Chengjun_Liu_7116099552_hw4

April 26, 2020

1. Did you review up to and including all the course content (excluding optional materials and guest lectures)? If you haven't, please do so first.

Yes, I have reviewed all course content materials.

2. Name three sources of randomness in a random forest model

1. Random selection and split of Datasets
2. Random selection of records from datasets for each tree
3. Random selection of features from datasets for each tree

3. Both random forest and boosting use an ensemble of trees. Which of these two builds trees sequentially?

The boosting trees build trees sequentially. Each tree after the first tree is built on minimizing the loss over the residuals left by the last tree.

4 What is the difference between support vector classifier and maximal margin classifier?

Maximal margin classifier requires that separating hyperplane exist to strictly split samples, while support vector classifier allows a soft boundary to accommodate some wrong classfications in some non-separable cases.

5 Explain the k-means algorithm in a few sentences

K-Means is an unsupervised algorithm to find the best divisions for specified K groups. By setting K starting points, dividing observations to the nearest starting points, calculation new central points using the average of the grouped observation, making another rounds of dividing observations to central points and calculating new points until centroids stop changing, we minimize the total distance of points to centroids and get the final divisions of observations to centroids as groups.

[ ]: