Fundamentals of Data Science
Homework Assignment 1

Anitha Subramanian
002509602

Question 1: Based on this information, what can you say about the inductive bias for this prediction task?

Answer: If k=3 and the distance between neighbors are measured by Euclidean distance, then The classification of an instance X will be most similar to the classification of other 3 instances that are nearby in Euclidean distance.

Question 1: Briefly describe at least one restriction and one preference bias for the kNN classifier that you will train for this task.

Answer: Restriction Bias - Non Parametric Regression - Should be able to model to predict the class labels in the Iris dataset by measuring the Euclidean distance between the neighbors.

Answer: Preference Bias - Locality - 3 near points are posited to be similar.
Equality - All features matter equally.

Question 2: Indicate the data types and scales of attributes (i.e., ID, sepal length (cm), sepal width (cm), petal length (cm), petal width (cm), species) in Iris dataset. (Both the data types and the data scales.)

Answer: ID - Qualitative, Nominal
Sepal length (cm) - Quantitative, Continuous, Ratio
Sepal Width (cm) - Quantitative, Continuous, Ratio
Petal length (cm) - Quantitative, Continuous, Ratio
Petal width (cm) - Quantitative, Continuous, Ratio
Species - Qualitative, Nominal

Question 3: Machine learning is often referred to as an ill-posed problem. What does this mean?

Answer: Machine learning algorithms search through all the patterns that exist between a set of descriptive features and target feature to find a best model that is consistent with the training data. There is a possibility for multiple models consistent with the training data. Because of this Machine learning is often referred as an ill-posed problem as there is typically not enough information in the training data to select a best model.

Question 4: 1. Propose a predictive data analytics solution which can be used to address this situation.

Fundamentals of Data Science
Homework Assignment 1

Answer: Customer Enjoyment Prediction - The analytics practitioner could build a model that predict the likelihood that a customer would enjoy a particular movie. Then the company can offer personalized recommendations of new release movies.

Determining goals and requirements comes under Business understanding of CRISP-DM Process model.

2. Describe the predictive model that will be built (Note: not the model itself but the task)

Answer: Initially a predictive model need to be built to predict the likelihood that a customer would enjoy a particular movie. This can be done by collecting customer data history from the database of the company with most viewed and favorited. Based on the genre, casting and imdb ratings (the one with high rating) the company can predict a particular movie that a customer would enjoy.

This will come under Data understanding , Data preparation, Modeling of CRISP-DM Process model.

3. Describe how this model will be used by the business

Answer: Once the model is ready with prediction of favorite movies, the company then put in place a service that personalized recommendations of new releases for its customers and thus reduce churn by enticing customers to stay with the service by offering them a better product.

This will come under Evaluation and Deployment of CRISP-DM Process model.

4. How using this model can solve the original customer churn problem.

Answer: Presumably, if the company offer a better service to its customers , fewer customers would churn. If the model predicted the likelihood of a particular movie that a customer would enjoy, then the company can offer personalized recommendations of new releases for its customers. This would definitely make the customers satisfied and happy to see their favorites listed in their account. By this way this model will help to solve the original customer churn problem.