Jon Organ

CS677

Homework 12, Seeds

**Question 1:**

Data subset split into training and testing data...

SVM Linear accuracy: 92.86%

SVM Linear confusion matrix:

Predicted 1 2

Actual

1 31 3

2 2 34

SVM Gaussian accuracy: 94.29%

SVM Gaussian confusion matrix:

Predicted 1 2

Actual

1 33 1

2 3 33

SVM Polynomial accuracy: 92.86%

SVM Polynomial confusion matrix:

Predicted 1 2

Actual

1 31 3

2 2 34

**Question 2:**

Naive Bayesian accuracy: 97.14%

Naive Bayesian confusion matrix:

Predicted 1 2

Actual

1 33 1

2 1 35

Model TP FP TN FN accuracy TPR TNR

SVM Linear 31 2 34 3 92.86 0.91 0.94

SVM Gaussian 33 3 33 1 94.29 0.97 0.92

SVM Polynomial 31 2 34 3 92.86 0.91 0.94

Naive Bayesian 33 1 35 1 97.14 0.97 0.97

**Question 3.1:**

Saving Q3 graph...

A graph with a line

Description automatically generated

Using the knee method, the point of diminishing returns seems to be around k = 5

**Question 3.2:**

Saving Q3.2 scatter plot...

A diagram of a scatter plot

Description automatically generated

Looking at multiple iterations of the Q3.2 scatter plot it immediately jumps out at me that most of the feature combinations seem to have a pattern that could be predicted by linear or logistic regression

**Question 3.3:**

Cluster 1

Label: Kama

Centroid: [14.57181818 0.88168864]

Cluster 2

Label: Rosa

Centroid: [15.54555556 0.87975926]

Cluster 3

Label: Canadian

Centroid: [13.01222222 0.84924667]

Cluster 4

Label: Rosa

Centroid: [16.52159091 0.88552955]

Cluster 5

Label: Canadian

Centroid: [13.6812 0.86365]

**Question 3.4:**

The overall accuracy for the new classifier is 82.38%

**Question 3.5:**

The accuracy of the new classifier for just the labels considered by SVM: 73.57%

Confusion matrix:

Predicted 1 2 3

Actual

1 45 0 25

2 12 58 0

Despite fluctuations in accuracy as a result of random initialization for k-means and group selection for the other classifiers, k-means is consistently worse than the other previous classifiers.