**IE 6318 Data Mining and Analytics**

**Homework 2**

Classification Using KNN & Decision Tree

1. Make a KNN classifier for the 3-class classification problem using the distance function you made in HW1 for Minkowski Distance. The KNN function performs classification based on the majority voting of K-nearest neighbors. Implement the KNN classifiers to the IRIS dataset using different KNN parameter settings: K = 3, 5, 7 for K-nearest neighbors, and r = 1, 2, 5 for the distance order of Minkowski Distance. For each parameter setting of K and r, perform the classification experiment using **5-fold cross validation**,

1) report the classification accuracies for the 9 KNN parameter settings.

2) find the best KNN parameter setting that generates the highest accuracy, report the corresponding confusion matrix for the best KNN parameter setting.

2. Design a simple decision tree using the attributes of petal width and petal length to classify two types of iris flowers: Versicolor and Virginica, as shown below. Assume the binary decision boundary on Petal Length is 43, and the decision boundary on Petal Width is 18. Make this simple decision tree into a decision tree classification function. The function should have two inputs: petal width and petal length, and one output: classification outcome Versicolor or Virginica. Implement the decision tree function to classify the 100 iris samples of Versicolor or Virginica, and report the classification accuracy and confusion matrix.

