

Q1, Q2: (omitted)

Q3: Use Instance Based learning scheme IBk for  $k=1$  and  $k=3$  (both with no distance weighting and with weight  $=1/\text{distance}$ ). Report on the confusion matrix obtained in each case and explain the role of  $k$  and the difference between no distance weighting and weight  $=1/\text{distance}$ .

*open the training dataset*

The screenshot shows the Weka software interface. The 'Preprocess' tab is selected. The 'Open file...' button is circled in red. The 'Filter' section shows 'Resample -5 7 -2 30.0' selected. The 'Current relation' is 'iris-weka.filters.unsupervised.attribute.Remove-R2,4-weka...'. The 'Attributes' section shows a list of attributes: '1 sepallength', '2 petallength', and '3 class'. The 'Selected attribute' section shows 'Name: sepallength', 'Missing: 0 (0%)', 'Distinct: 31', and 'Type: Numeric Unique: 11 (10%)'. A table of statistics is shown below:

Statistic	Value
Minimum	4.6
Maximum	7.9
Mean	6.2
StdDev	0.706

The 'Class: class (Nom)' dropdown is set to 'class (Nom)'. A histogram is displayed at the bottom right, showing the distribution of the 'sepallength' attribute. The x-axis ranges from 4.6 to 7.9, and the y-axis shows the frequency of instances in each bin. The histogram bars are colored red and cyan, with values 9, 20, 25, 34, 10, and 7 labeled above the bars.

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

- weka
  - classifiers
    - bayes
    - functions
    - JythonClassifier
    - lazy
      - IB1
      - IBk**
      - KStar
      - LBR
      - LWL
    - meta
    - mi
    - misc
    - rules
    - trees

```
LinearNNSearch -A {"weka.core.EuclideanDistance -R first-last"}

Classifier output

Name: weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neig
tion: iris-weka.filters.unsupervised.attribute.Remove-R2,4-w
ances: 105
tributes: 3
         sepallength
         petallength
         class
. mode: user supplied test set: size unknown (reading increme

Classifier model (full training set) ===

instance-based classifier
g 1 nearest neighbour(s) for classification

taken to build model: 0 seconds

Evaluation on test set ===
Summary ===

Correctly Classified Instances      42      93.3333 %
Incorrectly Classified Instances    3       6.6667 %
```

Filter... Remove filter Close

*open the test dataset*

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier is set to 'IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A {"weka.core.EuclideanDistance -R first-last}"'. In the 'Test options' section, 'Supplied test set' is selected and circled in orange. A dialog box titled 'weka.gui.GenericObjectEditor' is open, showing the configuration for the 'weka.classifiers.lazy.IBk' classifier. The 'KNN' value is set to 1 and circled in orange. The 'distanceWeighting' dropdown is also circled in orange and set to 'No distance weighting'. Other settings include 'crossValidate: False', 'debug: False', 'meanSquared: False', and 'nearestNeighbourSearchAlgorithm: LinearNNSearch -A "weka.core.Euclid"'. The dialog box has 'Open...', 'Save...', 'OK', and 'Cancel' buttons at the bottom.

**KNN:** the number of nearest neighbors

**distanceWeighting:** Some attribute will be more important than others. and it is usually reflected in the distance metric by some kind of attribute weighting.

# K = 1, no distance weighting

```
Scheme:      weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\"
Relation:    iris-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.instance.RemovePercentage-P30.0
Instances:   105
Attributes:  3
             sepallength
             petallength
             class
Test mode:   user supplied test set:  size unknown (reading incrementally)

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      42          93.3333 %
Incorrectly Classified Instances    3           6.6667 %
Kappa statistic                    0.8916
Mean absolute error                 0.0333
Root mean squared error            0.1061
Relative absolute error             8.0425 %
Root relative squared error        22.1291 %
Total Number of Instances          45

=== Detailed Accuracy By Class ===

             TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
             1       0       1          1       1          1       Iris-setosa
             1       0.086  0.769      1       0.87       1       Iris-versicolor
             0.88    0       1          0.88   0.936     1       Iris-virginica
Weighted Avg.  0.933  0.019  0.949      0.933  0.936     1

=== Confusion Matrix ===

 a  b  c  <-- classified as
10  0  0 | a = Iris-setosa
 0 10  0 | b = Iris-versicolor
 0  3 22 | c = Iris-virginica
```

