My Name (myNetID)

IE598 MLF F18

Module 6 Homework (Cross validation)

Using the Iris dataset, with 90% for training and 10% for test and the decision tree model that you submitted for Module 2:

**Part 1: Random test train splits**

1 2 3 4 5 6 7 8 9 10 mean std

in sample 0.993 0.978 0.993 0.985 0.978 0.993 0.993 0.993 0.993 1.000 0.990 0.007

out of sample 0.867 0.867 0.867 0.933 0.867 0.933 0.933 0.933 1.000 0.933 0.913 0.043

**Part 2: Cross validation**

1 2 3 4 5 6 7 8 9 10 mean std Out-of-sample-accuracy

0.933 1.0 0.867 0.933 0.933 1.0 1.0 0.917 0.917 0.917 0.942 0.042 1.0

**Part 3: Conclusions**

Run many test and find the possible % of the out put then run the test set base on it. CV is more effective less code more work done.

**Part 4: Appendix**

Link to github repo