Introduction to Intelligent Systems
Lab week 1

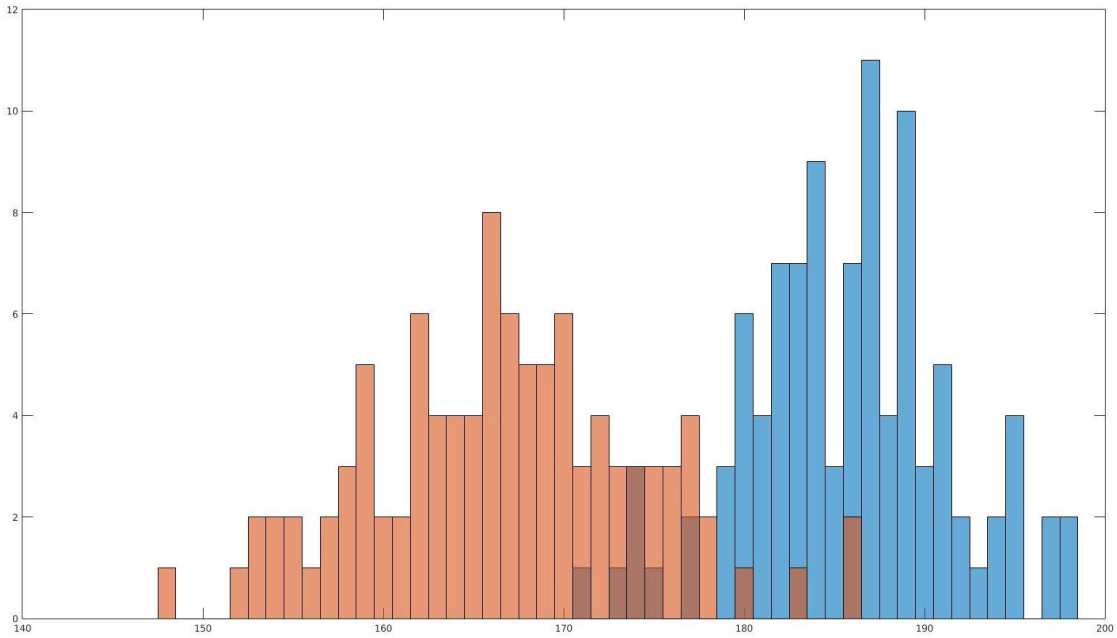
Diego Velasco Volkmann S2851059 Eduardo Faccin Vernier S3012875

October 31, 2015

Assignment 1:

Given are two sets of measured lengths (in cm) of men (length_men) and women (length_women) in the file lab1_1.mat.

- Plot histograms of both sets in one figure.



- Now choose the decision criterion at 170cm. How many men are classified incorrect? And how many women?

  0 men and 29 women, since no men has less than 170cm, but 29 women have more than 170cm.

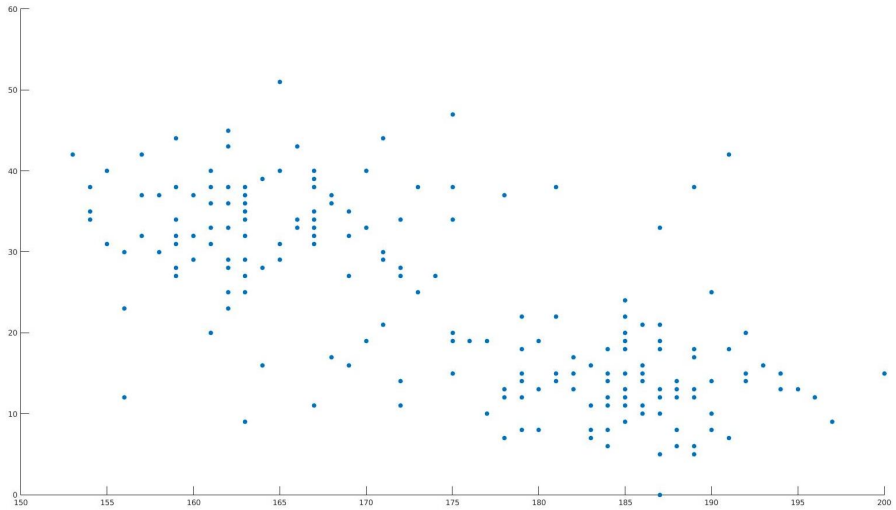
- What decision criterion should be used to minimize total number of misclassifications (sum over men and women)?

  177cm. Giving a total of 14 misclassifications (6 men and 8 women).

