AARHUS UNIVERSITET

Computer Engineering

Term:	**Exam Fall 2018**
Examination:	**Optimization and Data Analytics**
Date:	**14. Jan. 2019**
Duration:	**3 hours, 9-12**
Room:	**Trøjborgvej 82-84, Bygning 1912-1917, 8000 Aarhus C**

Aarhus University hands out:
4 pieces of checkered paper

Digital Exam
The exam questions will be available in "Digital Exam", and your exam answers must be handed in via "Digital Exam". Handwritten parts of the exam answers must be digitized and attached to your exam paper. In "Digital Exam", the exam answers must be uploaded in PDF format.

REMEMBER name and study number on all pages and in the file name when you upload (pdf).

Remember to upload via Digital Exam. You will get a receipt, immediately after you have uploaded correctly.

Remember to upload within the time limit, if you exceed the time limit, you must send in an application for an exemption.

Aids
All aids, like Computer, mathematical software, books, notes, pen and paper and internet is allowed. No form of communication or file sharing is allowed during the exam.

Exercise 1

Consider the following optimization problem:

Maximize:	$f(x_1, x_2) = 3x_1 + 2x_2$	
Subject to:	$2x_1 + x_2 \leq 100;$	(i)
	$x_1 + x_2 \leq 80;$	(ii)
	$x_1 \leq 40;$	(iii)
and	$x_1 \geq 0; x_2 \geq 0;$	

a) Sketch the feasible set in 2 dimensions, including a sketch of the level set: $L = \{(x_1, x_2) \in R^2 | f(x_1, x_2) = 0\}$.

b) Write out the simplex tableaux for the problem and show the first step needed to bring in a new variable into the solution (e.g. argue what column and row to choose, and what elementary operations are needed for the first reductions). Find the maximum for f on the feasible set (you may use MatLab).

c) Assume you are allowed to increase the right hand side of *only one* of the 3 inequalities above with a small amount. Which one will you choose, if the maximum should be increased the most? (argue for your choice).

Exercise 2

Consider the function: $g: R^2 \to R$, where $g\ (x_1, x_2) = e^{(x_1+1)x_2}$.

a) Find the partial derivatives of g and the directional derivative of g in the point $(2, 3)$ in the direction $d = (1, 1)$ (argue for your calculations).

b) Write up a linear expression in the variables x_1 and x_2 that approximates g in a neighborhood of the point $(2, 3)$ and use this expression to estimate the value of $g(2.1, 3.1)$ (argue for your calculations).

Now let $f(x_1, x_2) = (x_1 - 1)^2 + x_2^2$ and $h(x_1, x_2) = \frac{x_1^2}{9} + \frac{x_2^2}{4} - 1$.

c) Find the minimum and maximum of f subject to the constraint $h(x_1, x_2) = 0$ (argue for your calculations).

Exercise 3

Answer the following with **ONE sentence** per question.

a) List two useful terminating conditions for iterative optimization algorithms.

b) Sometimes simulated annealing accepts a **worse** candidate solution during its search. Why?

c) What is the difference between discrete and continuous optimization?

d) Is particle swarm optimization guaranteed to find a global optimum (Yes / No)?

e) A security guard has the job of patrolling a large building at night. They must repeatedly visit six important locations until their nightshift is over, referred to as locations A,B,C,D,E,F. The poor guard is a bit bored, so we will use **simulated annealing** to find the **longest path** to visit **all** locations A,B,C,D,E,F such that each location is visited **exactly once**. Assume you have a function *Dist* that takes two location labels as input and returns the distance between those locations, e.g. *Dist*(A,D) = 52.

i) Develop a representation of a candidate solution. Give an example of a candidate solution using your representation.

ii) Develop a neighbor algorithm for your candidate solution, such that, given candidate solution X your neighbor algorithm returns another "neighbor" candidate solution Y. Demonstrate with an example using candidate solution X and neighbor Y.

iii) Give a cost function, *f*, for a candidate solution. Show how to calculate the cost of your previous example of a candidate solution.

iv) Develop a linear cooling schedule, and give an example of temperature for the first two iterations.

v) In simulated annealing, when is the search algorithm more likely to accept a worse candidate solution?

vi) In lectures we discussed the following probability function:

$$\exp\left(\frac{-\left(f(Z^{(i)}) - f(X^{(i)})\right)}{T^{(i)}} \right)$$

…where $X^{(i)}$ is the candidate solution at iteration "i", $Z^{(i)}$ is the newly selected neighbor, $T^{(i)}$ is the temperature at iteration "i", and *f* is the cost function.

Can you use this probability function to solve the guard's path problem with your cost function *f*, or do you need to change this probability function? Explain your answer, and suggest a change if needed.

Exercise 4

A classification problem is formed by two classes. We are given a set of 2-dimensional data **X** = [$\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_4$]:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 & -2 \\ 2 & 0 & -1 & -1 \end{bmatrix}$$

each belonging to one of the two classes, as indicated in the class label vector **l** = [l_1 l_2 l_3 l_4]:

$$\boldsymbol{l} = \begin{bmatrix} 1 & 1 & 2 & 2 \end{bmatrix}$$

Using the above (training) vectors and the corresponding class labels, classify the following vectors:

$\mathbf{x}_5 = [2 \quad 0]^T$, $\mathbf{x}_6 = [0 \quad 0]^T$, $\mathbf{x}_7 = [-1 \quad 1]^T$, $\mathbf{x}_8 = [-2 \quad 0]^T$

using the following classifiers:

a) The Nearest Class Centroid (NCC) classifier
b) The Nearest Neighbor Classifier (using only one neighbor)

c) Calculate the projection vector **w** that you will obtain by applying the Linear Discriminant Analysis (LDA) method using the training vectors **X** = [$\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_4$].
d) Draw in Figure 1 the projection vector **w** of question c.

.

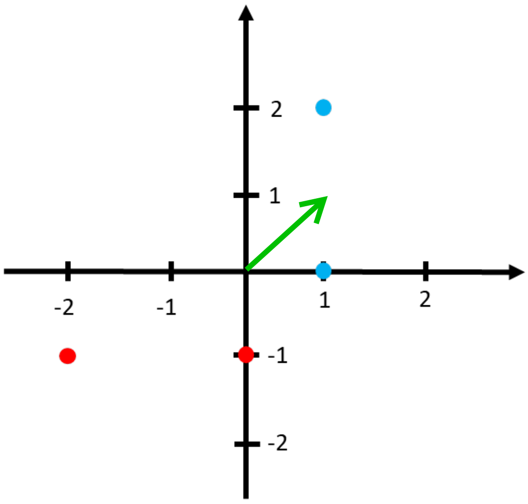

Fig. 1: Training samples: class one samples in blue and class two samples in red

Exercise 5

In a two-class classification problem, the distribution of the class-conditional probabilities $p(x|c_k)$, k=1,2 is given in Figure 2.

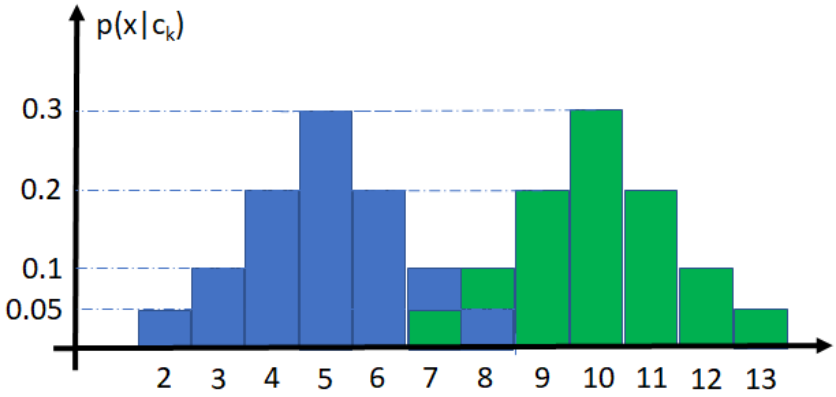

Fig. 2: Class-conditional probabilities of class 1 (blue) and class 2 (green)

The number of samples in class 1 and class 2 is equal to 500 and 1000, respectively.

a) Classify the following vectors (test) samples using the Bayes classification rule:

$$x_1 = 4, \quad x_2 = 7, \quad x_3 = 8 \text{ and } x_4 = 9$$

b) Consider that the classification risk for the two classes is given by the matrix:

$$\Lambda = \begin{bmatrix} 0 & 0.4 \\ 0.3 & 0 \end{bmatrix}$$

where $\Lambda_{ik} = \lambda((\alpha_i|c_k)$ is the risk of taking action α_i while the correct class is c_k. What is the classification result for the (test) samples x_i, i=1,…,4?

Exercise 6

The two classes of a binary classification problem are formed by the blue and red samples plotted in Figure 3.

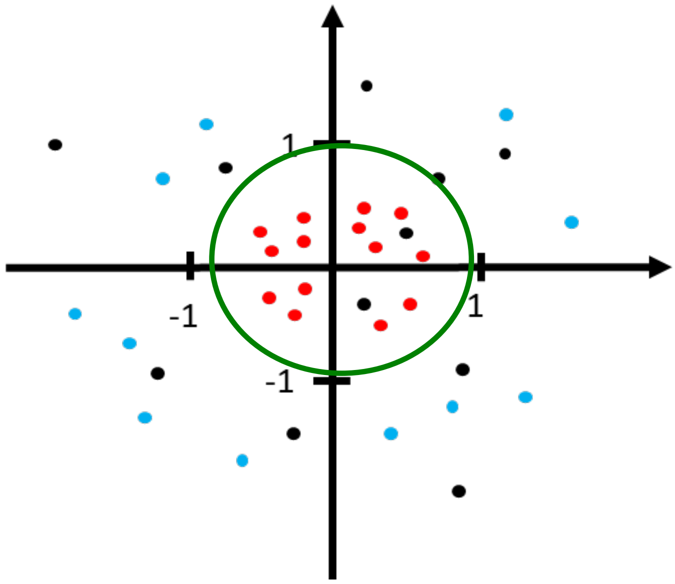

Fig. 3: Training samples of a two-class classification problem

a) Would you use Linear Discriminant Analysis in order to discriminate the two classes in a one-dimensional space? If yes, why. If no, why not?

b) Describe how we can use a Generalized Linear Discriminant Function in order to classify the test samples (plotted as black dots) using a linear classifier. Draw the linear classification function in the original feature space (the one shown in Figure 3) and the space where linear classification is performed.

c) Draw a neural network with one hidden layer that can be used for the above classification problem. In your drawing include all the parameters and activation functions. Explain (with equations) why we cannot use linear activation functions for its neurons in that case.