## Week 1 – Clustering

1. What is an example of unsupervised learning?
   1. K-means clustering.
2. What is NOT a reason that people use clustering?
   1. To identify clustering of variables and predict the response variable.
3. Which are clustering methods?
   1. Partitional clustering
   2. Hierarchical clustering
   3. Spectral clustering
4. What is the main difference between hard and soft clustering?
   1. In hard clustering, each data point either belongs to a cluster or not. In soft clustering, each datapoint is assigned a probability or likelihood of belonging to each cluster.
5. What is true?
   1. Agglomerative and divisive clustering are both hierarchical methods.
6. What are examples of clustering?
   1. Groups of people on a social network
   2. Genres of songs/movies in an audio library
   3. Segments of customers in the market
7. What are technical considerations in a clustering algorithm?
   1. Distance measures
   2. Initialisation
   3. The number of clusters
8. What are the steps of performing k-means clustering?
   1. Rescale the data for distance-based algorithms.
   2. Select the input variables and select the best K.
   3. Build the k-means model.
   4. Visualise the clustering plot and assign the data points to a cluster.
9. What is used to decide the number of clusters for hierarchical clustering?
   1. A dendrogram
10. How do you choose the best number of clusters for hierarchical clustering?
    1. The number of vertical lines in the dendrogram cut by a horizontal line that can transverse the maximum distance vertically without intersecting a cluster.
11. What is true about k-means and hierarchical clustering?
    1. K-means clustering requires prior knowledge about the number of clusters, whereas hierarchical clustering can use a dendrogram to determine the optimal number.
12. What is true about discriminant plots?
    1. The purpose of a discriminant plot is to visualise the data in a scatter plot and colour each data point by its cluster assignment.
    2. If the dataset is multi-dimensional, PCA can be performed on the variables and the data plotted against the first two PC coordinates.

## Week 2 – Decision Trees

1. What is a discriminative classifier?
   1. Neural networks.
2. If a boundary is considered memorised, it must be
   1. Too complex.
3. If a decision tree model is overfitting, what will help?
   1. Reducing the maximum depth.
4. When would it be best to use a decision tree?
   1. Predicting the default likelihood of loan applicants.
5. A decision tree is applicable only in the case of a binary response variable.
   1. False.
6. What is NOT relevant a tuning decision tree parameters?
   1. Maxsplit
7. Decision tree algorithms choose an attribute such that after the split there is a maximum increase in the entropy.
   1. False.
8. What is true about random forests?
   1. They are created from randomly selected partitions of the data.
9. If we need to predict the price of mushrooms based on where they were grown, where they were sold, their colour and their size, what is a useful model?
   1. A regression tree.

## Week 3 – Rules

1. What type of classifier is K-Nearest Neighbours?
   1. Non-parametric.
2. What is the right level of model complexity to determine the value of K?
   1. It is the point at which the training set accuracy is still increasing but the validation accuracy starts decreasing.
3. What is a desirable characteristic of K?
   1. An odd numbered value.
4. What helps when using KNN if there’s a lot of noise in the data?
   1. Larger values of K.
5. What are desirable conditions for applying KNN?
   1. Small to moderately-sized datasets.
   2. Small number of features.
   3. A categorical outcome.
6. What best describes the independence assumption?
   1. The joint probability condition can be decomposed into the product of the individual component probabilities.
7. What is a disadvantage of KNN?
   1. KNN does not learn a relationship between the input and output features like linear regression, logistic regression, decision trees etc. KNN starts learning when it sees the test data for making predictions.
8. Naïve Bayes is an unsupervised learning algorithm.
   1. False.
9. K in KNN is a parameter.
   1. False. (It’s a hyperparameter.)