Heiko Joshua Jungen jungen@student.chalmers.se 950418-T075
Elias Hult Pappas pappas@student.chalmers.se 940711-2417

TDA231 - Homework 2

Theoretical Problems

Problem 1.1

Class: “content”	1				1/2	0				1/2
Feature: “rich”	1	0	1	1	3/4	0	1	0	0	1/4
Feature: “married”	1	0	1	0	1/2	0	0	0	1	1/4
Feature: “healthy”	1	1	0	1	3/4	0	0	1	0	1/4

a) Probability to be content being “not rich”, “married” and “healthy”

$x^* = (0,\ 1,\ 1);\ c\ =\ 1$

$P(c = 1|x^*) = \frac{(1-0.75)*0.5*0.75*0.5}{(1-0.75)*0.5*0.75*0.5\ +\ (1-0.25)*0.25*0.25*0.5} = \frac{2}{3} \approx 67\%$

<u>Answer:</u> The probability of being content given the specified features is about **67%**.

b) Probability to be content being “not rich” and “married”

As we don’t know whether the person is healthy or unhealthy we have to consider the probabilities of both situations. That means we have to add up the probabilities of the person to be content when healthy and unhealthy.
The probability of being healthy is obtained by the ration of healthy people in the given observation. The probability of being unhealthy is the complementary probability. The mathematical equations are presented below.

$x^* = (x_1,\ x_2,\ x_3);\ x_1 = 0,\ x_2 = 1;$

$H:\ "Being\ healthy;\ P(H)\ =\ 0.5$

$P(c = 1|x_1 = 0,\ x_2 = 1) = P(c = 1|(0,\ 1,\ 0)) * P(\neg H)\ + P(c = 1|(0,\ 1,\ 1)) * P(H)$

$P(c = 1|(0,\ 1,\ 0)) = \frac{(1-0.75)*0.5*(1-0.75)*0.5}{(1-0.75)*0.5*(1-0.75)*0.5\ +\ (1-0.25)*0.25*(1-0.25)*0.5} = \frac{2}{11}$

$P(c = 1|x_1 = 0,\ x_2 = 1) = 0.5 * (\frac{2}{3} + \frac{2}{11}) = \frac{14}{33} \approx 42\%$

<u>Answer:</u> The probability of being content given the specified features is about **42%**.

Problem 1.2

In the given scenario, the attributes 1, 2 and 3 are dependent on another, because if one attribute takes the value 1 the other two need to be 0 in every case. The problem using naive bayes is, that the assumption of independence is therefore violated.

A solution to this problem are **multi-state variables**, opposed to binary variables. Attributes 1, 2 and 3 are merged into one variable with more than two states.

$x_1 =$ 0 if the customer is younger than 20, 1 if the customer if between 20 and 30, 2 if the customer is older than 30.
$x_2 =$ 1 if customer walks to work and 0 otherwise.

Thereby the dependency between the attributes is resolved, so we can continue with the naive Bayes method.

Heiko Joshua Jungen jungen@student.chalmers.se 950418-T075
Elias Hult Pappas pappas@student.chalmers.se 940711-2417

Matlab problems

Problem 2.1

a) Bayes expression and class conditional density

Below the Bayes expression for both classes are shown.

$$P(y_{new}=-1|x_{new},X,y)=\frac{P(x_{new}|y_{new}=-1,X,y)*P(y_{new}=-1)}{\sum_j P(x_{new}|y_{new}=j,X,y)*P(y_{new}=j)}=\frac{P(x_{new}|y_{new}=-1,X,y)}{\sum_j P(x_{new}|y_{new}=j,X,y)}$$

$$P(y_{new}=+1|x_{new},X,y)=\frac{P(x_{new}|y_{new}=+1,X,y)*P(y_{new}=+1)}{\sum_j P(x_{new}|y_{new}=j,X,y)*P(y_{new}=j)}=\frac{P(x_{new}|y_{new}=+1,X,y)}{\sum_j P(x_{new}|y_{new}=j,X,y)}$$

The class conditional density can be described by a gaussian distribution with the MLE as parameters for the normal distribution. **I** is the 3x3 identity matrix.

$$P(x_{new}|y_{new},X,y) = P(x_{new}|\hat{\mu}_j,\hat{\sigma}_j^2) = \sim N(\hat{\mu}_j,\hat{\sigma}_j^2*I)$$

$$\Rightarrow \frac{1}{\sqrt{2\pi^3*\hat{\sigma}_j^6}}*exp(-\frac{(x-\hat{\mu}_j)^T(x-\hat{\mu}_j)}{2\hat{\sigma}_j^2})$$

d) Five fold cross-validation

The five fold cross-validation shows that **both classifiers have an error rate of 0%**.

Problem 2.2

<u>Information:</u> File *problem_22.m* solves task a, b and c. *File problem_22_d.m* contains adjustments for task d.

c) Five fold cross-validation

Due to randomised folds these rates vary slightly, e.g a run resultet in error rate of precisely 5.64% for (a) and 23.18% for (b). Overall, the classification between class 5 and 8 has an error rate of about **6% for (a)** and about **23% for (b)**.

d) Improvements

The error rate for (b) can be improved by combining the feature vector of submatrices from the 16x16 original image matrix. Possible submatrix sizes are 2x2, 4x4 and 8x8. It turned out that a 4x4 size for submatrices is optimal for the error rate. Taking submatrices of that size will result in 16 matrices for which a feature vector of size 8x1 is calculated. These feature vectors are combined into a (16 Matrices * 8=) 128 feature vector. The resulting **error rate is about 6%**.

The former approach of (b) was to consider the variances of each row and column. By using submatrices the classifier considers again the variances of each row and column, but now in a specific area of the matrix. In other words, the classifier now tries to identify small sub-images instead of the whole image.