Mouhammadou Dabo

March 4, 2021

CS 1675: Intro to Machine Learning

Professor Milos Hauskrecht

Problem Assignment 5

Problem 1. Logistic regression model

a. N/A

b.

Confusion matrix for train set

$$\begin{bmatrix} 0.4879 & 0.1410 \\ 0.1651 & 0.2059 \end{bmatrix}$$

Confusion matrix for test set

$$\begin{bmatrix} 0.5284 & 0.1747 \\ 0.0961 & 0.2009 \end{bmatrix}$$

Training misclassification error

$$E = FP + FN$$

$$E = 0.1410 + 0.1651 = 0.3061$$

Testing misclassification error

$$E = FP + FN$$

$$E = 0.1747 + 0.0961 = 0.2707$$

Sensitivity of the model on the test set

$$SN = \frac{TP}{TP + FN} = \frac{0.2009}{0.2009 + 0.0961} = 0.6765$$

Specificity of the model on the test set

$$SP = \frac{TN}{TN + FP} = \frac{0.5284}{0.5284 + 0.1747} = 0.7516$$

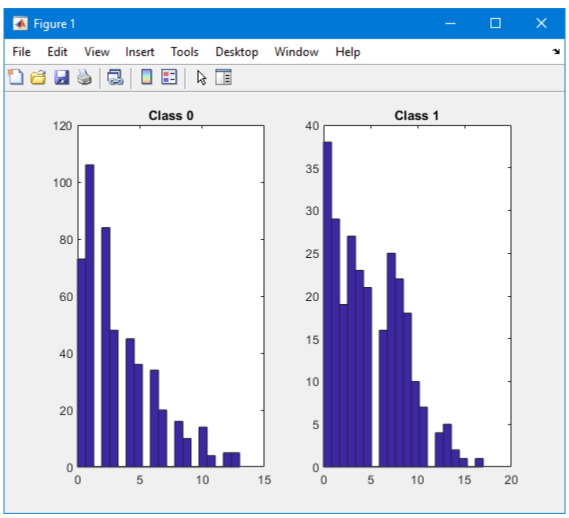
c.

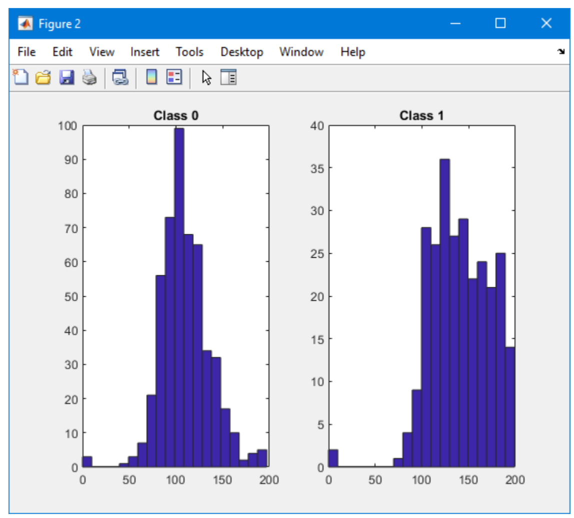
After experimenting with both the number of epochs and the learning rate, I found that the test error was stabilizing around 0.20, while the training error continually got lower to a value of 0.23; this was tested with epochs as high as 30,000, with a constant learning rate of .0001. Changing the initial weights resulted in worse training and test errors and was far less consistent.

