# Machine Learning HW 1 Binary classifier Due Tuesday, October 4 at end of day.

**Quiz on Thursday, October 6**

## DATA SET 1

|  |  |  |  |
| --- | --- | --- | --- |
| Width times Height | Type | Predict >6 | Predict >9 |
| 2 | Small | Small | Small |
| 3 | Small | Small | Small |
| 4 | Small | Small | Small |
| 5 | Large | Small | Small |
| 6 | Small | Small | Small |
| 8 | Large | Large | Small |
| 11 | Small | Large | Large |
| 12 | Large | Large | Large |
| 15 | Large | Large | Large |
| 17 | Large | Large | Large |

1. Consider a binary classifier with one parameter. Prediction is Large if the feature is greater than 6.
   1. Create the confusion matrix for data set 1.

|  |  |  |
| --- | --- | --- |
|  | Instance is positive | Instance is negative |
| Model output  is positive | 4 | 1 |
| Model output  is negative | 1 | 4 |

* 1. Compute accuracy, precision, sensitivity, and specificity.

Accuracy = (tp+tn)/sum = (4+4)/10 = 0.8

Precision = tp/(tp+fp) = 4/(4+1) = 0.8

Sensitivity = tp/(tp+fn) = 4/(4+1) = 0.8

Specificity = tn/(tn+fp) = 1/(4+1) = 0.8

* 1. Compute the F and G score

F Score: 2/ ((1/Precision) + (1/Sensitivity)) => 2/ ((1/0.8) + (1/0.8)) = 0.8

G Score: => = 0.8

1. Consider a binary classifier with one parameter. Prediction is Large if the feature is greater than 9.
   1. Create the confusion matrix for data set 1.

|  |  |  |
| --- | --- | --- |
|  | Instance is positive | Instance is negative |
| Model output  is positive | 3 | 1 |
| Model output  is negative | 2 | 4 |

* 1. Compute accuracy, precision, sensitivity, and specificity.

Accuracy = (tp+tn)/sum = (3+4)/sum = 0.7

Precision = tp/(tp+fp) = 3/(3+1) = 0.75

Sensitivity = tp/(tp+fn) = 3/(3+2) = 0.6

Specificity = tn/(tn+fp) = 4/(4+1) = 0.8

* 1. Compute the F and G score

F Score: 2/ ((1/Precision) + (1/Sensitivity)) => 2/ ((1/0.75) + (1/0.6)) = 0.66

G Score: => = 0.67

1. Which of the previous classifiers would you prefer for the data set 1? Explain your choice.

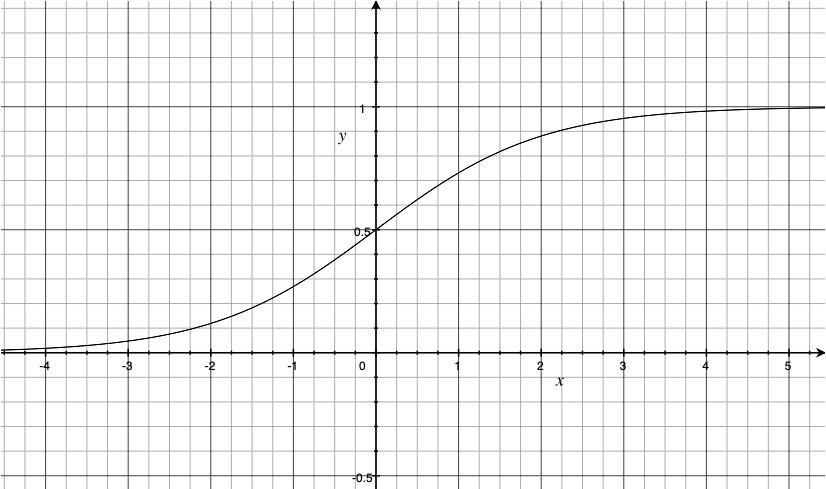
I would prefer the 1st classifier whose Prediction is Large if the feature is greater than 6 , as we should choose the classifer with highest Accuracy and highest F Score

1. The sigmoid function is defined as

*ex*

*S* *x*  1 *ex*

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)

