TÉCNICO LISBOA

Aprendizagem 2021/22

Homework III

Deadline 5/11/2021 23:59 via Fenix as PDF

- Homework limited to 5 pages (3. 5 −4pp for part I, 1−1.5pp for part II) according to the provided template
- Include your programming code as an Appendix (no page limits)
- Submission Gxxx.PDF in Fenix where xxx is your group number. Please note that it is possible to submit several times on Fenix to prevent last-minute problems. Yet, only the last submission is considered valid
- Exchange of ideas is encouraged. Yet, if copy is detected after automatic/manual clearance, homework is nullified and IST guidelines apply for content sharers and consumers, irrespectively of the underlying intent
- Please consult the FAQ before posting questions to your faculty hosts

I. Pen-and-paper [10v]

1) Consider a MLP classifier characterized by the following weights:

$$\mathbf{W}^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \mathbf{b}^{[1]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{W}^{[2]} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \mathbf{b}^{[2]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{W}^{[3]} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{b}^{[3]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

   a. [5v] Using the hyperbolic tangent activation function $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = tanh(x)$ on all units, and the squared error loss, perform a stochastic gradient descent update (with learning rate $\eta = 0.1$) for the training example $\mathbf{x} = (1,1,1,1,1)^T$ with positive target, i.e. $\mathbf{z} = (1, -1)^T$.

   b. [5v] Replacing the activation function on the output unit by softmax and the loss function by cross-entropy, perform a stochastic gradient descent update (with learning rate $\eta = 0.1$) for the same example. Note: under softmax, a positive target is defined as $\mathbf{z} = (1,0)^T$.

II. Programming and critical analysis [10v]

Use **sklearn** to answer the following questions. Consider a MLP with l_2 regularization, RELU activation functions in the hidden layers, and an architecture resembling that described in Part I, i.e. two hidden layers of size 3 and 2. Consider all the remaining MLP parameters as the defaults in sklearn (e.g. cross-entropy loss for classification and squared error loss for regression).

2) [5v] Using the `breast.w.arff` data from previous homeworks, show the confusion matrix of the aforementioned MLP in the presence and absence of early stopping. Briefly enumerate two reasons for the observed differences.
3) [5v] Using the `kin8nm.arff`, plot the distribution of the residues using boxplots in the presence and absence of regularization. Identify 4 strategies to minimize the observed error of the MLP regressor.

Note: consider a 5-CV with a fixed zero seed to answer (3) and (4).
`Kin8nm.arff` available at https://fenix.tecnico.ulisboa.pt/downloadFile/845043405555949/kin8nm.arff

END