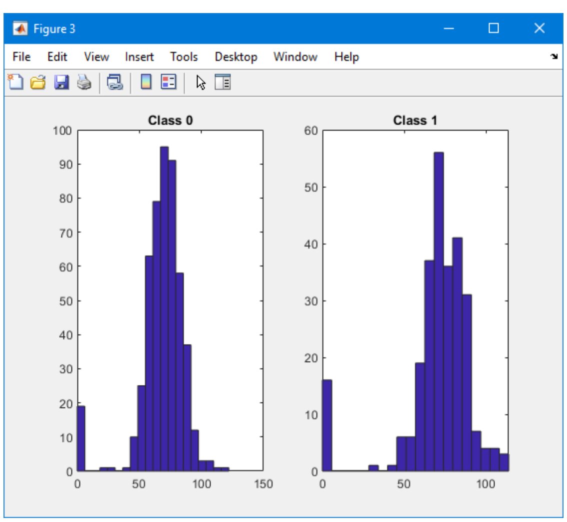
Problem 2. Naïve Bayes model

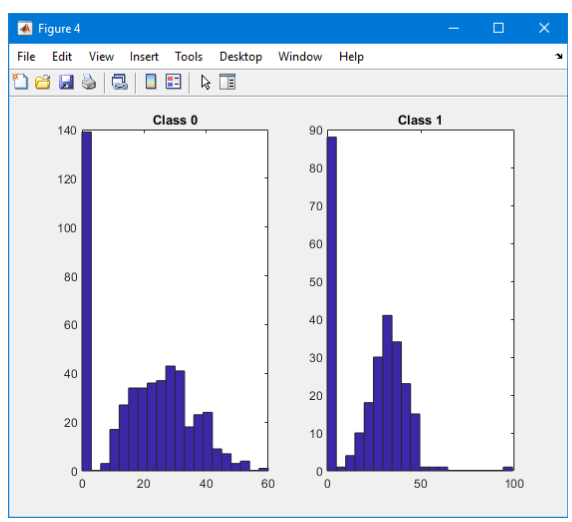
Problem 2.1. Exploratory data analysis

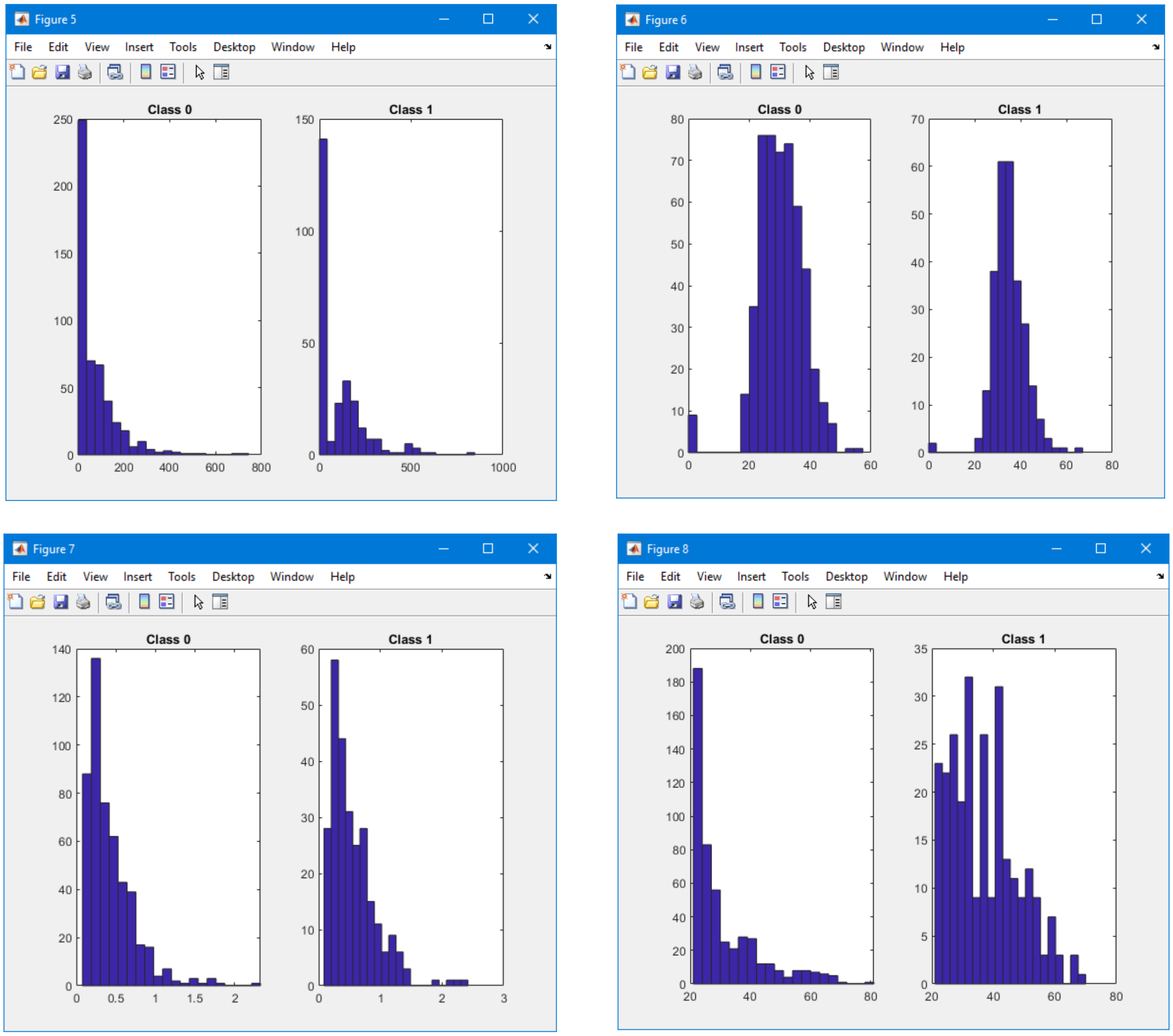
Part a











Histograms for Attributes 1-8 with Class 0 and Class 1

Part b.

The distribution/density to fit the values of attributes 1 to 8 in the pima dataset

Attribute	1	2	3	4	5	6	7	8
Distribution	Gamma	Normal	Normal	Normal	Gamma	Normal	Gamma	Gamma

Problem 2.2. Learning of the Naïve Bayes classifier

Part a.

main2_2.m

Part b.

Exponential estimate for inputs [1 5 7 8]

- Class 0
  - Input 1: 3.2419
  - Input 5: 67.7168
  - Input 7: 0.4164
  - Input 8: 31.1032
- Class 1
  - Input 1: 4.7100
  - Input 5: 103.7200
  - Input 7: 0.5491
  - Input 8: 37.1200

Univariate normal estimate for inputs [2 3 4 6]

- Class 0
  - Input 2: [109.6254, 26.2304]
  - Input 3: [67.5339, 18.6683]
  - Input 4: [19.7316, 14.5828]
  - Input 6: [30.3059, 7.7258]
- Class 1
  - Input 2: [141.3950, 33.6655]
  - Input 3: [70.19, 21.6213]
  - Input 4: [22.935, 17.8275033129112]
  - Input 6: [35.258, 7.3286]

Problem 2.3. Classification with the Naïve Bayes model

Part a.

predict_NB.m

Part b.

Training misclassification error

$$E = FP + FN = 0.2393$$

Testing misclassification error

$$E = FP + FN = 0.2271$$

Confusion matrix for train set

$$\begin{bmatrix} 0.5362 & 0.0928 \\ 0.1466 & 0.2245 \end{bmatrix}$$

Confusion matrix for test set

$$\begin{bmatrix} 0.6026 & 0.1004 \\ 0.1266 & 0.1703 \end{bmatrix}$$

Sensitivity of the model on the test set

$$SN = \frac{TP}{TP + FN} = 0.5735$$

Specificity of the model on the test set

$$SP = \frac{TN}{TN + FP} = 0.8571$$

Part c.

When comparing the results of the Naïve Bayes classifier with the results for the logistic regression model from Problem 1, the Naïve Bayes classifier performed better. The misclassification errors for both the training and testing data are lower than those of the logistic regression model.