*eW* *x**K* 

*p* *x*,*W* , *K*  

1

* *eW* *x**K* 

1. Fill in the following table using

*p* *x*,2,6 

1

2*x*6

. If p is > ½ predict Large. Use

*e*

 *e*2*x*6

the p value to compute the log-loss for each instance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Width  times Height | Type | **probability** | **prediction** | **Log-Loss** |
| 2 | Small | 0.00033535013046647816 | Small | 0.00033540637289577373 |
| 3 | Small | 0.0024726231566347748 | Small | 0.002475685137730443 |
| 4 | Small | 0.017986209962091555 | Small | 0.01814992791780973 |
| 5 | Large | 0.11920292202211755 | Small | 2.1269280110429727 |
| 6 | Small | 0.5 | Small | 0.6931471805599453 |
| 8 | Large | 0.9820137900379085 | Large | 0.01814992791780973 |
| 11 | Small | 0.9999546021312976 | Large | 10.000045398900186 |
| 12 | Large | 0.9999938558253978 | Large | 6.144193477725374e-06 |
| 15 | Large | 0.9999999847700205 | Large | 1.5229979615740706e-08 |
| 17 | Large | 0.9999999997210532 | Large | 2.7894675462274766e-10 |

1. Compute the average log-loss

Average : 1.28592377

1. Fill in the following table using. 𝑝(𝑥, 2,9) = !!(#$%)

!(#$%)

"#!

If p is > ½ predict Large

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Width times Height | Type | **probability** | **prediction** | **Log-Loss** |
| 2 | Small | 8.315280276641321e-07 | Small | 0.0001 |
| 3 | Small | 6.144174602214718e-06 | Small | 1e-05 |
| 4 | Small | 4.5397868702434395e-05 | Small | 5e-05 |
| 5 | Large | 0.00033535013046647816 | Small | 8.00034 |
| 6 | Small | 0.0024726231566347748 | Small | 0.00248 |
| 8 | Large | 0.11920292202211755 | Small | 2.12693 |
| 11 | Small | 0.9820137900379085 | Large | 4.01815 |
| 12 | Large | 0.9975273768433652 | Large | 0.00248 |
| 15 | Large | 0.9999938558253978 | Large | 1e-05 |
| 17 | Large | 0.9999998874648379 | Large | 0.0001 |

1. Compute the average log-loss

Average Loss = 1.415065

1. Based on the average log-loss, which classifier do you prefer?

Based on the calculated average log loss I would prefer the classifier with W=2 and K=6 as it is having minimum average log loss.

## DATA SET 2

|  |  |  |
| --- | --- | --- |
| Value | Type | Predict |
| 1 | Bike | Bike |
| 3 | Scooter | Bike |
| 5 | Bike | Bike |
| 7 | Bike | Bike |
| 9 | Bike | Bike |
| 11 | Scooter | Scooter |
| 13 | Bike | Scooter |
| 15 | Bike | Scooter |
| 17 | Scooter | Scooter |
| 19 | Bike | Bike |
| 21 | Scooter | Bike |
| 22 | Bike | Bike |
| 23 | Scooter | Bike |

1. Consider a binary classifier with two parameters. Prediction is Bike if the value is less than 10 or greater than 18. Scooter, otherwise.
   1. Create the confusion matrix for data set 2.

|  |  |  |
| --- | --- | --- |
|  | Instance is positive | Instance is negative |
| Model output  is positive | 6 | 3 |
| Model output  is negative | 2 | 2 |

* 1. Compute accuracy, precision, sensitivity, and specificity.

Accuracy = (tp+tn)/sum = (6+2)/13 = 0.615

Precision = tp/(tp+fp) = 6/(6+3) = 0.666

Sensitivity = tp/(tp+fn) = 6/(6+2) = 0.75

Specificity = tn/(tn+fp) = 2/(2+2) = 0.5

* 1. Compute the F and G score

F Score: 2/ ((1/Precision) + (1/Sensitivity)) => 2/ ((1/0.666) + (1/0.75)) = 0.702

G Score: => 6 = 0.703

I assumed that Bike was the positive and Scooter was the negative.