Exam: Introduction to Intelligent Systems 2013-10-28

NO OPEN BOOK! GEEN OPEN BOEK! - It is not allowed to use the course book(s) or slides or any other (printed, written or electronic) material during the exam. You may only use a simple electronic calculator. Give sufficient explanations to demonstrate how you come to a given solution or answer! The 'weight' of each problem is specified by a number of points, e.g. (1 p). You may give answers in English, Dutch or German language.
Be precise and write down equations where appropriate. Do not answer questions with just "Yes" or "No", always provide reasons/arguments for your answers!

---

1) **Hopfield Model (1 point)**
   Consider a Hopfield neural network with N fully connected neurons of the McCulloch-Pitts type: $S_i(t) \in \{-1,+1\}$, $(i = 1, 2, \ldots, N)$. These neurons display either maximal activity (+1) or minimal activity (-1).
   Given the synaptic weights $w_{ij} \in \Re$ $(i = 1, 2, \ldots, N; j = 1, 2, \ldots, N)$ with $w_{ii} = 0$ and activities $S_i(t) = \pm 1$ $(i = 1, 2, \ldots, N)$ at discrete time t, write down the update equation which defines the activity $S_j(t+1)$ in the next time step.
   Explain (in words and math) why connections with $w_{ij} > 0$ can be interpreted as *excitatory synapses* in this model.

2) **Vector Quantization (1.5 points)**
   Assume a set of N-dimensional vectors $\vec{\xi}^{\mu} \in \mathbb{R}^N$ $(\mu = 1,2,\ldots,P)$ is given. Consider unsupervised Vector Quantization with a fixed number K of prototypes.

   a) Explain the term *Quantization Error* in terms of Euclidean distance. You do not have to reproduce the equation from the lecture notes exactly, you can use words, but be as precise as possible!
   b) Argue in terms of simple graphical illustrations (example situations) that the outcome of Vector Quantization can depend strongly on the shape of existing clusters of data. Explain how VQ could place prototypes in "empty" regions where no data is observed at all.
   c) Assume K can be chosen freely, which value of K and which configuration of prototypes gives the lowest possible Quantization Error?

3) Supervised Learning (1 point)

a) Name and briefly explain two possible aims of supervised learning from high-dimensional data.

b) Name and explain one algorithm that can be used for (one of) these aims. Explain your example algorithm in words. You do not have to specify mathematical update equations or provide pseudo-code, but it should become clear which aim the algorithm achieves and how it works.

4) Overfitting (1 point)

a) Explain in words the problem of overfitting in terms of polynomial regression (curve fitting) as an example. Also explain the terms bias and variance in this context. Provide graphical illustrations of example situations and explain them in words.

c) In Feedforward Neural Networks, the wrong choice of which parameter could lead to overfitting? Compare, in words, the effect of extreme choices of this parameter.

5) Normal distributions. Maximum likelihood estimation. (2 points)

The following two sets of one-dimensional feature vectors x originate from two normal distributions:

$S_1 = \{8, 3, 15, -6, -14, -1\}$

$S_2 = \{18, 1, 95, 65, 32, 42, 34, 12, 28, 52\}$

a) Estimate the parameters (mean and standard deviation) of these distributions, using maximum likelihood estimation.

b) Specify the class conditional probability density functions $p(x|\omega_1)$ and $p(x|\omega_2)$.

c) Estimate the prior probabilities $P(\omega_1)$ and $P(\omega_2)$.

d) Give expressions for the posterior probabilities $P(\omega_1|x)$ and $P(\omega_2|x)$.

6) Bayesian Decision Theory. Decision criterion. (1.5 points)

For the normal distributions determined in the previous problem:

a. Using the Bayesian theory, set up an equation which can be used to determine the decision criterion which must be used.

b. Simplify this equation to a quadratic equation for x and find its solutions that are the values of the decision criterion.

7) **Hierarchical clustering (1 point).**

	O_1	O_2	O_3	O_4	O_5	O_6
O_1		1	4	5	7	7
O_2			3	5	6	6
O_3				2	3	3
O_4					4	5
O_5						1
O_6						

The following upper triangular matrix describes the dissimilarities between six objects. Use the algorithm presented in the lectures to derive a dendrogram for these objects. Assume that the dissimilarity between two clusters of points is defined by the dissimilarity of their least dissimilar elements.

Math reminder:

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$x_{1,2} = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

x	0.5	1	1.5	2	2.5	3	3.5	4
$\exp(-\frac{1}{2}x^2)$	0.8825	0.6065	0.3247	0.1353	0.0439	0.0111	0.0022	0.0003
$\ln(x)$	-0.6931	0	0.4055	0.6931	0.9163	1.0986	1.2528	1.3863