Aprendizagem 2025

Homework III

Deadline 15/10/2025 (Wednesday) 23:59 via Fenix as PDF

- Submit Gxxx.ZIP in Fenix where xxx is your group number. The ZIP should contain two files: Gxxx_report.pdf with your report and Gxxx_notebook.ipynb with your notebook demo according to the suggested templates
- It is possible to submit several times on Fenix to prevent last-minute problems. Yet, only the last submission is kept
- Exchange of ideas is encouraged. Yet, if copy is detected after automatic or 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 [12v]

For questions in this group, show your numerical results with 5 decimals or scientific notation.
Hint: we highly recommend the use of `numpy` (e.g., *linalg.pinv* for inverse) or other programmatic facilities to support the calculus involved in both questions (1) and (2).

Below is a training dataset D composed by two input variables and two output variables, one of which is numerical (y_{num}) and the other categorical (y_{class}). For this exercise, **consider a polynomial basis function** $\phi(y1,\ y2) = y1 \times y2$ that transforms the original space into a new one-dimensional space.

D	y_1	y_2	y_{num}	y_{class}
x_1	2	2	3.5	A
x_2	1	1	1.0	A
x_3	3	2	3.8	B
x_4	6	3	10.1	C
x_5	8	1	8.5	B

1. [2v] Learn a regression model on the **transformed feature space** using the OLS closed form solution to predict the continuous output y_{num}.

2. [2.5v] Repeat the previous exercise, but this time learn a Ridge regression with penalty factor $\lambda = 1$. Compare the learnt coefficients with the ones from the previous exercise and discuss how regularization affects them.

3. [2.5v] Given three new test observations and their corresponding y_{num} output $x_6 = (0,\ 2,\ 1)$, $x_7 = (3,\ 4,\ 6.2)$ and $x_8 = (5,\ 1,\ 3.6)$, compare the train and test MAE of the two models obtained in 1) and 2). Explain if the results go according to what is expected.

4. [5v] Consider an MLP to predict the output y_{class} characterized by the weights

$$W^{[1]} = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \\ 0.2 & 0.1 \end{bmatrix} \quad b^{[1]} = \begin{bmatrix} 0.1 \\ 0 \\ 0.1 \end{bmatrix} \quad W^{[2]} = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad b^{[2]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

the output activation function $softmax(Z_c^{[out]}) = \frac{e^{Z_c^{[out]}}}{\sum_{c'=1}^{|C|} e^{Z_{c'}^{[out]}}}$, the sigmoid activation function on the hidden layer(s) and the cross-entropy loss: $-\sum_{i=1}^{N}\sum_{c=1}^{|C|} t_c^{(i)} log(Z_c^{[out](i)})$, where t represents the target classes {A, B, C}. Consider also that the output layer of the MLP gives the predictions for the classes {A, B, C} in this order. Perform one stochastic gradient descent update to all the weights and biases with learning rate $\eta = 0.5$ using the training observation x_1. Explain, in words, what's the difference between using a sigmoid activation and no activation in terms of the representational capabilities of the learned classifier.

II. Programming and critical analysis [8v]

Consider the `rent.csv` dataset (available at the course's webpage), where the goal is to predict the price of rents based on location features and house measurements.

To answer question 5), average the performance of the models using 5-fold cross validation.

5. [2v] Train a Linear Regression model, an MLP Regressor with 2 hidden layers of 5 neurons each and no activation functions, and another MLP Regressor with 2 hidden layers of 5 neurons each using relu activation functions. Plot a boxplot of the test MAE of each model.

6. [3v] Compare a Linear Regression with a MLP with relu activation, and explain the impact and the importance of using activation functions in a MLP. Support your reasoning with the results from the boxplots.

7. [3v] With a MLP with relu activation, plot the loss function for each training iteration for both the training and validation set. Take the average across different folds. Compare the results and explain whether the model is overfitting, underfitting, or has good generalization.

END