Aprendizagem 2024
Lab 7-9: Neural Networks

Practical exercises

I. Multi-Layer Perceptron

1. Consider a network with three layers: 5 inputs, 3 hidden units and 2 outputs where all units use a sigmoid activation function.
   a) Initialize connection weights to 0.1 and biases to 0. Using the squared error loss do a stochastic gradient descent update (with learning rate η=1) for the training example
   $$\left\{x = [11000]^T, z = [10]^T\right\}$$
   b) Compute the MLP class for the query point $x_{new} = [10001]^T$

2. Consider a network with four layers with the following numbers of units 4, 4, 3, 3.
   Assume all units use the *hyperbolic tangent* activation function.
   a) Initialize all connection weights and biases to 0.1. Using the squared error loss do a *stochastic gradient descent* update (with learning rate η=0.1) for the training example:
   $$\left\{x = [1010]^T, z = [010]^T\right\}$$
   b) Reusing the computations from the previous exercise do a *gradient descent update* (with learning rate η=0.1) for the batch with the training example from the a) and the following:
   $$\left\{x = [00100]^T, z = [001]^T\right\}$$
   c) Consider the learned MLPs from a) and b). Which has smallest squared error? Which model has better classification accuracy?
   d) Compute the MLP class for the query point $x_{new} = [1110]^T$

3. Repeat the exact same exercise, but this time with following adaptations:
   - the output units have a *softmax* activation function
   - the error function is *cross-entropy*

   What are the major differences between using squared error and cross-entropy?

II. Model Complexity (*optional*)

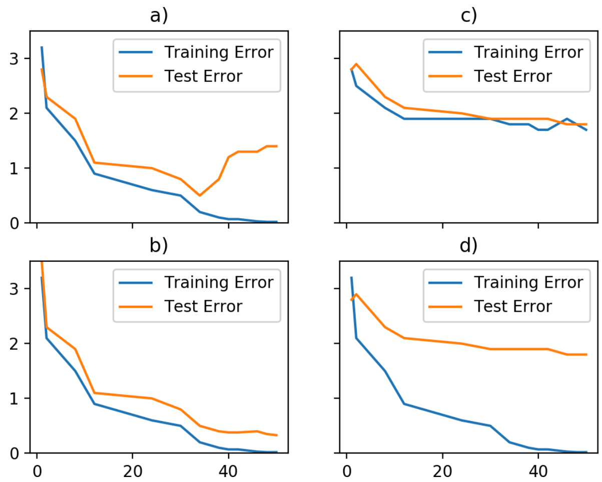
4. [*optional*] For the following scenarios which has the smallest number of parameters?
   a) three-dimensional real inputs classified by
      i. MLP with one hidden layer with the following units per layer 3 2 2
      ii. simple Bayesian classifier with multivariate gaussian likelihood function
   b) *N*-dimensional real inputs classified by
      i. perceptron
      ii. MLP with two hidden layers with the following units per layer N, $\frac{N}{2}$, $\frac{N}{2}$, 2
      iii. naive Bayes with Gaussian likelihoods

iv. simple Bayesian classifier with multivariate gaussian likelihood function

5. [*optional*] Choose between increase, decrease, maintain for each of the following factors:
   - training data
   - regularization
   - number of parameters

For each of the following four scenarios:

Justify each decision.



Programming quest

Resources: https://scikit-learn.org/stable/modules/neural_networks_supervised.html
as well as *Classification*, *Regression* and *Evaluation* notebooks

6. Consider a 10-fold CV, and MLPs with a single hidden layer with 5 nodes. Using *sklearn*:
   a) assess the classification accuracy of the MLP on the *iris* data using a cross-entropy loss
   b) assess the MAE of the MLP on the *housing* data using a squared error loss