CNG 562- Machine Learning

Assignment #2 (10 % of the Actual Mark)

This is a group assignment; max group size is 3.

Given the Iris dataset located in the sklearn library, we need to find the best strategy for classifying the dataset.

Task 1: Finding the best KNN predictor for the dataset.

1) We need to find the best distance metric to be used for the dataset, also the best value of k. Hence, compute the accuracy of KNN for raw data for the following distance metrics, and following k values

Distance	K=3	K=5	K=7	K=9	K=11
euclidean					
manhattan					
chebyshev					
mahalanobis					
minkowski					

2) Repeat the above task by transforming the raw data into z scores. What is the best classification strategy based on your findings?

3) Based on your findings, try the different weighted methods available in sklearn, and report your findings.

Task 2: Finding the best Naïve Bayes predictor for the dataset.

1) Try the Gaussian Naive Bayes classifier, babysit the model to get the best accuracy.
2) Try Multinomial Naive Bayes classifier, babysit the model to get the best accuracy.
3) Try Complement Naive Bayes, babysit the model to get the best accuracy.
4) Try Bernoulli Naive Bayes, babysit the model to get the best accuracy.
5) Try Categorical Naive Bayes, babysit the model to get the best accuracy.

Task 3: Finding the best decision tree predictor for the dataset.

1) Try the splits strategies against different split metrics (i.e., entropy and Gini)

2) Try different max_depth values, and report the best performing depth. Compare it to the default depth option.

3) Try different class weights, and report your findings.

Based on your experiments, what is the best classification strategy for iris dataset using the decision tree classifier?

Task 4: Finding the best SVM predictor for the dataset.

1) Try different kernel types, and report best performing type. Make sure you set the dependent parameters correctly.

2) Try a balanced class weight, and report your findings.

3) Try different gamma values, and report your findings.

Based on your experiments, what is the best classification strategy for iris dataset using the SVM classifier?

Task5: Try boosting for the worst and best predictors you found, report your findings.

1) Try different number of estimators, and report your findings for the best number.

2) Try different learning rates, and report your findings for the best learning rate.

Based on your experiments, what is the best boosting strategy for iris?