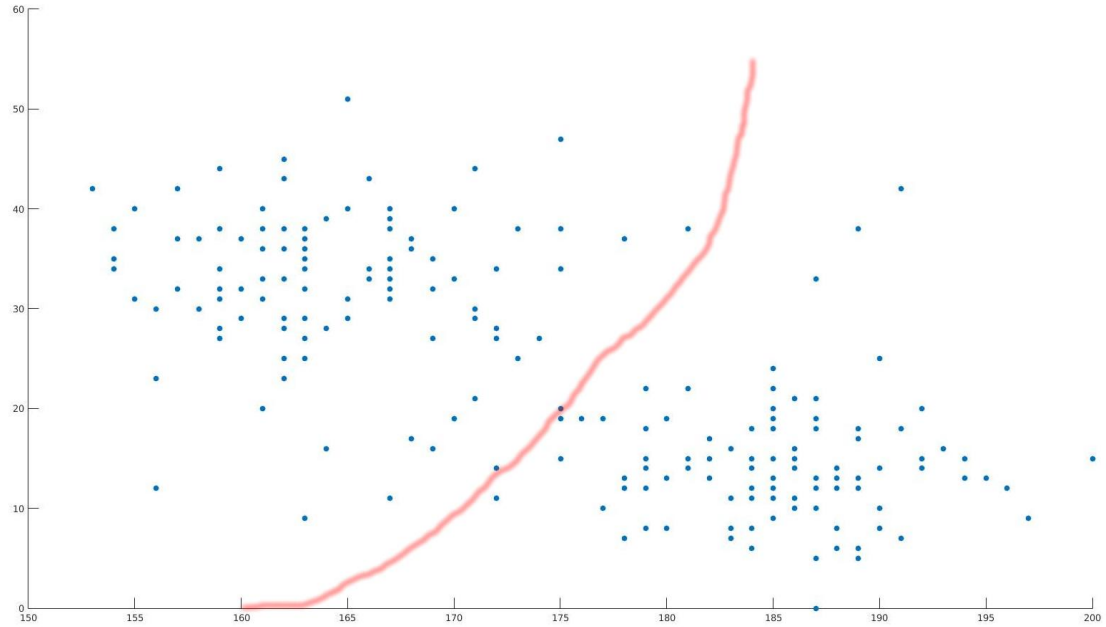
Assignment 2:

In the file lab1_2.mat a two dimensional array is given, consisting of measurements of the length (in cm) and the hair length (in cm) of 200 people.

- Plot the length versus the hair length.



- The measurements originate from 100 women and 100 men.Given the fact that in general men have shorter hair than women and men are taller, where would you draw the decision boundary(for example use a simple graphics editor to sketch it, or just plot a line between two points on top of your plot in matlab) and why ?



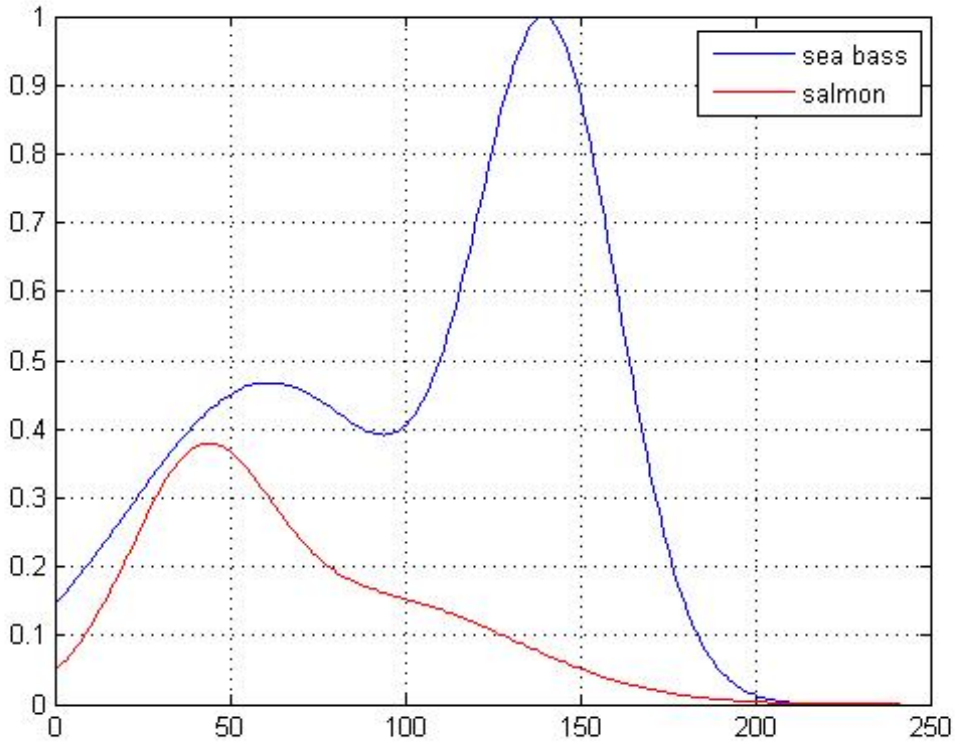
The decision boundary should be drawn in such a way to separate taller people with short hair, usually men, from the rest of the population.

Assignment 3:

Let's assume we obtained the class conditional probability of salmon and sea bass, describing the distribution of the length of the two classes. These are given by p_salmon and p_seabass at length 1 (in cm), all in the file lab1_3.mat.

- Given that sea bass is caught 3 times as often as salmon, calculate the posterior probabilities (don't forget to normalize to 1 for all lengths) and plot them.

  The likelihood of being a sea base is 75% or 0.75, while the likelihood of being a salmon is 25% or 0.25. Therefore, by multiplying the respective probabilities by the respective value, the plot is achieved. It is then normalized to 1 due to better representation.



