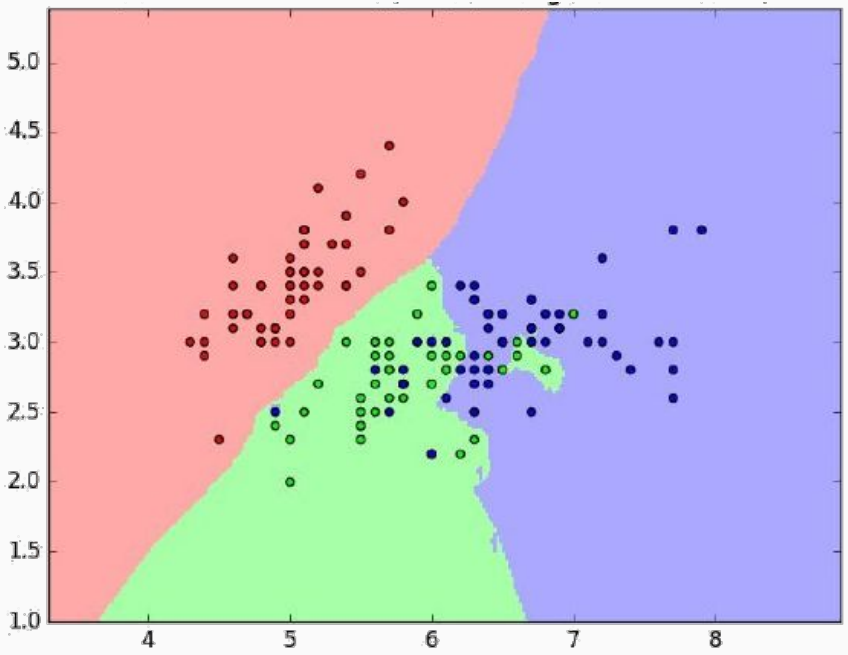
The target category of point ? varies with the number of the most adjacent nodes.

K-Nearest Neighbors (Cont.)

KNN is a non-parametric method which is usually used in datasets with irregular decision boundaries.

The KNN algorithm generally adopts the majority voting method for classification prediction and the average value method for regression prediction.

KNN requires a huge number of computations.



Thank You!

Next: 4.4 - Neural Networks