Problem 3 - Which Algorithm is Better? 10 points

- Mateus Silva Aragao , msa8779

You want to compare two algorithms for document retrieval. You need to answer this problem manually, showing explicit calculations. The ground truth and performance of the two algorithms is shown below for 100 samples (with relevant being positive and irrelevant being negative class):

1. Create confusion matrix for the two algorithms showing TP, FP, FN, TN. Note you need to compare ground truth labels from reference with corresponding labels from different algorithms to count these quantities. Follow the example discussed in class.

Answer:

For the First Algorithm:

	True positive value	true negative value
Predicted Positive Value	TP=25	FP=15
Predicted Negative Value	FN=30	TN=55

For the Second Algrithm:

	True positive value	true negative value
Predicted Positive Value	TP=20	FP=10
Predicted Negative Value	FN=10	TN=60

2. You are interested in finding the algorithm which has better performance on the negative classes. Your friend suggests to use Balanced accuracy instead of accuracy to identify the best algorithm. Your instructor suggests to use F-1 score instead. Who is right here and why? Support your answer with numbers.

Answer:

Given that our dataset is imbalanced with a larger number of negative lables. Using the F1 score is a better metric than using the Accuracy. Therefore the professor is correct.

F1 - score for algorithm 1 F1 = 25/(25 +(15+5)/2) = .714

F1 - score for algorithm 2 F1 = 20/(20 +(10+10)/2) = .66

Therefore, the algorithm 1, which has the highest F1 score, is the best algorithm.

3. Did the advise of your friend or your instructor helped you in identifying the right algorithm? If yes, you are good. If not, explain why the metrics suggested by them did not work.

Answer:

The advice of the professor helped with identifying the best algorithm.

4. List all the metric(s) do you think will give help you make the right selection? Check each algorithms ROC, Precision-Recall curve and the F1 -scores would be good metrics.