

Due October 13, 2017.

Turn in your homework

- as a PDF file
- named HW03_<LastName>_<FirstName>.pdf (no accents)
- at <http://tinyurl.com/ma2823-2017-hw>

Question 1

A naive binary classification algorithm always returns “positive” as an answer. We evaluate it on a data set containing 5 times more positive than negative examples. What are its precision and recall?

Solution:

- The proportion of positive examples is $5/6$ and that of negative examples $1/6$.
- The algorithm predicts that all samples are positive, hence has a true positive rate (= precision) of $5/6$.
- All samples are predicted positive, so by extension all positive samples are predicted positive, hence the recall is 100%.

Question 2

You have to implement a fraud detection algorithm for a bank. Undetected frauds are quite costly to the bank, compared to establishing that a transaction was, in fact, not fraudulent. What do you want to minimize:

- False positive rate
- False negative rate
- True positive rate
- Number of errors.

Solution: False negatives are the most expensive, so one should focus on minimizing that.