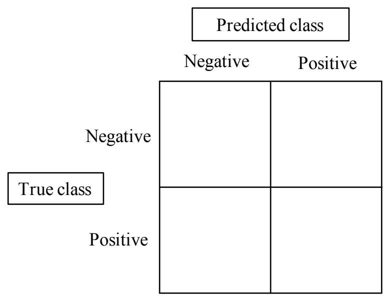
Evaluation Metrics *(find and work with a partner)*

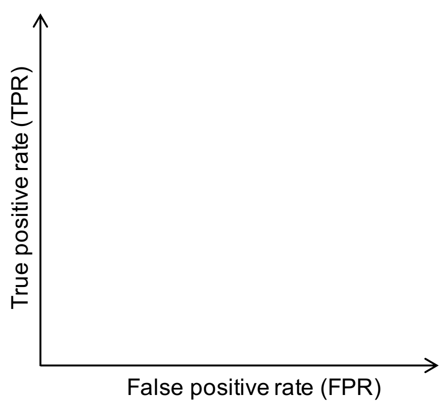
Say you google "ROC", wanting documents about ROC curves. Say there are 100 total documents to search through, and 10 are returned to you. Of these, 7 are about ROC curves. In reality, there are 20 documents about ROC curves.

1. Complete the confusion matrix below, adding the sum of each row to the right and the sum of each column to the bottom.



2. What is the precision? The recall (true positive rate)?

3. Compute the false positive rate and then plot this result on the ROC curve axes below.



4. With the same query, another search engine gives you 30 documents, 18 of which are about ROC curves. Add this point to your ROC curve above.

5. Add the other two points that are always on a ROC curve. In words, how would you describe the classifications strategies of these two points?

Bayesian Probability *(find and work with a partner)*

Clinical Trials Example. Say the probability of a disease (D) in the general population is 1 in 100, i.e. $P(D) = \frac{1}{100}$. This is our *prior* probability of the disease (i.e. without any data).

Furthermore, say we have test for this disease with 90% accuracy. We will call the results of the test positive ("pos") and negative ("neg"). 90% accuracy means that $P(\text{pos}|D) = \frac{9}{10}$, and $P(\text{neg}|H) = \frac{9}{10}$, where H means healthy.

What we actually want to know is: what is the probability of having the disease, given a positive test?

1. Apply Bayes rule to $P(D|\text{pos})$. Recall that Bayes rule says:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

2. We can often write the denominator as the sum of $P(a, B)$, for all options $a \in \text{vals}(A)$:

$$P(A|B) = \frac{P(A)P(B|A)}{\sum_{a\in \text{vals}(A)} P(a, B)}.$$

Use this idea to expand the denominator in the clinical trials example and compute a numerical value for $P(D|\text{pos})$.