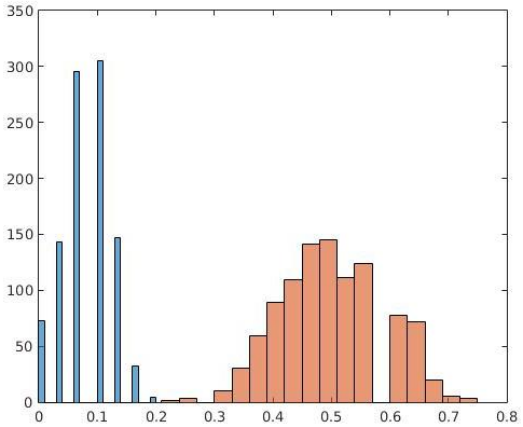
- Now a new fish is measured, which turns out to have a length of 8cm. According to your posterior probabilities, how would you classify this fish? And what if it's length is 20cm?

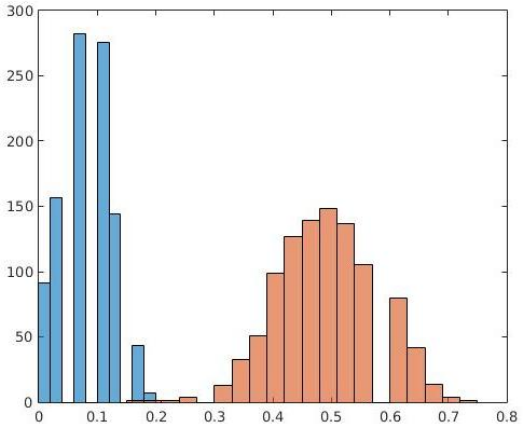
  Answer: In both cases it's more likely that the fish is of the sea bass kind.

Assignment 4:

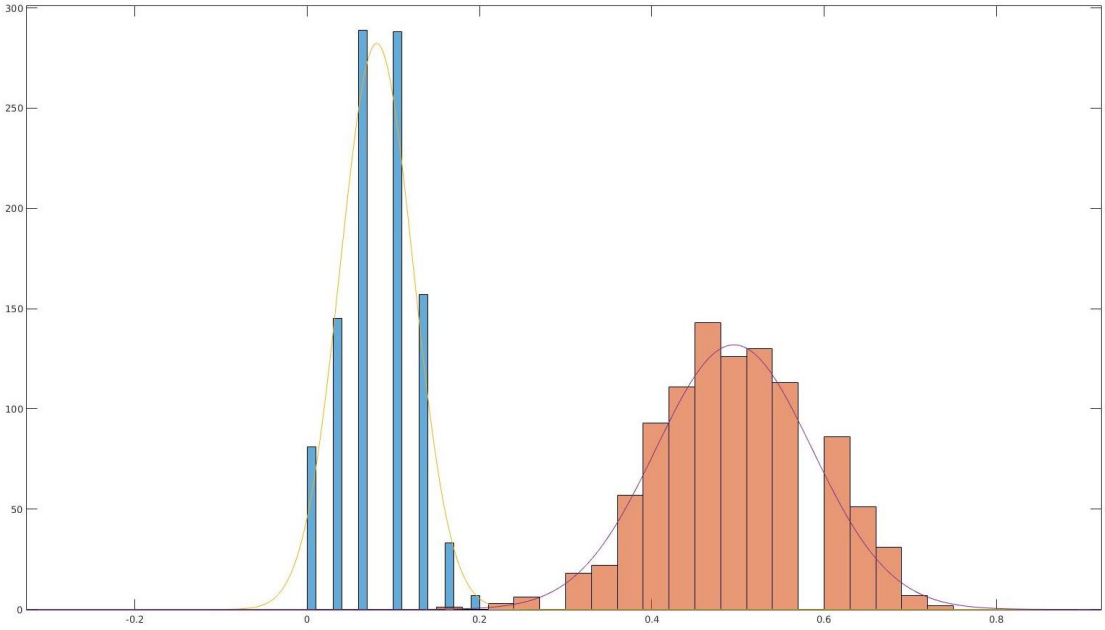
Consider the two dimensional binary arrays in files person01.mat to person20.mat. Each row of such an array person[i].m is a binary feature vector of 30 elements that is extracted from an iris image of a person that we call here person[i] (i = 1...20). Hence, each row is a 30-dimensional binary iris code of that person. There are 20 such iris codes of each person in the corresponding file person[i]; each row of the array is one such binary iris code.

- Plot the histograms of S and D in one figure with different colours. How much do the two histograms overlap?
  Since the comparison subjects are chosen randomly, the amount of overlap may vary. On the plot on the left for example, there is no overlap, but in most cases, we can identify some overlap around 0.2.





- Compute the means and the variances of the sets S and D. Add to the histograms of the previous question (4.3), plots of two normal distributions (Gaussian functions) with these means and variances. How well do the normal distributions fit the histograms?



The normal curves fit the histogram.

Division of work

The first and second assignments were done as a pair. Later Diego worked on assignment 3 and Eduardo on assignment 4.