Jon Organ

CS677

Homework 9, Diabetes

**Question 1.1:**

File read into data frames...

**Question 1.2:**

Saving Q1.2 0 Table...

A white rectangular object with black border

Description automatically generated

Saving Q1.2 1 Table...

A large number of rectangular boxes

Description automatically generated

**Question 1.3:**

a) Pregnancies and age have the highest correlation for healthy patients.

b) Skin thickness and age have the lowest correlation for healthy patients.

c) Insulin and skin thickness have the highest correlation for unhealthy patients.

d) BMI and pregnancies have the lowest correlation for unhealthy patients.

e) No, the correlated features are entirely different for each case.

**Question 1.4:**

Feature table:

Class μ(F1) σ(F1) μ(F2) σ(F2) μ(F3) σ(F3) μ(F4) σ(F4)

0 109.98 26.14 68.18 18.06 19.66 14.89 68.79 98.87

1 141.26 31.94 70.82 21.49 22.16 17.68 100.34 138.69

All 120.89 31.97 69.11 19.36 20.54 15.95 79.8 115.24

**Question 1.5:**

The mean of each of the features show a clear pattern with the mean of the all class acting as a simple classifier for each feature. The deviations don't show as much of a pattern

**Question 2:**

Does not appear to be in this assignment

**Question 3.1:**

Logistic regression classifier applied...

**Question 3.2:**

kNN classifier applied...

**Question 3.3:**

Naive Bayesian classifier applied...

**Question 3.4:**

Linear Discriminant Analysis classifier applied...

**Question 3.5:**

Quadratic Discriminant Analysis classifier applied...

**Question 3.6:**

Method TP FP TN FN accuracy TPR TNR

Logistic Reg. 224 73 58 29 0.73 0.89 0.44

k-NN (k = 1) 192 67 64 61 0.67 0.76 0.49

k-NN (k = 3) 206 68 63 47 0.7 0.81 0.48

k-NN (k = 5) 218 75 56 35 0.71 0.86 0.43

Naive Bayesian 224 76 55 29 0.73 0.89 0.42

Linear Discr. 222 72 59 31 0.73 0.88 0.45

Quadr. Discr. 224 77 54 29 0.72 0.89 0.41

Question 3.7:

Surprisingly enough all of the different classifiers had reletively similar performances. The largest discrepancy in accuracy is 6%, they all have TPRs of somewhere between 80-90% and TNRs somewhere between 40-50%