# K = 1, weight = 1/distance

```
Scheme: weka.classifiers.lazy.IBK -K 1 -W 0 -I -A "weka.core.neighboursearch.LinearMNSearch -A \"weka.core.EuclideanDistance -R first-last
Relation: iris-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.instance.RemovePercentage-P30.0
Instances: 105
Attributes: 3
    sepallength
    petallength
    class
Test mode: user supplied test set: size unknown (reading incrementally)

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      42      93.3333 %
Incorrectly Classified Instances     3      6.6667 %
Kappa statistic                     0.8916
Mean absolute error                  0.0223
Root mean squared error              0.1054
Relative absolute error               5.3936 %
Root relative squared error          21.9824 %
Total Number of Instances           45

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      1      0      1      1      1      1      Iris-setosa
      1      0.086  0.769     1      0.87     1      Iris-versicolor
      0.88     0      1      0.88  0.936     1      Iris-virginica
Weighted Avg.  0.933  0.019  0.949  0.933  0.936     1

=== Confusion Matrix ===

 a  b  c  <-- classified as
10  0  0 | a = Iris-setosa
 0 10  0 | b = Iris-versicolor
 0  3 22 | c = Iris-virginica
```

## K = 3, no distance weighting

```
Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\"
Relation:    iris-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.instance.RemovePercentage-P30.0
Instances:   105
Attributes:  3
             sepallength
             petallength
             class
Test mode:   user supplied test set:  size unknown (reading incrementally)

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      43          95.5556 %
Incorrectly Classified Instances     2           4.4444 %
Kappa statistic                     0.9268
Mean absolute error                  0.0534
Root mean squared error              0.1479
Relative absolute error              12.9213 %
Root relative squared error          30.8357 %
Total Number of Instances           45

=== Detailed Accuracy By Class ===

             TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
             -----  -----  -
             1         0         1           1         1           1         Iris-setosa
             1         0.057    0.833       1         0.909       1         Iris-versicolor
             0.92      0         1           0.92      0.958       1         Iris-virginica
Weighted Avg.  0.956    0.013      0.963      0.956      0.957       1

=== Confusion Matrix ===

 a  b  c  <-- classified as
10  0  0 | a = Iris-setosa
 0 10  0 | b = Iris-versicolor
 0  2 23 | c = Iris-virginica
```

## K = 3, weight = 1/distance

```
Scheme: weka.classifiers.lazy.IBk -K 3 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last
Relation: iris-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.instance.RemovePercentage-P30.0
Instances: 105
Attributes: 3
    sepallength
    petallength
    class
Test mode: user supplied test set: size unknown (reading incrementally)

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      45      100      %
Incorrectly Classified Instances     0         0      %
Kappa statistic                      1
Mean absolute error                  0.0262
Root mean squared error              0.1032
Relative absolute error              6.3465 %
Root relative squared error          21.5226 %
Total Number of Instances           45

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      1         0         1          1         1          1      Iris-setosa
      1         0         1          1         1          1      Iris-versicolor
      1         0         1          1         1          1      Iris-virginica
Weighted Avg.  1         0         1          1         1          1

=== Confusion Matrix ===

  a  b  c  <-- classified as
10  0  0 | a = Iris-setosa
 0 10  0 | b = Iris-versicolor
 0  0 25 | c = Iris-virginica
```

## K = 1, no distance weighting

Given a new instance in the test dataset, the new instance is compared with those instances in the training set using the distance metric and **only the nearest one** is used to assign the class to the new one.

1. Less Information (only one instance)
2. No information on the importance of the neighbors (no weighting on the attribute)

## K = 1, weight = 1/distance

So, as the distance is getting larger, the weighting makes the attribute take smaller effect on the classification.

1. Less Information (only one instance)
2. Information on the importance of the neighbors (weighting= 1/distance)

**No improvement on this case.**

$K = 3$ , no distance weighting

Given a new instance in the test dataset, the new instance is compared with those instances in the training set using the distance metric and only **three** nearest instances are used to assign the class to the new one.

1. More Information (**three** instances)
2. No information on the importance of the neighbors (no weighting on the attribute)

Improvement on this case

$K = 3$ , weight =  $1/\text{distance}$

1. More Information (**three** instance)
2. Information on the importance of the neighbors (weighting=  $1/\text{distance}$ )

So, as the distance is getting larger, the weighting makes the attribute take smaller effect on the classification.

Perfect classification on this case.