FUML Open Assessment

Y3858851

November 26, 2019

Question 1

MLE:
P(Y=0) = 0.648
P(X1=0|Y=0) = 0.6235
P(X1=0|Y=1) = 0.3864
P(X2=0|Y=0) = 0.4753
P(X2=0|Y=1) = 0.767
P(X3=0|Y=0) = 0.8117
P(X3=0|Y=1) = 0.4545

Bayesian:
P(Y=0) = 0.6474
P(X1=0|Y=0) = 0.7463
P(X1=0|Y=1) = 0.2537
P(X2=0|Y=0) = 0.5326
P(X2=0|Y=1) = 0.4674
P(X3=0|Y=0) = 0.7652
P(X3=0|Y=1) = 0.2348

Question 2

I used lasso regression to approximate the data. This is because the data was fit quite well by linear regression so I believe that a linear model is appropriate. Lasso regression was chosen over linear or ridge regression because it does feature selection, and given so many features with very few datapoints the models are prone to overfit. By using lasso egression, some of these features can be ignored which leads to a more robust model. The penalty of the regression was chosen via cross-validation.

Question 3

A logistic regression classifier would be appropriate as the classes are very clearly separated