Dipesh Mainali

Week 5 Quiz
Content: Chapter 2 “Introduction to Machine Learning with Python”

1. What is a decision boundary and what does a decision boundary chart show us?

   A **decision boundary** is a line, or surface in higher dimensions, that separates different classes in a classification problem.

   A **decision boundary chart** shows how a model divides the feature space into regions where it predicts different classes. This helps to visualize how a model is making decisions and which areas of the input space belong to each class.

2. What is the difference between classification and regression?

   Classification is the task of predicting a discrete value (class label), that can be binary or multiclass. It involves assigning data points to one of a set of predefined categories/classes. An example of (binary) classification is classifying an email as spam or not spam.

   Regression involves predicting a continuous value. It is used when the output is a numeric value instead of a class label. An example of regression is predicting house prices.

3. What is feature importance?

   Feature importance refers to the relative importance of each input feature for a model while making predictions. It indicates which features have the greatest influence on the model's decision, helping to identify which variables contribute most to the target prediction (output).

4. Do we prefer a simple or intricate model? Why?

   We, generally, prefer a **simple model** because it is less likely to overfit the training data. A simple model generalizes better to unseen data, whereas an intricate (complex) model may memorize the training data and perform poorly on new data due to overfitting. We can also avoid complex models by providing a variety of high-volume data.

5. You have trained a KNN model with 1 neighbor. If you test the model’s prediction accuracy using the testing data what happens?

   When using a **KNN model with 1 neighbor**, the model will look at the single closest data point in the feature space to make a prediction. This can lead to high variance and make the model highly sensitive to noise in the data. As a result, the model might perform very well on the training data but poorly on the testing data, especially if the training data has noise or outliers (a result of overfitting).