Pattern Recognition practical 1

Maikel Withagen Steven Bosch

September 2015

1 Assignment 1

1.1

Consider it done

1.2

To compute the pair-wise correlation coefficients we used the following command:

Input

```
load('lab1_1.mat')
corrcoef(lab1_1)
```

This yields us the following table of correlation coefficients:

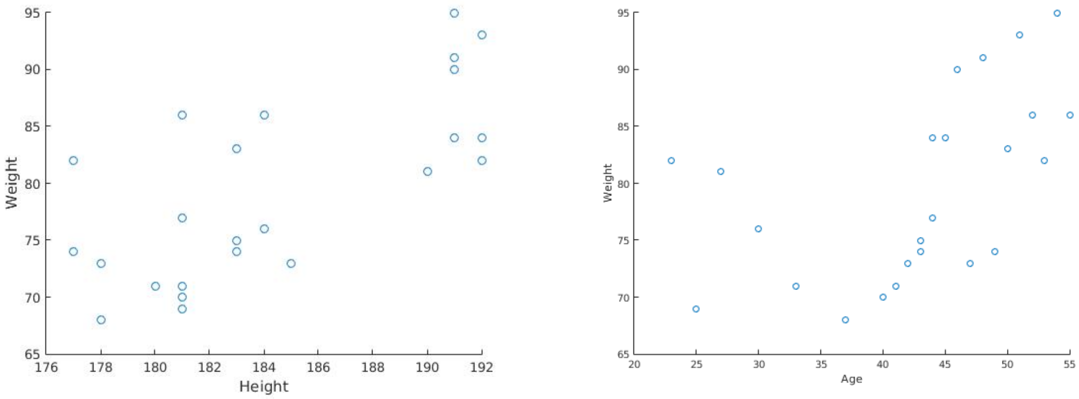
Table 1: *Pair-wise correlation coefficients*

	Length	Age	Weight
Length	1	-0.0615	0.7156
Age	-0.615	1	0.5142
Weight	0.7156	0.5142	1

1.3

The two features for which the correlation is the largest are the first and third column, respectively the height and the weight.

The two features for which the correlation is the second largest are the second and third column, respectively the height and the weight.



(a) Scatterplot of weight to length

(b) Scatterplot of weight to age

Figure 1

From a scatterplot alone it is hard to draw conclusions about any possible relationships between the different features. We do get indications though; figure 1a shows that there is likely to be a correlation between the weight and the height. An increase in weight seems to correspond to a (somewhat linear) increase in height. A similar kind of relationship can be seen in figure 1b, between the factors weight and age.

2 Assignment 2

2.3 Sets S and D

The following subsections show the code we used to acquire the 1000 Hamming distances for set S and D.

a)

Code for set S

```
hd_s = zeros(1,1000);
for i = 1:1000
    person = randi([1,20]);
    row1 = randi([1,20]);
    row2 = row1;
    while(row1 == row2)
        row2 = randi([1,20]);
    end
    load(sprintf('person%02d.mat',person));

    hd_s(i) = sum(abs(iriscode(row1,:) - iriscode(row2,:)));
end
hd_s = hd_s/30;
```

b)

Code for set D

```
hd_d = zeros(1,1000);
for i = 1:1000
    person1 = randi([1,20]);
    row1 = randi([1,20]);
    row2 = randi([1,20]);
    person2 = person1;
    while(person1 == person2)
        person2 = randi([1,20]);
    end
    load(sprintf('person%02d.mat',person1));
    x = iriscode(row1,:);
    load(sprintf('person%02d.mat',person2));
    y = iriscode(row2,:);
    hd_d(i) = sum(abs(x-y));
end
hd_d = hd_d/30;
```

2.4 Histogram

Figure 2 shows the histogram of sets S and D. The figure shows that the two distributions overlap very little, most of it around $hd = 5, 6$.

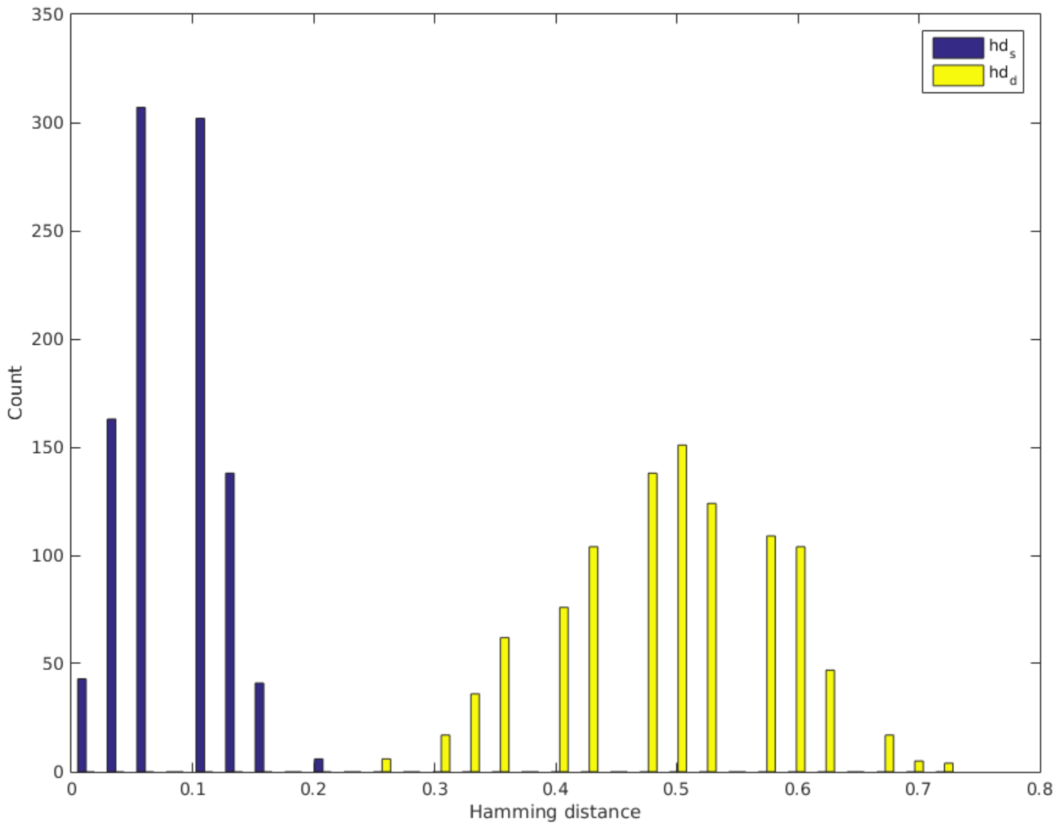


Figure 2: Histogram of sets S and D (30 bins

2.5

The means and variances of both of the sets are the following:

Set	Mean	Variance	Standard deviation
S	0.0825	0.0016	0.0398
D	0.4946	0.0079	0.0886

Table 2: Means and variances of both of the sets

The prior probability of two bits being the same between persons the following is: $1 - 0.4946 = 0.5054$. Using the formula $n = p(1-p)/\sigma^2$ we get $n = (0.4946 * (1 - 0.4946))/(0.0886^2) \approx 31.84$. From this we can conclude that 30 bits are not sufficient to properly distinguish between two irises.

2.6

2.7

2.8