Term:	**Reexamination Winter 2017**
Test in:	**Optimization and Data Analytics**
Date:	**1. June 2018**
Form:	**3 hours (9.00 - 12.00)**
Room:	**Åbogade 15, bygning 5511, lokale 022, 8200 Århus N**

Aarhus University hands out:
4 pieces of white 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.
No form of communication or file sharing is allowed during the exam.

Exercise 1

Consider the following optimization problem:

Maximize:	$f(x) = 6x_1 + 5x_2$	
Subject to:	$x_1 + 4x_2 \leq 16;$	(i)
	$6x_1 + 4x_2 \leq 30;$	(ii)
	$2x_1 - 5x_2 \leq 6;$	(iii)
and	$x_1 \geq 0; x_2 \geq 0;$	

a) Sketch the feasible set in 2 dimensions.

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) Find the maximum point $(x_1, x_{2,})$ of f, when x_1, x_2 are both ***integers***. Show your solution in the sketch from question a). Argue for your answer.

Exercise 2

Consider the function: $f: R^2 \rightarrow R$, where $f\ (x) = x_1 \cdot x_2 - x_1$.

a) Find the gradient of f and the directional derivate in the direction $d = (1, 1)$ in the point $(x_1, x_2) = (1, 2)$. Argue for your calculations.

Now let f be subject to the constraint, $h(x_1, x_2) = 0$, where $h(x_1, x_{2,}) = x_1^2 + x_2^2 - 4$.

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

c) Argue that the maximum of f on the set $D = \{(x_1, x_{2,}) | h(x_1, x_{2,}) \leq 0\}$ is the maximum found in b) (Note: D includes both the circle AND the interior of the circle).

Exercise 3

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

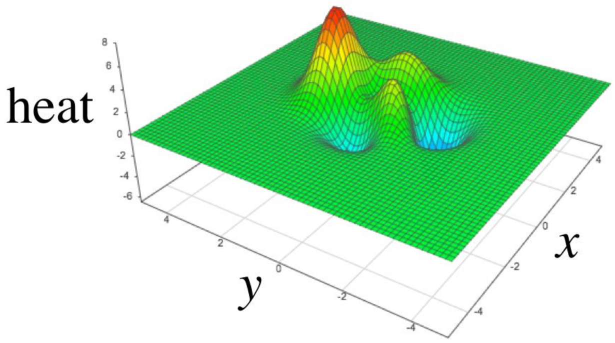
a) In the context of optimization, is simulated annealing guaranteed to find the global optimum (Yes/No)?

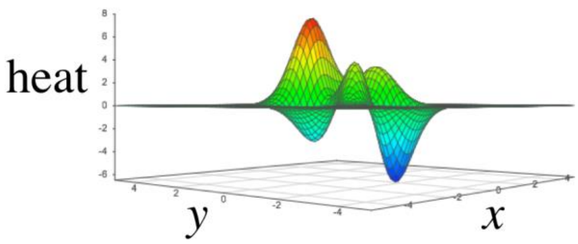
b) What is continuous optimization? (optional: you may give an example of a continuous optimization problem to help explain your answer).

c) What does it mean if an optimization method is "stochastic"? Give one example of a stochastic optimization method we discussed in lectures.

d) Some optimization methods use a "population of candidates" during the search. What does this mean? Give one example of such an optimization method.

e) We are helping a cat find the warmest location in the room to have a sleep. Each location in the room is represented by two real coordinates (*x*,*y*), and each location gives a heat score described by a function *f*. The distribution of heat across the room according to function *f* is illustrated in the graphs below.





We want to use a particle swarm to find the warmest location in the room, i.e. the (x,y) coordinates that have the highest "heat" score.

i) Develop a representation of a candidate solution as a "particle", including a fitness function.

The particle swarm update formula is:

$$v_i(t+1) = wv_i(t) + c_1r_1[\hat{x}_i(t) - x_i(t)] + c_2r_2[g(t) - x_i(t)]$$

ii) What does "v_i" refer to in this formula?

iii) What does "t" refer to in this formula?

iv) Given an initial particle population P of size N, describe the THREE main steps of your particle swarm search for ONE iteration (about one sentence per step).

