AARHUS UNIVERSITET

Computer Engineering

Term:	**Exam Fall 2017**
Examination:	**Optimization and Data Analytics**
Date:	**15. Jan. 2018**
Duration:	**3 hours, 9-12**
Room:	**Xxx**
1 cover plus paper for draft and fair copy will be handed out.	

Please notice:
You can use Blackboard to hand in your exam answers electronically in PDF-format.
Please indicate on the exam cover, whether you have handed in your answers hand written, electronic or both.

REMEMBER name and study number on all pages and in the document/file name when you upload (pdf).
It is your own responsibility to have knowledge of the rules for electronic hand in and if necessary, to be able to download to your own USB memory key in the unlikely event that Blackboard is down.

Exercise 1

Consider the following optimization problem:

Maximize:	$f(x) = x_1 + x_2$	
Subject to:	$1x_1 + 2x_2 \leq 12;$	(i)
	$x_1 + 5x_2 \leq 25;$	(ii)
	$x_1 \leq 6;$	(iii)
and	$x_1 \geq 0; x_2 \geq 0;$	

a) Sketch the feasible set in 2 dimensions, including the level set: $L = \{x \in R^n | f(x) = 1\}$.

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).

Now we change inequality ii) to the following inequality: $x_1 + 5x_2 \leq 30$. We keep inequality i) and iii).

c) Does this change the maximum for f on new the feasible set? Argue for your answer (try to argue without doing a new simplex tableau calculation).

Exercise 2

Consider the function: $f: R^3 \rightarrow R$, where $f\ (x) = x_1 - x_2 - x_3$.

a) Find the gradient of f and the directional derivate in the direction $d = (2, 3, 4)$ (argue for your calculations).

Now let f be subject to the constraint, $h(x_1, x_2, x_3) = 0$, where $h\left(x_1, x_{2,} x_3\right) = \frac{x_1^2}{4} + \frac{x_2^2}{9} + \frac{x_3^2}{3} - 1$.

b) Find the maximum and minimum for f over the feasible set $\mathcal{F} = \left\{\left(x_1, x_{2,} x_3\right) | h\left(x_1, x_{2,} x_3\right) = 0\right\}$ (argue for your calculations).

Exercise 3

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

a) What is the main difference between the Newton and Quasi-Newton method?

b) What is discrete optimization? Name one example of a well-known discrete optimization problem.

c) What is the main feature of global search methods?

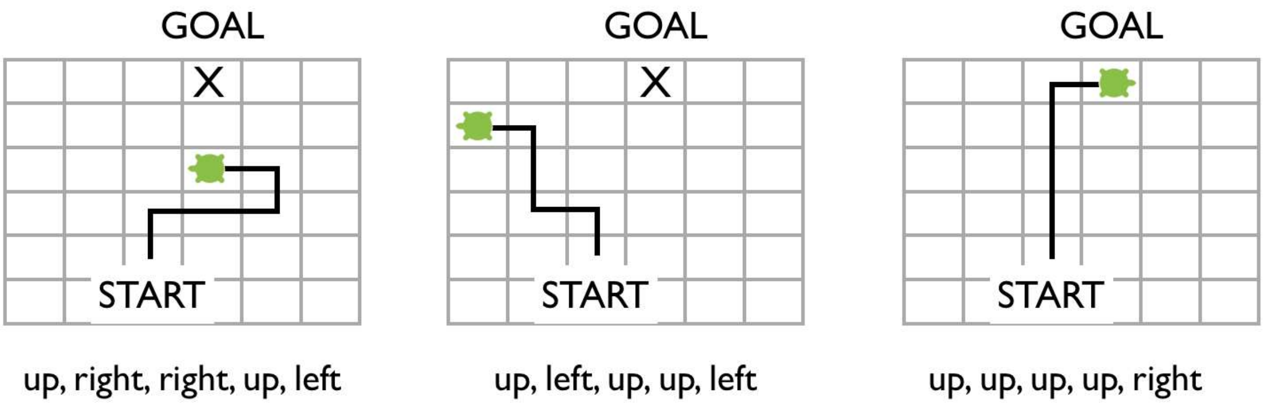
d) In lectures we discussed the following optimization methods:

Particle Swarm	Newton
Simulated Annealing	Genetic Algorithms
Quasi-Newton	Conjugate Gradient

Add these methods to the tables below according to their properties (some methods will be added to more than one table)

Deterministic	Stochastic	Global search	Population of candidate solutions per iteration

e) We are helping a robot turtle find some treasure. A "turtle path" always begins at the START square and consists of five steps. At each step the turtle moves one square, either: up, down, left, or right. Below are some examples of turtle paths.



We want to use a genetic algorithm to find instructions so the turtle ends up at the GOAL (where the treasure is) once it stops moving. Each "turtle path" is a candidate solution.

i) Develop a representation of a five step "turtle path" as a chromosome. Explain your representation, and give one example of a chromosome with the corresponding turtle path.

ii) Suggest an objective function that gives an appropriate cost to each candidate solution such that any turtle path that ends at the goal minimises the cost. Give the costs for the three example turtle paths in the pictures above.

iii) Give an example of single-point cross-over using your chromosomes (choose an example that clearly demonstrates this).

iv) Give an example of mutation using your chromosomes.

v) Given an initial population P of size N, explain the steps of your genetic algorithm search for ONE generation (about one or two sentences per algorithm step).

Exercise 4

A classification problem is formed by two classes. We are given a set of 2-dimensional data $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_7]$:

$$\boldsymbol{X} = \begin{bmatrix} 0 & 1 & 1 & 3 & 3 & 4 & 4 \\ 1 & 0 & 1 & 1 & 2 & 1 & 2 \end{bmatrix}$$

each belonging to one of the two classes, as indicated in the class label vector $\mathbf{l} = [l_1\ l_2\ \ldots\ l_7]$:

$$\boldsymbol{l} = [1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 2]$$

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

$$\mathbf{x}_8 = [1 \quad 2]^T, \quad \mathbf{x}_9 = [3 \quad 0]^T, \quad \mathbf{x}_{10} = [2 \quad 1]^T$$

using the following classifiers:

a) The Nearest Class Centroid (NCC) classifier
b) The Nearest Neighbor Classifier (using only one neighbor)
c) The Bayes-based classification scheme, where:

$$p(\boldsymbol{x}_i|c_k) = \frac{\|\boldsymbol{x}_i - \boldsymbol{m}_k\|_2^{-2}}{\sum_{m=1}^{K}\|\boldsymbol{x}_i - \boldsymbol{m}_l\|_2^{-2}},$$

where $\boldsymbol{m}_k$ is the class mean vector of class c_k, k=1,2 and $\|v\|_2^{-2} = 1 / (v^T v)$.

d) Compare (qualitatively) the decision functions obtained by using the NCC classifier and the above Bayes-based classification scheme.

Exercise 5

a) Draw a neural network solving a 3-class classification problem using training data $x_i \in \mathbb{R}^5, i = 1, \dots, N$. The neural network is formed by 2 hidden layers (having L_1 and L_2 neurons, respectively). For each neuron, you may or not use the bias input.
b) Show that the use of the linear activation function (for all neurons) makes the above network equivalent to a two-layer (no hidden layers) network.
c) Based on the above, describe why it is important to use non-linear activation functions in neural networks. You can base your answer using the structure of the network in question a.

Exercise 6

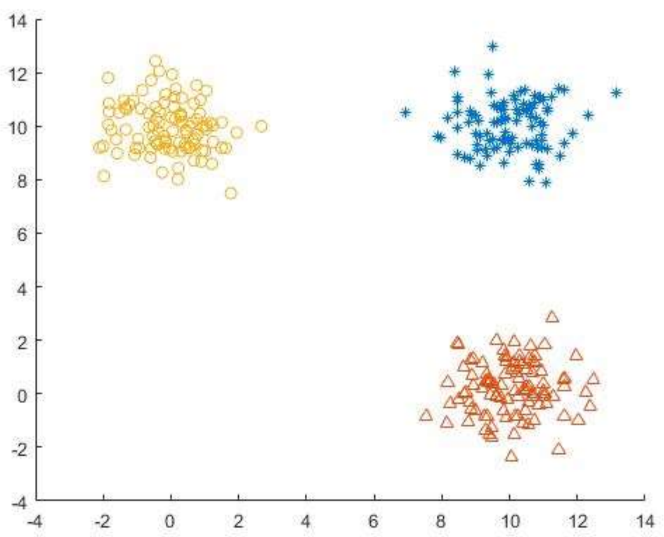
The following Figures illustrate data coming from four classes of a 3-class classification problem. Samples belonging to the same class each class are illustrated with the same color. Draw decision functions based on:

a) the One-versus-Rest (or One-versus-All) multi-class classification scheme
b) the One-versus-One multi-class classification scheme

For each multi-class classification scheme (a) and (b), draw the corresponding decision function parameter vectors (**w**).

c) For both the above multi-class classification schemes, based on the decision functions you drew for question a and b, indicate the regions of ambiguity.

One-versus-Rest



One-versus-One