Vikram Tandon
1001553980

REPORT

PROBLEM A (ATNT data)

KNN Classifier Accuracy:

2-fold CV : 0.78000000000000003, 0.71499999999999997 **Average**:0.7475
3-fold CV: 0.86567164179104472, 0.84962406015037595, 0.75187969924812026
Average:0.822391800397
5-fold CV: 0.92500000000000004, 0.90000000000000002, 0.82499999999999996, 0.83750000000000002, 0.90000000000000002 **Average**:0.8775
10-fold CV: 0.90000000000000002, 0.90000000000000002, 0.80000000000000004, 0.875, 0.84999999999999998, 0.97499999999999998, 0.84999999999999998, 0.92500000000000004, 0.92500000000000004, 0.92500000000000004
Average:0.8925

SVM Classifier Accuracy:

2-fold CV : 0.0050000000000000001, 0.01 **Average**:0.0075
3-fold CV:0.007462686567164179, 0.03007518796992481, 0.0075187969924812026
Average:0.0150188905099
5-fold CV: 0.0, 0.0, 0.0, 0.0, 0.0 **Average**:0.0
10-fold CV:0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
Average: 0.0

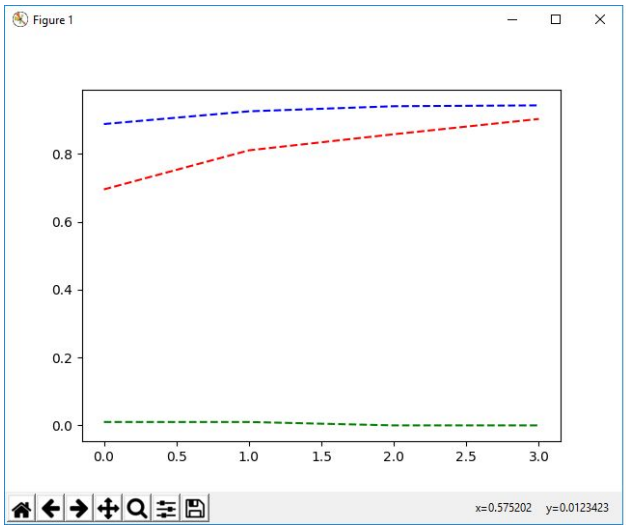
Centroid Classifier Accuracy:

2-fold CV : 0.90000000000000002, 0.92500000000000004 **Average**:0.9125
3-fold CV:0.94776119402985071, 0.93984962406015038, 0.8571428571428571
Average:0.914917891744
5-fold CV: 0.9375, 0.98750000000000004, 0.91249999999999998, 0.9375, 0.875
Average:0.93
10-fold CV:1.0, 0.92500000000000004, 0.97499999999999998, 0.90000000000000002, 0.90000000000000002, 0.875, 0.92500000000000004, 0.94999999999999996, 0.97499999999999998, 0.82499999999999996
Average: 0.925

Trend Observation:

Vikram Tandon
1001553980

1. SVM(Green trend) increases initially but then does not give substantial result with the increase in number of folds.
2. Knn(Red Trend) increases with the number of folds.It is always less than or equal to Centroid.
3. Centroid(Blue Trend) decreases initially but gradually remains near to constant with the increase in number of folds.



PROBLEM B (hand-written-letter)

KNN Classifier Accuracy:

2-fold CV : 0.73175542406311633, 0.63905325443786987
Average:0.68540433925
3-fold CV: 0.69526627218934911, 0.74556213017751483, 0.76923076923076927
Average:0.822391800397
5-fold CV: 0.75369458128078815, 0.77832512315270941, 0.75862068965517238, 0.70935960591133007, 0.74257425742574257 **Average**:0.748514851485
10-fold CV: 0.78431372549019607, 0.72549019607843135, 0.77450980392156865, 0.7647058823529412, 0.68316831683168322, 0.7722772277227723, 0.83168316831683164, 0.80198019801980203, 0.70297029702970293, 0.78217821782178221
Average:0.762327703359

Vikram Tandon
1001553980

SVM Classifier Accuracy:

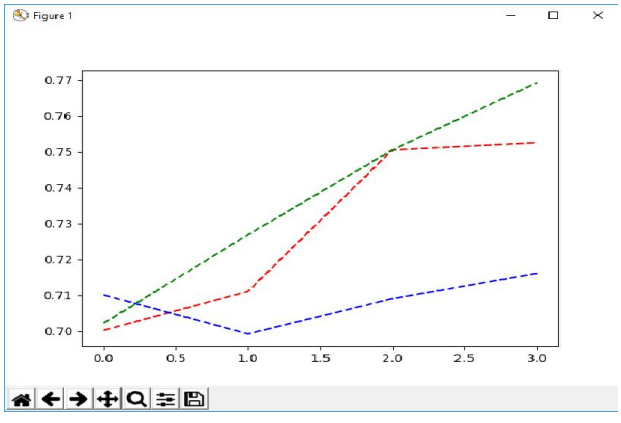
2-fold CV : 0.7100591715976331, 0.69230769230769229 **Average**:0.701183431953
3-fold CV:0.74852071005917165, 0.75739644970414199, 0.71893491124260356
Average:0.741617357002
5-fold CV: 0.77339901477832518, 0.76847290640394084, 0.74384236453201968, 0.76847290640394084, 0.73267326732673266 **Average**:0.757372091889
10-fold CV:0.80392156862745101, 0.76470588235294112, 0.78431372549019607, 0.79411764705882348, 0.80198019801980203, 0.7722772277227723, 0.76237623762376239, 0.71287128712871284, 0.74257425742574257, 0.7722772277227723
Average: 0.771141525917

Centroid Classifier Accuracy:

2-fold CV : 0.67455621301775148, 0.71992110453648916 **Average**:0.697238658777
3-fold CV:0.72189349112426038, 0.69230769230769229, 0.72189349112426038
Average:0.712031558185
5-fold CV: 0.68472906403940892, 0.69950738916256161, 0.77832512315270941, 0.72906403940886699, 0.68316831683168322 **Average**:0.714958786519
10-fold CV:0.72549019607843135, 0.72549019607843135, 0.71568627450980393, 0.74509803921568629, 0.65346534653465349, 0.71287128712871284, 0.72277227722772275, 0.75247524752475248, 0.7722772277227723, 0.70297029702970293
Average: 0.722859638905

Trend Observation:

1. SVM(Green trend) increases with the number of folds.
2. Knn(Red Trend) increases with the number of folds initially but ten decreases gradually.It is always less than or equal to SVM.
3. Centroid(Blue Trend) decreases initially but gradually increases with the number of folds.Its is always less than SVM.



Vikram Tandon
1001553980

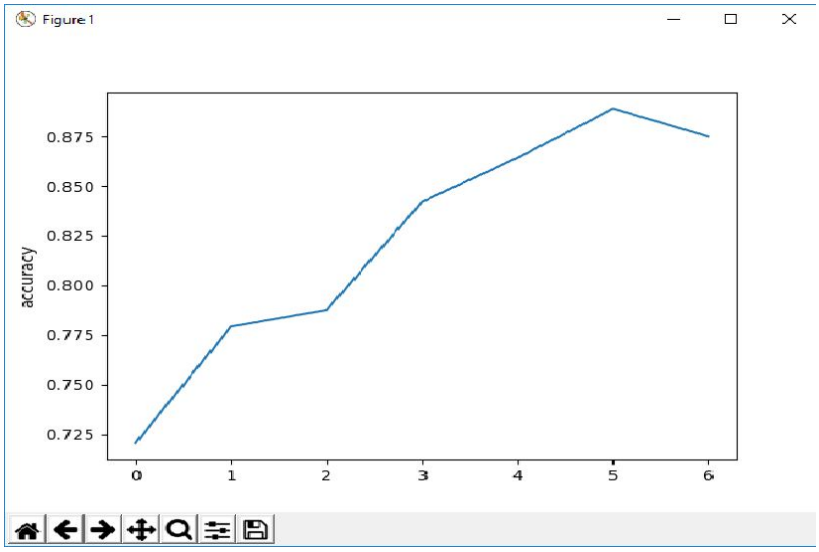
PROBLEM D

Seven different data-split cases, running the centroid classifier computing test image classification accuracy as follows:

```
input = 'qrstuvwxyz'
```

- (train=5 test=34) = 0.720588235294
- (train=10 test=29) = 0.779310344828
- (train=15 test=24) = 0.7875
- (train=20 test=19) = 0.842105263158
- (train=25 test=14) = 0.864285714286
- (train=30 test=9) = 0.888888888889
- (train=35 test=4) = 0.875

Plotted above 7 accuracy on one curve in figure below:



Trend Observation:

- As the train data increases the accuracy improves substantially in first 6 of 7 cases.
- In the last case the accuracy decreases a bit.

Vikram Tandon
1001553980

PROBLEM E

Seven different data-split cases, running the centroid classifier computing test image classification accuracy as follows:

`input = 'abcdefghij'`

- (train=5 test=34) = 0.738235294118
- (train=10 test=29) = 0.824137931034
- (train=15 test=24) = 0.779166666667
- (train=20 test=19) = 0.805263157895
- (train=25 test=14) = 0.8
- (train=30 test=9) = 0.811111111111
- (train=35 test=4) = 0.825

Trend Observation:

- As the train data increases the accuracy improves abruptly initially for the first case.
- When train data is increased after the first case as shown above accuracy decreases and increases alternatively.

Plotted above 7 accuracy on one curve in figure below: