1. A binary classifier was evaluated using a set of 1,000 test examples in which 50% of all examples are negative. It was found that the classifier has 60% sensitivity and 70 % accuracy. Write the confusion matrix.
2. Using the confusion matrix created in the previous exercise, compute the classifier’s precision, F1-measure, and specificity.
3. To test a binary classifier, a data set consisting of 100 positive and 400 negative examples was used. It turned out that the ROC curve goes through the point TPR = FPR = 0.2. Calculate Precision at this point.
4. Imagine we are classifying dollar bills as authentic (0) or counterfeit (1). Suppose we encounter many more authentic bills than counterfeit bills, e.g., 99% of bills are authentic and 1% are counterfeit.
   1. If we used a naïve algorithm for classification that always classified every bill as authentic, what accuracy would we have? Can you commit the result?
   2. Suppose for one thousand bills, a classifier predicts 30 to be counterfeit and 970 to be authentic. Of the 30 counterfeit predictions, 20 are actually counterfeit. Of the 970 authentic predictions, 965 are actually authentic. Fill in a confusion matrix.
   3. What does recall and precision mean in this case?
   4. What does high recall and high precision mean in this case?
   5. What is the recall and what is the precision associated with our counterfeit bill classifier?
   6. It is desirable for a classifier to have high recall and high precision. However, there is often a trade-off between the two, where a classifier’s recall can be improved at a cost to its precision, or the classifier’s precision can be improved at a cost to its recall. For a given classifier, if you tweak its parameters so as to increase the number of positive predictions that the classifier makes (in our example, you cause the classifier to predict a larger number of bills as counterfeit), then which of the following are likely:
      1. Recall will increase and precision will decrease
      2. Precision will increase and recall will decrease
   7. What is the F1 score for our counterfeit bill classifier above?
   8. Instead, suppose we use a naïve classifier that predicts all 1000 bills are counterfeit even though only 30 are actually counterfeit. What is the F1 score for this naïve classifier?  
       What can we conclude about this classifier?
5. Here is a covid test example: In a Covid test of 1000 patients, there were 45 positive tests, of which 30 patients had covid and 15 were falsely tested positive. Of the 955 negative tests there were 5 that were incorrect, these patients had covid but were tested negatively. Draw the confusion matrix and calculate the accuracy, precision, recall, sensitivity and F1 score from the matrix.
6. Ball example. Imagine we are playing a game where you have to guess if the next item is a ball or not. You are aware it happens around half the time. You make a guess, the item is a ball or nor a ball, and you are awarded one or zero points. Here is an example dataset of results of 10 guesses, the correct answers and points awarded. The prediction ball was made 4 times and it was correct 3 out of those 4 times. The prediction no ball was made 6 times, with 3 out of those 6 attempts correct. What is the resulting confusion matrix? Calculate the accuracy, precision, recall, sensitivity and F1 score from the matrix.
7. The table below provides a training data set containing five examples, a feature X, and a qualitative target attribute Y. What is the variance of the target attribute by the linear regression on x if the regression parameter estimates are 0 = 5 and 1 = −1?
8. 20 cases. In exercise 1 there are only 20 cases where 8 patients are diagnosed correctly as positive and 4 incorrectly. There were 8 patients diagnosed with a negative result, 5 correctly and 3 incorrectly. Complete the confusion matrix first, and then calculate the evaluation metrics.
9. 100 cases. In exercise 2 there are 100 cases where only 5 positive cases were found, and there was only one case that was incorrectly negative. Of the five positive cases, the actual results showed 3 correct and 2 incorrect cases.
10. 100 samples. In exercise 3 there are also 100 samples. Of the 60 positive samples there were 45 correctly identified positive cases, whilst there were also 35 correctly identified negative cases.
11. 128 tests. In exercise 4, of the 75 positive tests, 9 were false, whilst the 53 negative tests included 22 that were false.
12. 200 tests. In exercise 5, of the 122 positive tests, only 2 were false, whilst the 78 negative tests included only 5 that were false.