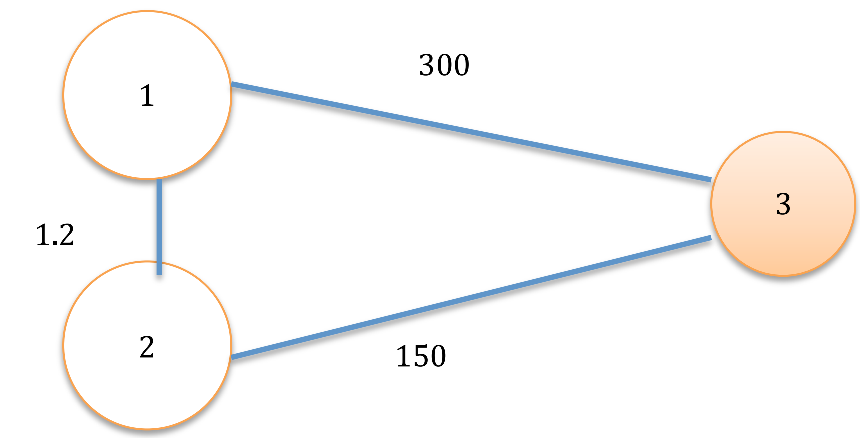
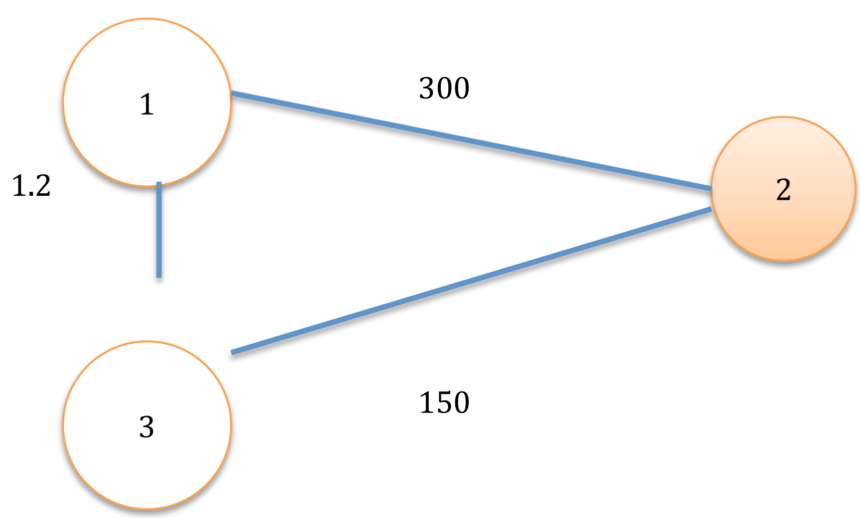
EECS 349 Homework 4

Professor Downey

By: Ted Wu

1.





Theta= 1.1

Q = 2

Consider the shape above. The two circles are a distance of 1.2 apart. The colored circle is 300 and 150 distance away from the other two circles.

For hierarchical clustering, each circle gets its own cluster. On the next iteration, the two closest clusters, the white circles, will merge into one cluster, which is the correct clustering.

For sequential clustering with q=2 and $\Theta = 1.1$ because the maximum number of clusters has been met, the colored circle will be clustered with the closest cluster, which is one of the circles which is wrong.

In the other case where the order is white circle, colored circle, white circle, the first circle will have its own cluster then on the next iteration, the colored circle will have its own cluster since d>1.1, the colored circle will have its own cluster. On the next iteration, the other circle will be clustered with closest cluster and will have a correct clustering.

2.

Clustering Techniques

a. K-means is less applicable to nominal attributes than the other unsupervised learning methods because k-means relies explicitly on Euclidean mean. Nominal values will not be applicable but other methods rely on distance measures

K- means rely on mean, making it the worse learning algorithm.

b. You need to use k-node to fix it. For k-mode, you cluster around data points by the most common responses for each question, and you group them by distance, where distance is number of answers to questions that disappear. In this case, the mode does not establish a mathematical relationship between examples that doesn’t exist, while still allowing you to group similar examples together.

3.

SVMs are used as an easier and efficient way to find boundaries that can separate data to solve classification problems. However sometimes there is inseparable data like a circle in 2d with adjacent dots of alternating colors.

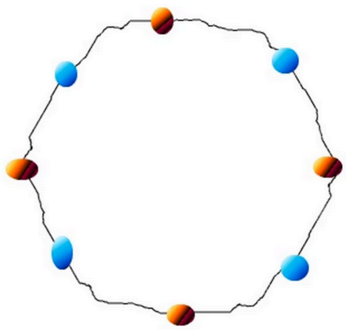


Figure 1. 2d with adjacent dots (inseparable data)

3. Since linear machine can only classify data in linear separable feature space, kernels are important because the kernel functions induce feature space by mapping training data into higher dimensional space where data can be separated. Kernels are valuable because they allow SVMs to transform data. Perceptions are unable to do so, and can only correctly classify datasets that are linearly separable. For example, perceptions cannot classify XOR.

   SVMs also maximize margin around the decision boundary, which is important because by maximizing the margin, points near the decision surface represent more uncertain classification decisions. A classifier with large margin can give you a safety margin so even if there is a slight error in measurement it will not cause a misclassification.

4. There are 1000 examples with the first column as the first attribute, which is a set of random integers from 0-100. The second column is the classification, which is essentially the number in the previous column mod 2 (i.e. 1 if the number is odd,

0 if it is even).

Decision Tree(J48)

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         531               53.1    %
Incorrectly Classified Instances       469               46.9    %
Kappa statistic                          0.0089
Mean absolute error                      0.4951
Root mean squared error                  0.4975
Relative absolute error                 99.317  %
Root relative squared error             99.6482 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.992     0.529     1         0.692      0.504    0
                 0.008     0         1         0.008     0.017      0.504    1
Weighted Avg.    0.531     0.523     0.752     0.531     0.373      0.504
```

Nearest Neighbor(IBk)

```
=== Summary ===

Correctly Classified Instances        1000              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0.0001
Root mean squared error                  0.0002
Relative absolute error                  0.0291 %
Root relative squared error              0.0356 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0          1          1        1          1        0
                 1         0          1          1        1          1        1
Weighted Avg.    1         0          1          1        1          1
```

Multi-layer perceptron

```
=== Summary ===

Correctly Classified Instances         509               50.9    %
Incorrectly Classified Instances       491               49.1    %
Kappa statistic                         -0.0068
Mean absolute error                      0.4997
Root mean squared error                  0.5015
Relative absolute error                100.2276 %
Root relative squared error            100.4459 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.725     0.732      0.525     0.725     0.609      0.495    0
                 0.268     0.275      0.467     0.268     0.341      0.495    1
Weighted Avg.    0.509     0.516      0.497     0.509     0.482      0.495
```

Naïve Bayes

```
=== Summary ===

Correctly Classified Instances        1000              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0.0001
Root mean squared error                  0.0002
Relative absolute error                  0.0291 %
Root relative squared error              0.0356 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0          1          1         1          1        0
                 1         0          1          1         1          1        1
Weighted Avg.    1         0          1          1         1          1
```

For the extra credit, train the 1000 examples using XOR function with first two columns as the XOR inputs and last as the XOR output