- 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).

1. Consider the problem of learning a regression model from 4 bivariate observations
   $\left\{ \begin{pmatrix} 0.7 \\ -0.3 \end{pmatrix}, \begin{pmatrix} 0.4 \\ 0.5 \end{pmatrix}, \begin{pmatrix} -0.2 \\ 0.8 \end{pmatrix}, \begin{pmatrix} -0.4 \\ 0.3 \end{pmatrix} \right\}$ with targets $(0.8, 0.6, 0.3, 0.3)$.

   a. [4v] Given the radial basis function, $\phi_j(x) = exp\left(-\frac{\|\mathbf{x} - c_j\|^2}{2} \right)$, that transforms the original space onto a new space characterized by the similarity of the original observations to the following data points, $\left\{ c_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, c_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, c_3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$.

      Learn the Ridge regression (l_2 regularization) using the closed solution with $\lambda = 0.1$.

   b. [2v] Compute the training RMSE for the learnt regression.

2. [6v] Consider a MLP classifier of three outcomes $- A$, B and $C -$ characterized by the weights,

   $$W^{[1]} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, b^{[1]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, W^{[2]} = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}, b^{[2]} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, W^{[3]} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}, b^{[3]} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

   the activation $f(x) = \frac{e^{0.5x-2} - e^{-0.5x+2}}{e^{0.5x-2} + e^{-0.5x+2}} = tanh(0.5x - 2)$ for every unit, and squared error loss

   $\frac{1}{2}\|\mathbf{z} - \hat{\mathbf{z}}\|_2^2$. Perform one batch gradient descent update (with learning rate $\eta = 0.1$) for training

   observations $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ and $\mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$ with targets B and A, respectively.

II. Programming and critical analysis [8v]

Consider the `winequality-red.csv` dataset (available at the webpage) where the goal is to estimate the quality (sensory appreciation) of a wine based on physicochemical inputs.

Using a 80-20 training-test split with a fixed seed (`random_state=0`), you are asked to learn MLP regressors to answer the following questions.

Given their stochastic behavior, average the performance of each MLP from 10 runs
(for reproducibility consider seeding the MLPs with `random_state` $\in \{1..10\}$).

1) [3.5v] Learn a MLP regressor with 2 hidden layers of size 10, rectifier linear unit activation on all nodes, and early stopping with 20% of training data set aside for validation. All remaining parameters (e.g., loss, batch size, regularization term, solver) should be set as default. Plot the distribution of the residues (in absolute value) using a histogram.

2) [1.5v] Since we are in the presence of a *integer regression* task, a recommended trick is to round and bound estimates. Assess the impact of these operations on the MAE of the MLP learnt in previous question.

3) [1.5v] Similarly assess the impact on RMSE from replacing early stopping by a well-defined number of iterations in {20,50,100,200} (where one iteration corresponds to a batch).

4) [1.5v] Critically comment the results obtained in previous question, hypothesizing at least one reason why early stopping favors and/or worsens performance.

END