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

**ASSIGNMENT 4**

Name: Samama Imtiaz Butt Roll Number: 18L-1882

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

In this assignment, I have mapped the training label 5 to -1 and 2 to 1 respectively. If the predicted value is greater than 0 then the label would be 2 otherwise 5. Its pseudo code is as follows:

#Replace the Train Labels with -1 and 1

If(tain\_label == 2){

train\_label = 1

}else{

train\_label = -1}

#Suppose our predictions are store in prediction variable

If(prediction < 0{

Print (“Lable is 5 ”)

}else{

Print (“Lable is 2 ”)}

Another way of doing this is to set a threshold of 3.5 so if the prediction is less than 3.5 predict the label 2 otherwise 5. In this case, we don’t need to map the labels to -1 and 1.

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

**RESULTS**

|  |  |  |
| --- | --- | --- |
| **RESULTS FOR TRAINING DATA** | **RESULTS FOR TEST DATA** | **Predicted value of y for test data** |
| Lambda = 0.001  Actual 2 Acutal5 Total  Predicted 2 TP = 150 FP = 0 150  Predicted 5 FN = 0 TN = 125 125  Total -> 150 125 275 | Lambda = 0.001  Actual 2 Actual 5 Total  Predicted 2 TP =140 FP =36 175  Predicted 5 FN= 10 TN=89 100  Total -> 150 125 275 | Row 1 = 2  Row 10 = 5  Row 15 = 5  Row 100 = 5  Row 120 = 2  Row 200 = 5 |
| Lambda = 1  Actual 2 Actual 5 Total  Predicted 2 TP =150 FP =0 150  Predicted 5 FN =0 TN =125 125  Total -> 150 125 275 | Lambda = 1  Actual 2 Actual 5 Total  Predicted 2 TP =142 FP =8 150  Predicted 5 FN =8 TN =117 125  Total -> 150 125 275 | Row 1 = 2  Row 10 = 5  Row 15 = 5  Row 100 = 5  Row 120 = 5  Row 200 = 5 |
| Lambda = 1000  Actual 2 Acutal5 Total  Predicted 2 TP = 150 FP = 9 159  Predicted 5 FN = 0 TN = 116 116  Total -> 150 125 275 | Lambda = 1000  Actual 2 Actual 5 Total  Predicted 2 TP = 149 FP = 14 163  Predicted 5 FN = 1 TN =111 112  Total -> 150 125 275 | Row 1 = 2  Row 10 = 5  Row 15 = 5  Row 100 = 5  Row 120 = 5  Row 200 = 5 |
| Lambda = 10 (should have a good result)  Actual 2 Actual 5 Total  Predicted 2 TP =150 FP = 0 150  Predicted 5 FN = 0 TN = 125 125  Total -> 150 150 275 | Lambda = 10  Actual 2 Actual 5 Total  Predicted 2 TP = 147 FP = 4 151  Predicted 5 FN = 3 TN = 121 124  Total -> 150 125 275 | Row 1 = 2  Row 10 = 5  Row 15 = 5  Row 100 = 5  Row 120 = 5  Row 200 = 5 |

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

From the results above we can observe that if we take lambda very small (0.001) then our model over fits the training data if it is large (1000) then it under fits the test data. So, we have to find the value of lambda which neither under fits nor over fits the data which is in this case is 10.