

Assignment 2

Classification of results is very important in medical domain. We also need to know how well the decision boundary and the results work. Thus, we need to find a test to determine the accuracy of classification. There are two methods to decide that

- 1) Confusion Matrix
- 2) Sensitivity & Specificity

Latter is usually used to classify just two classes which is usually the case in medical domain. Its either the test is positive or negative. Hence, it becomes a good method to determine the accuracy of the tests based on two parameters sensitivity and specificity.

1. Sensitivity – measures the proportion of positives that are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).
2. Specificity- measures the proportion of negatives that are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition).

In this assignment, we were allocated two tasks

- 1) Calculate the sensitivity and specificity with decision boundary 1.5
- 2) Calculate the sensitivity and specificity with varying decision boundary from -1 to 5 and plot ROC.

Implementation of both the tasks were done in two different matlab scripts.

Task 1:

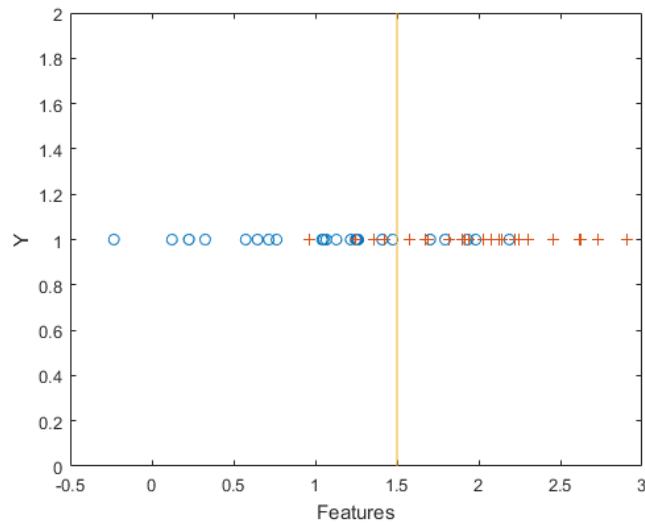
1. Data set was generated using randn with given specifications
2. Boundary was initialized to 1.5
3. Now loop was iterated for 25 samples to calculate TP, TN, FP, FN.
4. Sensitivity and specificity were calculated based on the formula given in the book.

Expectations:

The result was given so no other expectations. Just that, we got a clearer idea of values of sensitivity and specificity.

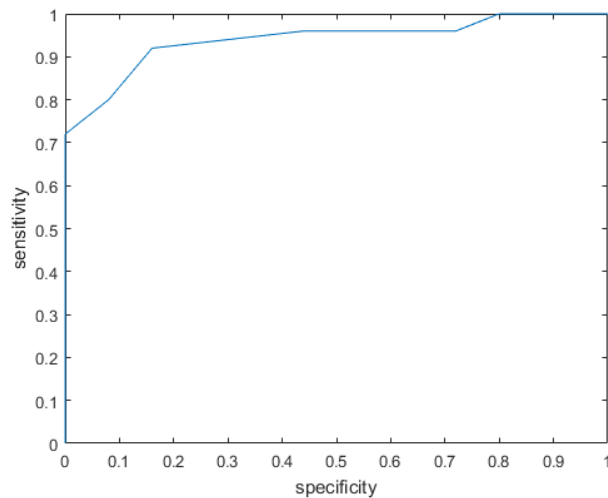
Results:

1. In this task, we plotted the graph below with two random variable set and boundary line and found the sensitivity and specificity.
2. In the graph below Y axis is not important and the important data is varying on X axis. Threshold $x_b = 1.5$. Circle elements to the left of boundary line are true positives while ones to the right are false negatives. Similarly, + elements to the right of x_b are True negatives while once to the right are False positives.
3. Sensitivity and specificity were calculated based on above parameters with formula given in the book.
4. Obviously, as the feature set changes, all the parameters change as well.



Task 2:

1. Data set was generated using randn with given specifications
2. Boundary was initialized as a variable ranging from -1 to 5 in steps 0.2
3. Then, two loops were defined with one loop inside the other.
4. Outer loop iterating over the steps in decision boundary and inner loop iterating over samples
5. Sensitivity and specificity for boundaries -1 to 5 were calculated and stored in an array.
6. Then, we plot ROC (Sensitivity vs Specificity).



Expectations:

I expected the curve to be smoother but the results were same as expected.

Results:

1. In this task, sensitivity and specificity for $x_b = -1$ to 5 were plotted.

2. After careful observation, we find out that when all the elements are to the right of the boundary, sensitivity = 0 and specificity = 1. When all elements are to the left, sensitivity = 1 and specificity = 0
3. Both the values increase eventually and while plotting sensitivity v/s specificity we get a graph as above.