Machine Learning Binary classifier home work

Due: Start of class Feb

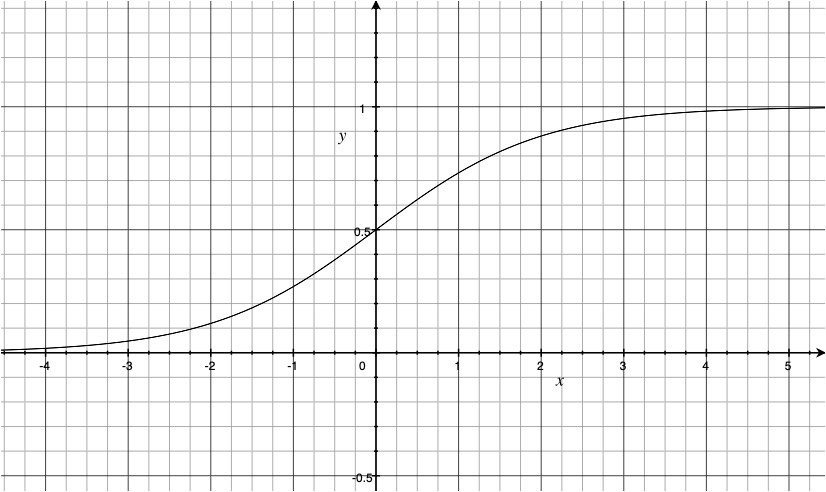
**DATA SET 1**

|  |  |
| --- | --- |
| Value | Type |
| 1 | Red |
| 1 | Red |
| 2 | Red |
| 4 | Blue |
| 5 | Red |
| 7 | Blue |
| 10 | Blue |
| 14 | Blue |
| 15 | Blue |
| 16 | Blue |

1. Consider a binary classifier with one parameter. Prediction is Blue if the value is greater than 3.
   1. Create the confusion matrix for data set 1.
   2. Compute accuracy, precision, sensitivity, and specificity.
   3. Compute the F and G score
2. Consider a binary classifier with one parameter. Prediction is Blue if the value is greater than 6.
   1. Create the confusion matrix for data set 1.
   2. Compute accuracy, precision, sensitivity, and specificity.
   3. Compute the F and G score
3. Which of the previous classifiers would you prefer for the data set 1? Explain your choice.
4. The sigmoid function is defined as



It has an S shape and limits to 0 and 1.



This makes it suitable to use as a classifier with two parameters that returns a probability. W specifies the steepness of the transition. (Larger W is steeper). K specifies the middle of sigmoid. (Where the probability is 0.5)



a) Fill in the following table using . If p is > ½ predict Blue

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | Type | probability | prediction | Log-Loss |
| 1 | Red |  |  |  |
| 1 | Red |  |  |  |
| 2 | Red |  |  |  |
| 4 | Blue |  |  |  |
| 5 | Red |  |  |  |
| 7 | Blue |  |  |  |
| 10 | Blue |  |  |  |
| 14 | Blue |  |  |  |
| 15 | Blue |  |  |  |
| 16 | Blue |  |  |  |

b) Compute the average log-loss

c) Fill in the following table using . If p is > ½ predict Blue

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | Type | probability | prediction | Log-Loss |
| 1 | Red |  |  |  |
| 1 | Red |  |  |  |
| 2 | Red |  |  |  |
| 4 | Blue |  |  |  |
| 5 | Red |  |  |  |
| 7 | Blue |  |  |  |
| 10 | Blue |  |  |  |
| 14 | Blue |  |  |  |
| 15 | Blue |  |  |  |
| 16 | Blue |  |  |  |

d) Compute the average log-loss

e) Based on the average log-loss, which classifier do you prefer?

**DATA SET 2**

|  |  |
| --- | --- |
| Value | Type |
| 1 | Red |
| 1 | Red |
| 2 | Red |
| 3 | Blue |
| 4 | Red |
| 5 | Blue |
| 7 | Blue |
| 10 | Blue |
| 14 | Red |
| 15 | Blue |
| 16 | Red |
| 19 | Red |
| 20 | Red |

1. Consider a binary classifier with two parameters. Prediction is Red if the value is less than 3 or greater than 15. Blue, otherwise.
   1. Create the confusion matrix for data set 2.
   2. Compute accuracy, precision, sensitivity, and specificity.
   3. Compute the F and G score