**Assingment 2.**

1. After generating the sets S and D we saved their values in a .mat file for later use. We used function pdist2 for calculated the normalized hamming distances.

numPersons = 20;

codeLength = 30;

fileNameLength = 12;

size = 1000;

%generating filenames

Names = repmat(char(0),20,12);

for i = 1:20

Names(i, :) = sprintf('person%02d.mat',i);

end

%generating set S

S = zeros(1, size);

for i = 1:size

index = randi(numPersons);

randomFileName = Names(index, :);

person = load(randomFileName);

row1 = person.iriscode(randi(20), :);

row2 = person.iriscode(randi(20), :);

S(i) = pdist2(row1, row2, 'hamming');

end

%generating set D

D = zeros(1, size);

for i = 1:size

index = randi(numPersons);

randomFileName = Names(index, :);

person = load(randomFileName);

row1 = person.iriscode(randi(20), :);

index = randi(numPersons);

randomFileName = Names(index, :);

person = load(randomFileName);

row2 = person.iriscode(randi(20), :);

D(i) = pdist2(row1, row2, 'hamming');

end

save('hdsets.mat', 'S', 'D');

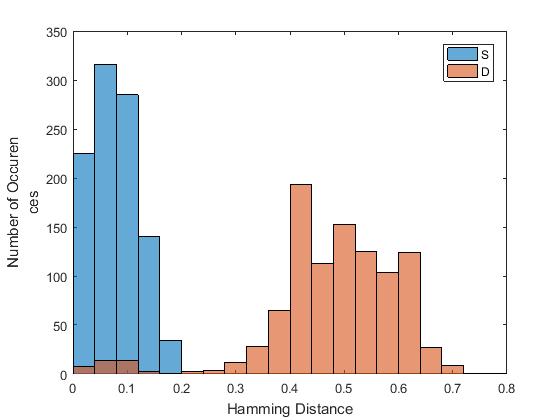
2. We constructed the histograms with binwidth of 0.4:

histogram(S, 'BinWidth', 0.04);

hold on

histogram(D, 'BinWidth', 0.04);

In the following picture, histograms constructed for sets S and D are shown. The two histograms have very small overlap. The values in set S are considerably smaller than values in set D.



3.

3.a. The following MATLAB code calculates the means of sets S and D:

meanS = 0; meanD = 0;

for i = 1:size

meanS = meanS + S(i);

meanD = meanD + D(i);

end

meanS = meanS / size;

meanD = meanD / size;

The following MATLAB code calculates the variances of sets S and D:

varS = 0; varD = 0;

for i = 1:size

varS = varS + (S(i) - meanS)^2;

varD = varD + (D(i) - meanD)^2;

end

varS = varS / size;

varD = varD / size;

The results are shown in the table below:

|  |  |  |
| --- | --- | --- |
|  | mean | variance |
| set S | 0.0786 | 0.0017 |
| set D | 0.4778 | 0.0143 |

3.b. The probability that two bits (in the same position) of the iris codes of two different persons are different can be estimated using the mean value of set D.

Set D contains normalized hamming distances between iris codes of different people, which is the number of positions in the iris codes where the values of bits are different, divided bu the number of bits. If we multiply the mean value by the number of bits, which is 30, we will get the average number of different bits (in the same position) of iris codes of two different people. Since there are 30 bits in total, the probability that bits on a randomly chosen position are different can be estimated using the average number of different bits divided by the number of bits, which is the mean value of set D: 0.4778.

3.c. Let’s construct 20 iriscodes of length 30 randomly. Obviously, the number of statistically dependent bits in these iriscodes will be 0.

randomCodes = zeros(20, 30);

for i = 1:20

for j = 1:30

randomCodes(i, j) = randi(2) - 1;

end

end

Each row of randomCodes contains a random iris code of length 30. Let’s construct set K of 1000 elements, each of which will be a normalized hamming distance between two randomly chosen rows of randomCodes.

K = zeros(1, size);

for i = 1:size

row1 = randomCodes(randi(20), :);

row2 = randomCodes(randi(20), :);

K(i) = pdist2(row1, row2, 'hamming');

end

The mean of set K is equal to 0.4796. For comparison, mean of set S is 0.0786 and the mean of set D is 0.4778.

4. If we plot graphs of normal distributions using the mean and standard deviation of sets S and D, we get the following results:

MATLAB code:

xs = [0:0.01:0.2];

normS = normpdf(xs, meanS, sqrt(varS));

xd = [0:0.01:0.8];

normD = normpdf(xd, meanD, sqrt(varD));

figure;

plot(xs,normS);

hold on

plot(xd, normD);

