Binary Classification Metrics

Dr. Kalidas Y., IIT Tirupati

By the end of this lecture, you will understand what are the common metrics in a classification problem

51) key phrase.. “Positive class”

- The class of data points ***that is of our interest***
- You define it!

52) key phrase.. “Negative class”

- The class of data points of that is ***not of our interest***
- You define it!

53) key phrase… “True Positive”

- The ground truth data corresponding to “positive class’

54) key phrase… “True Negative”

- The ground truth data corresponding to “negative class’

55) key phrase… “Predicted Positive”

- The predictions that are “positive”

56) key phrase… “Predicted Negative”

- The predictions that are “negative”

57) key phrase… “False Positive”

- The predictions that are “positive”, but actual ground truth says they are “negative”

57) key phrase… “False Negative”

- The predictions that are “negative”, but actual ground truth says they are “positive”

58) key phrase… “Precision”

- “true positives” ÷ “positive predictions”

59) key phrase… “Recall”

- “true positives among positive predictions” ÷ “true positives”

60) key phrase… “Correct Predictions”

- true positives among predictive positives and
- true negatives among predicted negatives

61) key phrase… “Wrong Predictions”

- False positives among predictive positives and
- False negatives among predicted negatives

62) key phrase… “Accuracy”

- Correct Predictions ÷ Total number of points in data set

63) key phrase… “Confusion Matrix”

- Human to Decide
  - on which yi = 1 (positive) and
  - on which yj = 0 (negative)
- Get a Data Set
  - x1 = (1,1), y1 = 1
  - x2 = (1,2), y2 = 1
  - x3 = (1,3), y3 = 1
  - x4 = (2,1), y4 = 0
  - x5 = (3,1), y5 = 0
- Model Predictions
  - y1’ = 1
  - y2’ = 1
  - y3’ = 0
  - y4’ = 1
  - y5’ = 0

- Correct predictions
  - True Positives (TP) = {1, 2}
  - True Negatives (TN) = {5}
- Wrong predictions
  - False Positives (FP) = {4}
  - False Negatives (FN) = {3}
- Total = 5 points

	Pred +	Pred -	
Given +	TP = 2	FN = 1	**Recall = TP/(TP+FN)**
Given -	FP = 1	TN = 1	
	Precision = TP/(TP+FP)		**Accuracy = (TP+TN)/Total**

F1 score = $\frac{2\,p*r}{(p+r)}$

64) key phrase… “Threshold on Score”

- Human to Decide
  - on which yi = 1 (positive) and
  - on which yj = 0 (negative)
- Get a Data Set
  - x1 = (1,1), y1 = 1
  - x2 = (1,2), y2 = 1
  - x3 = (1,3), y3 = 1
  - x4 = (2,1), y4 = 0
  - x5 = (3,1), y5 = 0
- Model Predictions
  - y1’ = 1
  - y2’ = 1
  - y3’ = 0
  - y4’ = 1
  - y5’ = 0

- Human to Decide
  - on which yi = 1 (positive) and
  - on which yj = 0 (negative)
- Get a Data Set
  - x1 = (1,1), y1 = 1
  - x2 = (1,2), y2 = 1
  - x3 = (1,3), y3 = 1
  - x4 = (2,1), y4 = 0
  - x5 = (3,1), y5 = 0
- Model Prediction Scores
  - y1_score = 0.9
  - y2_score = 0.7
  - y3_score = 0.3
  - y4_score = 0.8
  - y5_score = 0.4

Choose a Threshold

Above which are predicted +

below which are predicted -

For example 0.55 ???

We will see more details in the next lecture…