Machine Learning

KTU S7 CSE CS469

November 14, 2020

Contents

Chapter 1

Module 5

1.1 Markov Models

Markov Assumption: When the state at $t + 1$ only depends on t, regardless of the states of the previous weeks. Given the present state, the future is independent of the past.

1.1.1 Discrete Markov Chain

Markov Chains that have an initial process.

1.2 Hidden Markov Model

When the Markov Process is happening but is hidden from the rest of the world.

1.2.1 General Case

A Hidden Markov Model is characterized by the following

1. The number of states in the model, say N, let the states be $S_1, S_2, S_3 \ldots,$
2. The number of distinct observation symbols M

The values of N, M are implicitily defined in A, B and π. So a HMM is completely defined by parameter set

$$\lambda = (A, B, \pi)$$

1.2.2 Three basic problems

- Evaluation Problem
- Finding state sequence problem
- Learning model parameters problem

1.2.3 HMM Speech Recognition

1.3 Combining Multiple Learners

1.3.1 Model Combination Schemes

- Voting
- Bagging
- Boosting

Chapter 2

Module 6

Clustering is grouping similar data points together. Objects in the same group(cluster) are similar to those in the same group and not to the ones in the other groups.

2.1 K-Means Clustering

It's the simplest unsupervised learning algorithm. Let's say we need to make k clusters. We take k points as the centers and associate the points nearest to each center to it. We now take the average of the data point and replace the center with it. We repeat until it converges into some set points.

2.1.1 Computing Clusters

1. Take two points as arbitary
2. Compute distance between them and each point, assign the point to the one which is minimum. The equation to find the distance between points are.

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

3. Calculate new centers which is the average of the centers. Let c be the the number of points. Then the new points will be calculated as

$$\frac{x_1 + x_2 + x_3}{c}$$

4. Calculate the new distance from the new centers for each point.
5. Continue until two consecutive centers are the same.

2.1.2 Disadvantages of K-Means Algorithm

2.1.3 Applications

- Image Segmentation
- Data Compression

2.2   Expectation Maximisation Algorithm

2.3   Hierarchial Clustering

In Hierarchial Clustering, the clusters are arranged in a hierarchy, where the higher clusters are formed by merging the clusters at the next lower level. At the lowest level the cluster will contain one single data point and at the highest cluster the entire points. Whether to merge two clusters depend on the distance between them.

2.3.1   Dendograms

Dendograms are binary trees which represent Hierarchial Clusters.

2.3.2   Methods of Hierarchial Clustering

- Agglomerative Method
- Divisive Method

2.3.3   Agglomerative Method

2.3.4   Divisive Method

2.3.5   Measures of Distance

2.4   Density Based Clustering

2.4.1   DBSCAN Algorithm