
INSTITUTO SUPERIOR TÉCNICO
MACHINE LEARNING

!!! **Yellow** !!! Exame: 1 14 June 2021

- You have 2:00 h plus 15+15 min for authentication, upload and download. *Panic not.*
- The exam is open book
- 20 points total.
- Write your number and name at the top of each page.
- Write all formulas! Present all the computations and justifications for your answers.
- The exam should be hand written or with a tablet and pencil (hand written)

1. (3 pts) Linear Regression

a) (2pts)

Training set consists on 5 observations (sample) with dimension *D = 2*

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 8 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 0 \\ -10 \end{pmatrix}$$

and the corresponding target

$$t_1=1,\ t_2=2,\ t_3=-2\ t_4=2,\ t_5=4.$$

Compute the nonlinear regression with the basis function being the Euclidean length of the vector $\mathbf{x}_i$

$$\phi_1 = \|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2}$$

For example, with

$$\phi_{11} = \|\boldsymbol{x}_1\|_2 = \sqrt{1^2 + 9^2} = \sqrt{82} = 9.05539$$

and with l_2 regularization, $\log(\lambda/4) = 0$.

b) (1 pts)

How could we solve the nonlinear regression of with l_1 regularization?

2. (6 pts) Neural Network

Given the weights

$$W^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$b^{[1]} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$W^{[2]} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$b^{[2]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and the activation function: $f(x)$=exp(0.1*x) for all units/neurons, including in the output layer using the usual (as in the practical lectures) squared error loss / error

$$E[w] = \frac{1}{2}\sum (o - t)^2 = \frac{1}{2}\sum (t - o)^2$$

do one stochastic gradient descent update (with learning rate $\eta = 1$) for the training example:

$\mathbf{x}=(0,0,0,0,0)^T$ and the target $\mathbf{t}=(0,0)^T$

Determine the new weights and biases.

3. (5 pts) RBF

(a) (4 pts)

The training set is described by three vectors

$$\boldsymbol{x_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{x_2} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{x_3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and the corresponding target of two classes

$t_1 = 1, t_2 = 1, t_3 = -1$

Determine the parameters of the RBF network. Use an RBF with k-Means clustering with k=2 and σ =1 for the clusters and one output unit implemented as the original Rosenblatt perceptron.

The cluster centers are initialized with

$$\boldsymbol{c_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{c_2} = \begin{pmatrix} 0 \\ 0 \\ 1.1 \end{pmatrix}$$

*Hint you can determine the cluster centers of k-Means by drawing the projected values in 2 dim coordinate system. It is enough to indicate what are the cluster centers after applying k-Means algorithm**<u>. No computation for k-Means algorithm are required!</u>***

Initialize all weights of the perceptron to one (including the bias). Use a learning rate of one for simplicity. Apply the perceptron learning algorithm (the original Rosenblatt model with sign()) for one epoch.

(b) (1 pts)

If we used a perceptron on the original data (no hidden layer) could it solve the problem? Indicate why or why not.

4. (4 pts) Model Complexity

How many trainable parameters are in each of the following models with the input of 10x10 gray image:

- CNN with:
    - conv(window=2x3, stride=1, channels=20, no padding),
    - max_pooling(window=2, stride=2),
    - fully-connected output layer with 2 units
- EM Gaussian clustering with 4 clusters
- Bayesian Network where each pixel is a binary variable conditionally dependent on the pixel above it (if it exists) and on a binary class variable
- MLP with two nonlinear hidden layers [20,20] and one output unit.

5. (2 pts) VC Dimension

Consider a problem with two-dimensional inputs. Also consider two models:

1) Perceptron.
2) MLP with two hidden layers [4,3] and one output unit. All hidden units have a linear activation function f(x)=x and the output unit has the sign function as activation.

Use the fact that the VC dimension of a d-dimensional Perceptron is equal to d+1 to prove the true VC dimension of model 2.