Exam Pattern Recognition
Tuesday, March 12, 2013
14.00-17.00 hrs

General Remarks

1. Hand in your answers to part A (statistical pattern recognition) and part B (geometrical pattern recognition) on separate sheets of paper.
2. Put your name and student number on every sheet.
3. You are allowed to use a calculator.
   It is not allowed to consult books, notes, telephones, etc.
4. Always show how you arrived at the result of your calculations. Always explain your answer, used symbols, etc. Be precise.
5. If you only retake one of the two parts (either A or B) you have $1\frac{1}{2}$ hours time; otherwise you have 3 hours time.

Part A: Statistical Pattern Recognition

Question 1 Short Questions (16 points)

(a) (4 points) In neural networks, what is *weight decay*?

(b) (4 points) In support vector machines we minimize the objective function

$$C \sum_{n=1}^{N} \xi_n + \frac{1}{2}||\mathbf{w}||^2.$$

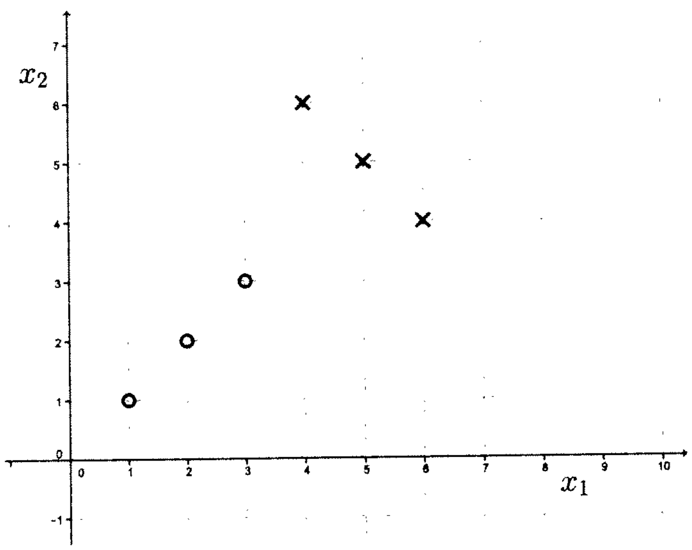
What is the purpose of C in this expression?

(c) (4 points) In linear regression, suppose we want to predict the sales price of a house, using the lot size (measured in square meters), and whether (desire.loc=1) or not (desire.loc=0) the house is built on a desirable location. We want to fit a model that has the following property: *the price per square meter does not depend on the location of the house, but people pay a fixed premium for a house on a desirable location.* Specify a regression function that captures this property.

(d) (4 points) What is the *curse of dimensionality*? Give an example of this phenomenon.

Question 2 Classification (20 Points)

We are given the following six data points:



The circles represent observations from class 1, and the crosses observations from class 2.

(a) (12 points) Give the linear discriminant functions $a_1(\mathbf{x})$ and $a_2(\mathbf{x})$, where the linear discriminant function of class k is given by:

$$a_k(\mathbf{x}) = \bar{\mathbf{x}}_k^\top \hat{\boldsymbol{\Sigma}}_{\text{pooled}}^{-1} \mathbf{x} - \tfrac{1}{2} \bar{\mathbf{x}}_k^\top \hat{\boldsymbol{\Sigma}}_{\text{pooled}}^{-1} \bar{\mathbf{x}}_k + \ln \frac{N_k}{N}.$$

(b) (2 points) Give the decision boundary $a(\mathbf{x}) = 0$ corresponding to your answer at (a).

(c) (3 points) Give the decision boundary produced by the support vector machine with linear kernel and perfect separation. You don't need to give a formal proof, an intuitive geometric argument to justify the given decision boundary suffices.

(d) (3 points) Give the support vectors corresponding to your answer at (c).

Question 3 Optimization (14 points)

Consider the logistic regression model

$$P(t_n = 1|x_n) = \frac{e^{w_0+w_1x_n}}{1+e^{w_0+w_1x_n}}, \qquad n = 1,\ldots,N,$$

and the two observations ($N = 2$):

n	x_n	t_n
1	2	0
2	3	1

To find values of w_0 and w_1 for which the cross-entropy error function (negative log likelihood) is at a minimum, we apply the method of gradient descent, with step size $\eta = 0.1$ and initial values $w_0^{(0)} = -4$ and $w_1^{(0)} = 3$.

(a) (2 points) Compute $P(t_n = 1|x_n)$ for $n = 1, 2$, using weight values $w_0^{(0)}$ and $w_1^{(0)}$. Give the result to an accuracy of two decimal places.

(b) (12 points) Find the value of $w_1^{(1)}$ using gradient descent on the two given observations. Clearly show how you arrived at the answer.

Hint: You can model logistic regression as a simple neural network, and use your knowledge of gradient descent in neural networks. The new weight value $w_1^{(1)}$ is obtained by processing the two observations together (batch processing) rather than one at a time (sequential processing).

If you don't get the hint, you may also explicitly use the cross-entropy error function for logistic regression, which is given by:

$$E(w_0, w_1) = \sum_{n=1}^{N} \left\{t_n \ln(1+e^{-w_0-w_1x_n}) + (1-t_n)\ln(1+e^{w_0+w_1x_n})\right\}$$

(Turn over for part B.)

Part B: Geometrical Pattern Recognition

1. (General) 5 points
   What is the formulation for an approximate optimization problem between two patterns A and B?

2. (Annulus) 10 points
   Describe an algorithm to compute the smallest width annulus of a finite set of points in the plane.

3. (Weighted point sets) 15 points

   Explain what the Earth Mover's Distance is and how it works. Is it for partial or complete, and 1-1 or n-m matching?

4. (Curve matching) 20 points

   (a) (10 pts) Give an algorithm to compute the Hausdorff distance between two sets of line segments. Draw a picture and explain why it works.

   (b) (10 pts) Explain that the lower left corner of the axis parallel bounding box is a reference point for finding an approximate translation that minimizes the Hausdorff distance between 2D curves. Explain what its approximation factor is, with the help of a picture.