**DS 501: STATISTICAL & MATHEMATICAL METHODS FOR DATA SCIENCE**

**ASSIGNMENT 6**

Name: Ali Nauman Roll Number: 18L1863

**QUESTION 1:** How did you map the values of predictions from the perceptron model to labels? Give an exact mathematical expression or pseudo code.

I have taken the absolute difference of prediction value from 2 and 5. Label is selected based on smallest distance from both of the differences:

**if abs(val-2) < abs(val-5):**

**label = 2**

**else:**

**label = 5**

**QUESTION 2:** (ASSUME 2 IS THE POSITVE CLASS. MARKS DEDUCTED IF YOU DO NOT ASSUME THIS) **RESULTS**

|  |  |
| --- | --- |
| **RESULTS FOR TRAINING DATA** | **RESULTS FOR TEST DATA** |
| = 0.001, = 0,iterations = 100  Actual 2 Actual 5 Total  Predicted 2 TP = **145** FP = **4** **149**  Predicted 5 FN = **5** TN = **121** **126**  Total -> **150** **125** **275**  BAC = **0.967** | = 0.001, = 0,iterations = 100  Actual 2 Actual 5 Total  Predicted 2 TP = **103** FP = **32** **135**  Predicted 5 FN = **47** TN = **93** **140**  Total -> **150** **125** **275**  BAC = **0.715** |
| = 0.7, = 0.9,iterations = 100  Actual 2 Actual 5 Total  Predicted 2 TP = **NAN** FP = **NAN NAN**  Predicted 5 FN = **NAN** TN = **NAN NAN**  Total -> **NAN** **NAN NAN**  BAC = **NAN** | = 0.1, = 0.5,iterations = 100  Actual 2 Actual 5 Total  Predicted 2 TP = **NAN** FP = **NAN NAN**  Predicted 5 FN = **NAN** TN = **NAN** **NAN**  Total -> **NAN** **NAN** **NAN**  BAC = **NAN** |
| = **0.0001** = **0.1**, iterations = **100**  Actual 2 Actual 5 Total  Predicted 2 TP = **111** FP = **33** **144**  Predicted 5 FN = **39** TN = **92** **131**  Total -> **150** **125** **275**  BAC = **0.738** | = **0.0001** = **0.1**, iterations = **100**  Actual 2 Actual 5 Total  Predicted 2 TP = **96** FP = **47** **143**  Predicted 5 FN = **54** TN = **78** **132**  Total -> **150** **125** **275**  BAC = **0.632** |
| = **0.001** = **0.9**, iterations = **100**  Actual 2 Actual 5 Total  Predicted 2 TP = **150** FP = **0** **150**  Predicted 5 FN = **0** TN = **125** **125**  Total -> **150** **125** **275** BAC = **1** | = **0.001** = **0.9**, iterations = **100**  Actual 2 Actual 5 Total  Predicted 2 TP = **108** FP = **15** **123**  Predicted 5 FN = **42** TN = **110** **152**  Total -> **150** **125** **275** BAC = **0.8** |

**QUESTION 3:** Give YOUR opinion or conclusions about the results

As we can see from the results above, increase in the learning rate causes the **BAC** accuracy to decrease. As we move towards a higher value of learning rate (e.g. 0.7), it gives us **NAN** values. Lower learning rate allows us to take small steps. Increasing momentum (e.g. 0.9), gives us a better **BAC** on the test set, as it is displayed from the above results. Number of iterations have been kept constant for consistency.