Technische Hochschule Ostwestfalen-Lippe
Fachbereich Elektrotechnik und Technische Informatik
Prof. Dr. Helene Dörksen

# Exercises and Labs 3 for Lecture “Authentication „ (M.Sc.)

**Exercise 3.1** Three coins are tossed simultaneously. What is the probability of them showing two heads?

**Exercise 3.2** (a) What is the probability of getting at least one “6” if three dice are rolled? (b) What is the probability of getting at least one “6” in four rolls of a single die? (c) What is the probability of getting at least one double-6 in 24 throws of a pair of dice?

**Exercise 3.3** Consider a family with two children. Given that one of the children is a boy, what is the probability that both children are boys?

**Exercise 3.4** In a game show (Monty Hall’s Let’s Make a Deal) there are three closed doors. Behind one of these doors is a car, and behind the other two are goats. The contestant picks a door and Monty opens one of the remaining doors to reveal a goat. The contestant is then given the option to switch doors. Is it to their advantage to do so? (What is the prior probability of winning the car? What is the (posterior) probability of winning the car if they stay with their first choice? What if they decide to switch?).

**Exercise 3.5** Suppose that a rare disease affects 1 out of every 1000 people in a population, i.e., the prior probability is 1/1000. And suppose that there is a good, but not perfect, test for the disease. For a person who has the disease, the test comes back positive 99% of the time (sensitivity = 0.99) and if for a person who does not have the disease the test is negative 98% of the time (specificity = 0.98). You have just tested positive; what are your chances of having the disease? (Try by calculation, and then check using CondprobEx3.xls, found in the same download area).

*Remark*: The Excel file, CondprobEx3.xls finds the posterior (or predictive) probability of a positive test, $P(A|B)$, and the posterior (or predictive) probability of a negative test (NPV), $P(\bar{A}|\bar{B})$, given the sensitivity and specificity of the test and the prior probability of the disease (Formulation I) or the contingency table (Formulation II).

**Exercise 3.6** Consider following example: About 1% of women who participate in routine breast screening have breast cancer. About 80% of those with breast cancer get a positive test result and 9.6% without breast cancer also get a positive result. A woman gets a positive test result. What is the probability that she has breast cancer? We calculate: The women who have a positive mammography results and actually have breast cancer amount to 80% of 1%, or 0.8%, of all the women who were tested. The women who do not have breast cancer, but still show a (false) positive mammography results, amount to 9.6% of 99%, or 9.504%, of all the women who were tested. Therefore, total positive mammography results amount to $0.8\% + 9.504\%$, or 10.304%, of all the women who were tested. Of this percentage, the women who actually have cancer (0.8%) are in the distinct minority. The probability of a woman who has a positive test result actually having cancer, $P(A|B)$, is $0.8/10.304 = 0.0776$ or 7.76%.

We can see this also by taking, e.g. a population of $100,000$. About 1% of these have breast cancer, i.e., $A = 1,000$. Further, 80% of these test positive ($TP = 800$, and the remaining 200 are $FN$); 9.6% of those without breast cancer ($FP + TN = 99,000$) get a positive result ($FP$): therefore $FP$ is 9504. The probability of a woman who has a positive test result actually having cancer is $TP/(TP + FP) = 7.76\%$.

Now, the woman in example tested positive and her posterior probability of having breast cancer was calculated to be 7.76%. If she decides to go for another test and she tests positive a second time, what is the probability of her having breast cancer? (And what if a third, fourth, fifth test was positive, what would the corresponding probabilities be? Of course this is an unlikely scenario since each test exposes her to x-rays and a consequent risk of actually causing cancer). Try by calculation, and then check using CondprobEx3.xls.

**Exercise 3.7** Consider the following feature vector:

$$X = \begin{bmatrix} 7 & 4 & 3 \\ 4 & 1 & 8 \\ 6 & 3 & 5 \\ 8 & 6 & 1 \\ 8 & 5 & 7 \\ 7 & 2 & 9 \\ 8 & 2 & 2 \\ 7 & 4 & 5 \\ 9 & 5 & 8 \\ 5 & 3 & 3 \end{bmatrix}$$

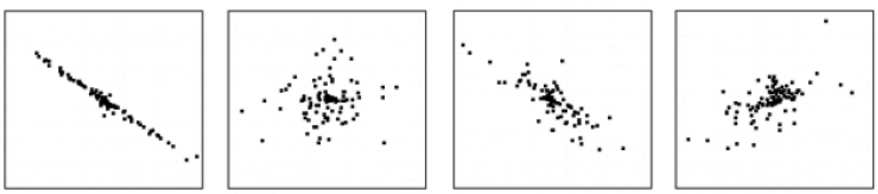
representing a set of ten observations of three features of an object. Calculate the covariance matrix.

**Exercise 3.8** We want to classify a given person as male or female based on height, weight and foot size. The data from an example training set (assumed Gaussian) is:

| Sex | Height (feet) | Weight (lbs) | Foot size (inches) |
|---|---|---|---|
| M | 6 | 180 | 12 |
| M | 5.92 (5'11") | 190 | 11 |
| M | 5.58 (5' 7") | 170 | 12 |
| M | 5.92 (5'11") | 165 | 10 |
| F | 5 | 100 | 6 |
| F | 5.5 (5' 6") | 150 | 8 |
| F | 5.42 (5' 6') | 130 | 7 |
| F | 5.75 (5' 9') | 150 | 9 |

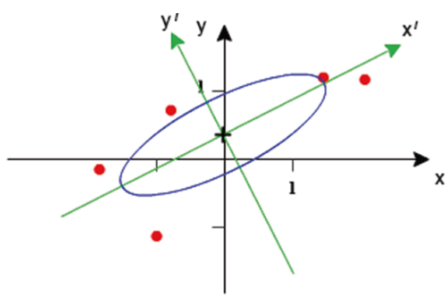
How would you classify a sample with height = 6', weight = 130 lbs., and foot size= 8"? Use (i) the naïve Bayes' classifier (i.e. assuming no covariance between the features) (ii) use the covariance matrix.

**Exercise 3.9** Match the scatter plot (A – D) in figures to the correct correlation coefficient (i) 0.14; (ii) -0.99; (iii) 0.43; (iv) -0.77.



(A) (B) (C) (D)

**Exercise 3.10** Given following data points in feature space (assumed bivariate Gaussian) and a corresponding isocontour:




The coordinates are:

$$(x, y) = \begin{pmatrix} 2.06 & 1.18 \\ 1.46 & 1.20 \\ -0.78 & 0.72 \\ -1.84 & -0.16 \\ -1.00 & -1.12 \end{pmatrix}$$

Construct the covariance matrix in the $(x, y)$ coordinate system from the data shown. Diagonalize the matrix and get the coordinates in $(x', y')$. Use $[V, D] = eig(A)$ in MatLab: the eigenvalues of $A$ are obtained in $D$, and the columns of $V$ are the eigenvectors of $A$.

**Exercise 3.11** Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}$.