BANA 290 Week 7 – In Class Exercise

Team Members:

1. Load diabetes.arff in Weka. In this exercise, you need to compare different of testing techniques. Pick a classification technique and compare how choosing different test options vary the accuracy measures. Pick at least 3 values for cross validation between 5-20 folds and 3 values for percentage split between 20-90%, and contrast the accuracy measures.

Tabulate your findings in the following table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classifier | Training Set (Accuracy) | C.V. (folds) | C.V. Accuracy | Training Split (%) | T.S. Accuracy |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

2. Now fix the C.V. folds that give you the maximum accuracy in Part 1. Compare 3 different classification techniques using C.V. Tabulate the best performance of each of the 3 chosen classifiers (Note you should change the preprocessing steps and the options of the classifiers to increase the accuracy as much as possible).

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Accuracy | FP (Type 1 error) | FN (Type 2 error) |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

3. Now load the segment-challenge.arff file. It contains characteristics of images that can be used to classify them into 1 of 7 classes – “brickface,” “foliage,” “sky,” “cement,” “window,” “path,” and “grass.” Choose a classifier, and compute the confusion matrix. Which type of images are the most difficult to classify? Why?

4. For the best model in Part 3, plot the ROC curve and find the area under the curve. How does the threshold change when the Type 1 Error cost is 2 and Type 2 Error cost is 1? How does the threshold change when the Type 1 Error cost is 1 and Type 2 Error cost is 2?

5. Load segment-challenge.arff. Try K-means clustering with 2 clusters. Compute the Cluster Validity (C.V.). Increment k and compute C.V. Try with 5-7 values of k and choose the best number of clusters using the elbow method.

|  |  |
| --- | --- |
| k | C.V. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |