Madhudeep Petwal

**********NOTE**********

Ony a single instance of confusion matrix is shown in this document as number of confusion matrix are large. A separate file is there for each data set starting with name out***.txt for confusion matrix and other results. For confusion matrix convension is as follows

	Class1	Class2	Class 3
Class1			
Class 2			
Class 3			

Questions to be answered

1. Number of Data sets chosen = 3
2. Distance function chosen = Euclidian distance

1. IRIS DATA SET

1. Features 5.
2. Number of Instances: 150 (50 in each of three classes)
3. Number of Attributes: 4 numeric, predictive attributes and the class
4. Attribute Information:
   1. sepal length in cm
   2. sepal width in cm
   3. petal length in cm
   4. petal width in cm
   5. class:
      -- Iris Setosa
      -- Iris Versicolour
      -- Iris Virginica

2. WINE DATA SET

1. Features 13 ans 3 classes
1) Alcohol
2) Malic acid
3) Ash
4) Alcalinity of ash
5) Magnesium
6) Total phenols

7) Flavanoids
8) Nonflavanoid phenols
9) Proanthocyanins
10)Color intensity
11)Hue
12)OD280/OD315 of diluted wines
13)Proline

3. **BREASTCANCER DATA SET**

1. Number of Instances: 699 (as of 15 July 1992)
2. Number of Attributes: 10 plus the class attribute
3. Attribute Information: (class attribute has been moved to last column)

#	Attribute	Domain
1.	Sample code number	id number
2.	Clump Thickness	1 - 10
3.	Uniformity of Cell Size	1 - 10
4.	Uniformity of Cell Shape	1 - 10
5.	Marginal Adhesion	1 - 10
6.	Single Epithelial Cell Size	1 - 10
7.	Bare Nuclei	1 - 10
8.	Bland Chromatin	1 - 10
9.	Normal Nucleoli	1 - 10
10.	Mitoses	1 - 10
11.	Class:	(2 for benign, 4 for malignant)

RESULTS

1. IRIS DATA SET

Knn classifier for IRIS DATA for K=1 random subsampling
confusion_matrix =

```
  25    0    0
   0   25    0
   0    3   22
```

Mean for K=1 is 95.466667
Standard Deviation for K=1 is 1.799863

Knn classifier for IRIS DATA for K=3 random subsampling
confusion_matrix =

```
  25    0    0
   0   25    2
   0    3   20
```

Mean for K=3 is 94.266667
Standard Deviation for K=3 is 3.445520

Knn classifier for IRIS DATA for K=1 five folds cross validation
confusion_matrix =

```
  11    0    0
   0    8    1
   0    0   10
```

1	2	3	4	5	6	7	8	9	10
96.108	96.108	95.441	96.129	95.441	94.796	95.462	95.462	96.065	96.086

Grand Mean for 5 folds for K=1 is 95.709677
Standard Deviation for 5 folds for K=1 0.454872

Knn classifier for IRIS DATA for K=3 five folds cross validation
mean_matrix =

1	2	3	4	5	6	7	8	9	10
96.108	96.108	95.441	96.129	95.441	94.796	95.462	95.462	96.065	96.086

Grand Mean for 5 folds for K=3 is 94.559140
Standard Deviation for 5 folds for K=3 0.919241

2. WINE DATA SET

Knn classifier for WINE DATA for K=1 random subsampling
confusion_matrix =

25 0 4
2 21 12
3 9 13
Mean for K=1 is 72.359551
Standard Deviation for K=1 is 3.561008

Knn classifier for WINE DATA for K=3 random subsampling
confusion_matrix =
25 1 3
3 22 6
5 12 12
Mean for K=3 is 68.314607
Standard Deviation for K=3 is 3.885034

Knn classifier for WINE DATA for K=1 five folds cross validation
confusion_matrix =
13 0 1
2 5 3
1 6 4

1	2	3	4	5	6	7	8	9	10
74.598	74.181	77.502	77.501	74.088	76.390	75.353	75.833	78.074	76.343

Grand Mean for 5 folds for K=1 is 75.986358
Standard Deviation for 5 folds for K=1 1.431713

Knn classifier for Wine DATA for K=3 five folds cross validation
confusion_matrix =
18 1 0
0 4 3
0 4 5

mean_matrix =

1	2	3	4	5	6	7	8	9	10
71.990	69.195	70.876	67.171	72.467	69.704	70.862	72.004	70.984	72.052

Grand Mean for 5 folds for K=3 is 70.730588
Standard Deviation for 5 folds for K=3 1.638302

3. BreastCancer DATA set

Knn classifier for BreastCancer DATA for K=1 random subsampling
confusion_matrix =
  172 50
  85 43
Mean for K=1 is 60.742857
Standard Deviation for K=1 is 1.915209

Knn classifier for BreastCancer DATA for K=3 random subsampling
confusion_matrix =
  157 85
  60 48
Mean for K=3 is 59.857143
Standard Deviation for K=3 is 1.381773

Knn classifier for BreastCancer DATA for K=1 five folds cross validation
confusion_matrix =
  63 27
  25 24

1	2	3	4	5	6	7	8	9	10
62.330	58.714	61.488	63.521	60.733	63.341	63.648	59.591	62.466	61.190

Grand Mean for 5 folds for K=1 is 61.702201
Standard Deviation for 5 folds for K=1 1.679720

Knn classifier for BreastCancer DATA for K=3 five folds cross validation
confusion_matrix =
  64 27
  34 14

1	2	3	4	5	6	7	8	9	10
58.424	62.207	59.166	61.197	60.587	60.759	60.456	60.884	61.475	62.349

Grand Mean for 5 folds for K=3 is 60.750503

Standard Deviation for 5 folds for K=3 1.222605

DECISION BOUNDARY