HW 3

CS 7616 Pattern Recognition

Anisha Gartia

GTID 903136557

KNN

For the first part of the assignment, we have implemented k-nearest neighbor classification algorithm. Here, we classify a data point based on majority vote of k of the neighbors nearest to the data point. Hence, our model is a parametric model, where we need to learn the value of k. One of the most common issues of classification [problems is over-fitting. To avoid this problem, we train a model over random subsets of the data and average the resultant accuracy, a concept called Cross-Validation.

We do 2-fold and 5-fold cross-validation for all data sets, and additionally, for wine dataset, we do Leave One Out(LOO) cross-validation. LOO has a higher number of subsets, thus resulting is a higher accuracy model. But, as it is computation intensive, we perform it only on wine dataset.

For wine data, as we are not given pre-defined train and tests, we randomly perform 80-20 split on the entire data (20% is reserved for test). MNIST and Office have predefined train and test sets, thus, we proceed to cross-validation. Here, we compute the accuracy of each cross-validation subset (5 subsets each of size 20%, 50%, and 80% along with 100%) for a number of k values and take the average of all of them. The k which highest average accuracy is ultimately selected. As the number of features is large for MNIST and Office dataset, we have performed LDA to reduce the dimensionality of a dataset without losing any information.

In addition to k, the defining component of k-NN model is the distance metric for computing the nearest neighbors. As we use LDA, out data is transformed to a space where each class is closer together with respect to variance. Thus, we use Mahalanobis distance, which gives us the variance-proximity instead of euclidean. This gives us more closely related k nearest neighbours and thus higher accuracy. We use the unbuilt LMNN transformation function of sci-kit learn library which does the above for us.

For almost all the datasets, we see higher cross-validation error as compared to test error. This is because the size of each subset is reduced to a large extent for some of them.

For each of 20, 50 and 80% subset size, we calculate the stability of accuracy of the five subsets. We do this by comparing the variance between the subsets for each subset size. The trend often noted is that the accuracy stays more stable as the number of training data increases in the subset. The reasoning behind this can be given as follows. For higher number of training data points, we can build a more accurate model every time. Thus, there will be a higher probability of seeing all the features that define the data. Thus, we also see that the average accuracy increases with increasing size of a subset.

Wine Dataset

Best K = 3

Final Test Accuracy Score = 1.0
CV accuracy = 0.994130139661

Cross Validation Error = 0.586986033925 %
Test Error percentage for best k = 0.0 %

Average accuracy with respect to k:

K = 1	0.99690512
K = 3	0.99413014
K = 5	0.99330357

Average Accuracy with respect to Folds:
Comparing the accuracies for each fold, we see Leave one out Worked best

2 Fold	0.98706336
5 fold	0.99849073
Leave one out	0.99878474

Stability of accuracy for each subset size:
Variance in accuracy of between the subsets is

d= 20%	1.97782817e-04
d = 50%	6.40464492e-06
d = 80%	2.51309737e-06

Thus we see the stability increases with increasing size of d.

Accuracy with respect to the size of the training subset

d= 20%	0.98730159
d = 50%	0.99873463
d = 80%	0.99764843

MNIST Dataset

Best K = 5
CV accuracy = 0.963827004483
Final Test Accuracy Score = 0.959671396565

Cross Validation Error = 3.6172995517 %
Test Error percentage for best k = 4.03286034354 %

Accuracy with respect to k:

K = 1	0.96226061
K = 3	0.96345013
K = 5	0.963827

Accuracy with respect to F-Folds:
Comparing the accuracies for each fold, we see Leave one out Worked best

2 Fold	0.96310389
5 fold	0.9632546

Stability of accuracy for each subset size:
Variance in accuracy of between the subsets of same size is

d= 20%	1.48390101e-06
d = 50%	3.49854371e-07
d = 80%	3.32901321e-07

Thus we see the stability increases with increasing size of d.

Accuracy with respect to the size of the training subset

d= 20%	0.96275086
d = 50%	0.96283713
d = 80%	0.96389474

Office Dataset

Best K = 10

CV accuracy 0.651362506107

Final Test Accuracy Score = 0.552800974252

Cross Validation Error = 34.8637493893 %

Test Error percentage for best k = 44.7199025748 %

Accuracy with respect to k:

K = 1	0.65507802
K = 3	0.65462056
K = 5	0.65136251

Accuracy with respect to 2 Fold, 5 fold:
Comparing the accuracies for each fold, we see Leave one out Worked best

2 Fold	0.64453023
5 fold	0.66284382

Stability of accuracy for each subset size:

Variance in accuracy of between the subsets of same size is

d= 20%	2.70209214e-05
d = 50%	2.89656808e-06
d = 80%	5.52469003e-07

Accuracy with respect to the size of the training subset

d= 20%	0.62485974
d = 50%	0.6552153
d = 80%	0.67665533

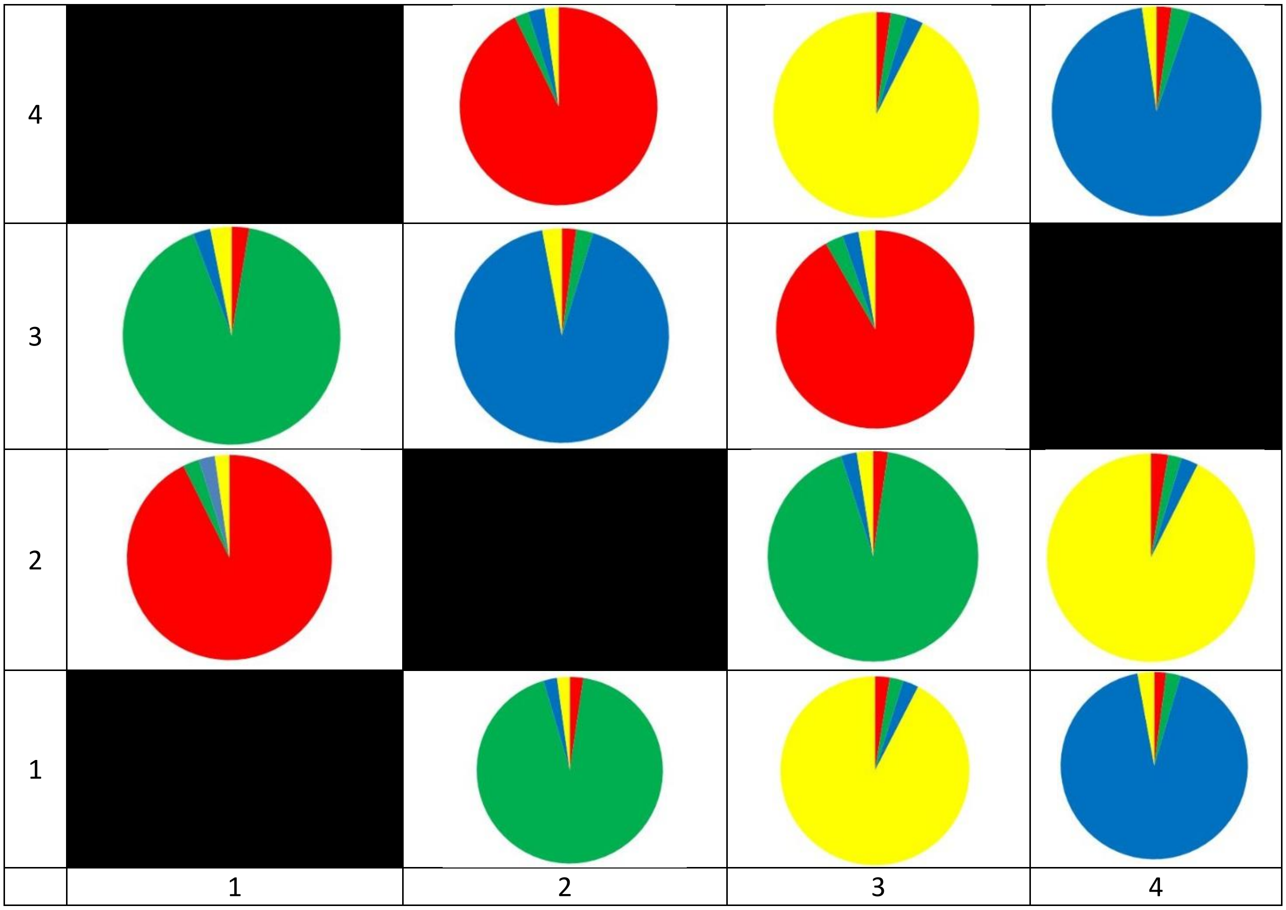
HMM

Observation Probabilities:

(probability of observing a particular output 'o')

States \ Observattions	Red	Blue	Green	Yellow
(1,1)	0	0	0	0
(1,2)	0.9257272	0.02569896	0.02541655	0.02287489
(1,3)	0.02608953	0.9163949	0.02579306	0.03142603
(1,4)	0	0	0	0
(2,1)	0.02336449	0.93090788	0.02336449	0.02202937
(2,2)	0	0	0	0
(2,3)	0.02170088	0.0255132	0.92316716	0.02932551
(2,4)	0.92699441	0.02266706	0.0273771	0.02266706
(3,1)	0.02519685	0.02362205	0.02677165	0.92409449
(3,2)	0.02300939	0.92764154	0.02422041	0.02482592
(3,3)	0.91659139	0.0298103	0.02619693	0.02710027
(3,4)	0.02267442	0.025	0.02703488	0.925
(4,1)	0.02067498	0.02493159	0.92520523	0.02888416
(4,2)	0.02658004	0.02096869	0.02658004	0.9255759
(4,3)	0	0	0	0
(4,4)	0.0230701	0.02928128	0.9248743	0.02247856

Visualisation



In the above observation probability, we see that even though the highest probability in each state is the actual the color as given in the grid, we see a small portion belonging to the other colors. This is because of the transition between states. The observation probability is more accurate that the small world given, as it includes probabilities of occurrences of all the observation in addition to the ones given in small world. Although these are small, in addition to transition probability, we can accurately find a model.

Transition probability:

(probability of traversing from one state to another)

```
0.        , 0.        , 0.        , 0.        , 0.        ,
         0.        , 0.        , 0.        , 0.        , 0.        ,
         0.        , 0.        , 0.        , 0.        , 0.        ,
         0.        ],
       [ 0.        , 0.7627551 , 0.23696145, 0.        , 0.        ,
         0.        , 0.        , 0.        , 0.        , 0.        ,
         0.        , 0.        , 0.        , 0.        , 0.        ,
         0.        ],
```

```
[ 0.        , 0.24769551, 0.49509367, 0.        , 0.        ,
  0.        , 0.25691347, 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.73673606,
  0.        , 0.        , 0.        , 0.26292814, 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.25228681, 0.        , 0.        ,
  0.        , 0.25671289, 0.25051638, 0.        , 0.        ,
  0.24018885, 0.        , 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.24800473, 0.49275791, 0.        , 0.        ,
  0.        , 0.25894177, 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.24628517,
  0.        , 0.        , 0.        , 0.24944673, 0.25735062,
  0.        , 0.        , 0.24660133, 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.24467438, 0.24893488,
  0.25745587, 0.        , 0.        , 0.24863055, 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.24825916, 0.        , 0.        , 0.25734181,
  0.2479564 , 0.24613987, 0.        , 0.        , 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.25087822, 0.        , 0.        ,
  0.24180328, 0.2602459 , 0.        , 0.        , 0.        ,
  0.24677986],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.24091603, 0.        ,
  0.        , 0.        , 0.50198473, 0.25679389, 0.        ,
  0.        ],
[ 0.        , 0.        , 0.        , 0.        , 0.        ,
  0.        , 0.        , 0.        , 0.        , 0.23915627,
  0.        , 0.        , 0.25044563, 0.51010101, 0.        ,
  0.        ],
```

```
[ 0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ],
[ 0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
  0.        ,  0.25193337,  0.        ,  0.        ,  0.        ,
  0.74776919]
```

Average error of test sequence:

We see the following error(range 0 to 1) probability for the given test sequences:

```
[ 0.135],
    [ 0.105],
    [ 0.255],
    [ 0.16 ],
    [ 0.17 ],
    [ 0.26 ],
    [ 0.12 ],
    [ 0.18 ],
    [ 0.175],
    [ 0.225],
    [ 0.19 ],
    [ 0.28 ],
    [ 0.165],
    [ 0.13 ],
    [ 0.115],
    [ 0.155],
    [ 0.185],
    [ 0.095],
    [ 0.24 ],
    [ 0.195],
    [ 0.17 ],
    [ 0.175],
    [ 0.2  ],
    [ 0.19 ],
    [ 0.135],
    [ 0.235],
    [ 0.175],
    [ 0.21 ],
    [ 0.095],
    [ 0.175],
    [ 0.195],
    [ 0.235],
    [ 0.17 ],
```

[0.145],
[0.19],
[0.145],
[0.225],
[0.105],
[0.19],
[0.245],
[0.225],
[0.11],
[0.145],
[0.185],
[0.2],
[0.125],
[0.165],
[0.265],
[0.17],
[0.17],
[0.235],
[0.145],
[0.165],
[0.16],
[0.11],
[0.205],
[0.23],
[0.185],
[0.195],
[0.16],
[0.15],
[0.19],
[0.105],
[0.175],
[0.225],
[0.24],
[0.135],
[0.21],
[0.185],
[0.13],
[0.13],
[0.17],
[0.255],
[0.155],
[0.205],
[0.175],
[0.195],
[0.175],
[0.13],
[0.205],
[0.16],

```
[ 0.36 ],
[ 0.16 ],
[ 0.18 ],
[ 0.23 ],
[ 0.115],
[ 0.12 ],
[ 0.31 ],
[ 0.285],
[ 0.185],
[ 0.11 ],
[ 0.205],
[ 0.17 ],
[ 0.215],
[ 0.14 ],
[ 0.115],
[ 0.215],
[ 0.105],
[ 0.24 ],
[ 0.12 ],
[ 0.155],
[ 0.22 ],
[ 0.195],
[ 0.215],
[ 0.17 ],
[ 0.25 ],
[ 0.205],
[ 0.255],
[ 0.12 ],
[ 0.1  ],
[ 0.105],
[ 0.21 ],
[ 0.125],
[ 0.275],
[ 0.21 ],
[ 0.2  ],
[ 0.17 ],
[ 0.25 ],
[ 0.175],
[ 0.165],
[ 0.08 ],
[ 0.225],
[ 0.235],
[ 0.31 ],
[ 0.245],
[ 0.13 ],
[ 0.2  ],
[ 0.185],
[ 0.22 ],
```

```
[ 0.22 ],
[ 0.175],
[ 0.155],
[ 0.12 ],
[ 0.225],
[ 0.22 ],
[ 0.18 ],
[ 0.145],
[ 0.26 ],
[ 0.195],
[ 0.095],
[ 0.21 ],
[ 0.135],
[ 0.095],
[ 0.115],
[ 0.07 ],
[ 0.255],
[ 0.245],
[ 0.16 ],
[ 0.22 ],
[ 0.15 ],
[ 0.155],
[ 0.185],
[ 0.15 ],
[ 0.31 ],
[ 0.255],
[ 0.135],
[ 0.295],
[ 0.195],
[ 0.39 ],
[ 0.225],
[ 0.165],
[ 0.2  ],
[ 0.11 ],
[ 0.155],
[ 0.23 ],
[ 0.2  ],
[ 0.13 ],
[ 0.28 ],
[ 0.185],
[ 0.23 ],
[ 0.18 ],
[ 0.16 ],
[ 0.225],
[ 0.28 ],
[ 0.315],
[ 0.215],
[ 0.16 ],
```

```
 [ 0.215],
 [ 0.23 ],
 [ 0.165],
 [ 0.26 ],
 [ 0.205],
 [ 0.13 ],
 [ 0.19 ],
 [ 0.185],
 [ 0.195],
 [ 0.255],
 [ 0.19 ],
 [ 0.175],
 [ 0.28 ],
 [ 0.235],
 [ 0.225],
 [ 0.105],
 [ 0.285],
 [ 0.18 ],
 [ 0.185],
 [ 0.11 ],
 [ 0.23 ],
 [ 0.13 ],
 [ 0.12 ]])
```

Discussion:

Hidden Markov model is a means to generate a model that tells us the most probable observations of new states. We learn the model based on the training data, which gives us possible coordinates and its observations. The transition between states is an important factor in deciding the next state. Thus although we often see a certain observation for a certain state, it may be majorly because the previous state is the same. Thus in our observation probability, we see that each state has smaller portions of all the possible observations.

HMM consists of two phases. The forward phase is where we learn the model. And next is the Viterbi which is for predicting states. In the forward phase, we learn the transition and observation probabilities for each sequence, for each observation. We have initialization of the first observation, and induction of the remaining.

The Viterbi phase is for predicting the observation of test sequences. Here we perform back tracking, starting from the last observation and going backwards each observation to reach the first. Thus the four sub stages are initialization, recursion, and termination and back tracking.

In our small world, HMM works as the training set has all the states. Thus we are able to build an all-inclusive transition and observation matrix each. Also, the probability of the robot observing the true color is 90% which will give us a high probability of back tracking in Viterbi.