Exercise 4

Question 1

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

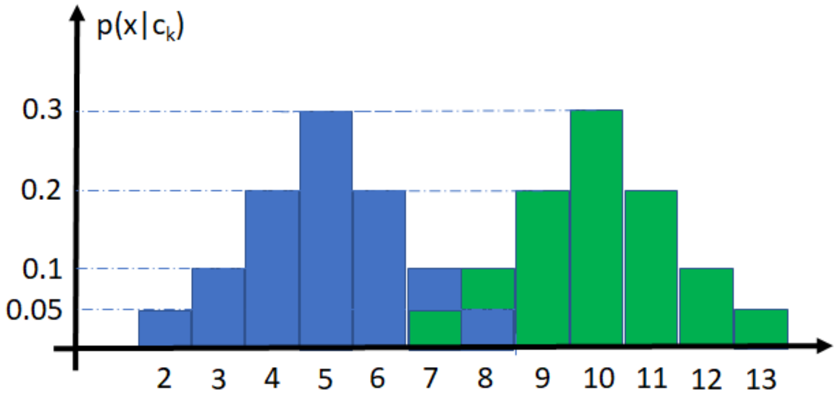


Fig. 1: Class-conditional probabilities of class 1 (blue) and class 2 (green)

The number of samples in class 1 and class 2 is equal to 100 and 200, respectively.

a) Classify (using the trained classifier) the following vectors (test) samples:

$$x_1 = 3, \quad x_2 = 7, \quad x_3 = 8 \text{ and } x_4 = 9$$

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

$$\Lambda = \begin{bmatrix} 0 & 0.3 \\ 0.2 & 0 \end{bmatrix}$$

where $\Lambda_{ij} = \lambda((\alpha_i|c_k)$ 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?

Question 2

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

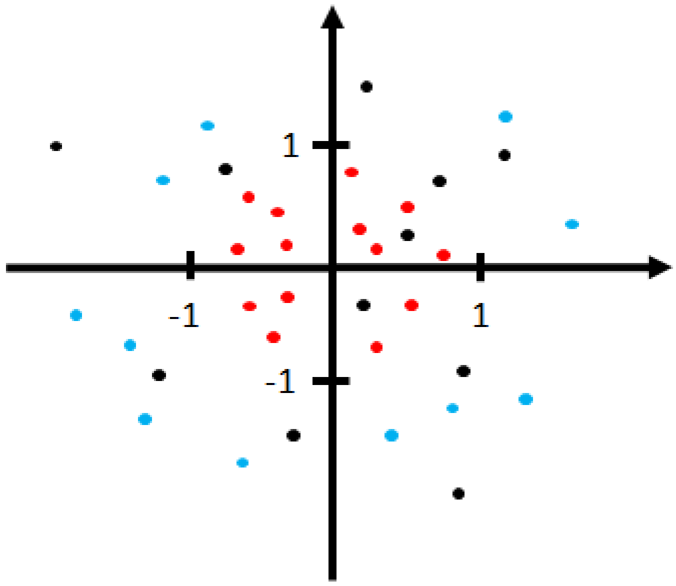


Fig. 2: Training samples of a two-class classification problem

a) 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.
b) Write two data transformations that can be used for the application of the process in a).
c) For each of the data transformations described in question (b), draw the transformed training data and the decision function.

Question 3

a) Draw a neural network solving a 5-class classification problem using training data $\boldsymbol{x}_i \in \mathbb{R}^5, i = 1, \ldots, N$. The neural network is formed by 2 hidden layers.

b) Express the output of the network $\boldsymbol{o}_i$ with respect to the input vector $\boldsymbol{x}_i$ and describe the classification rule based on which $\boldsymbol{x}_i$ will be assigned to one of the 5 classes.

c) Show that the use of the linear activation function (for all layers) makes the above network equivalent to a two-layer (no hidden layers) network.

d) Based on the above, describe why it is important to use non-linear activation functions in neural networks.