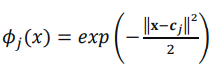
* 1. **Pen-and-paper** [12v]

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).

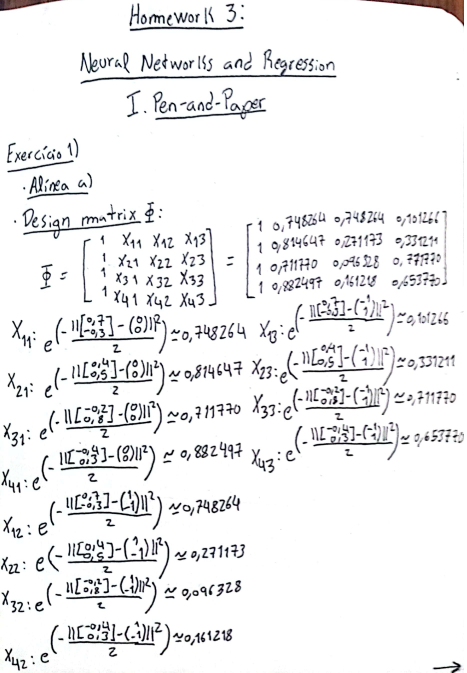
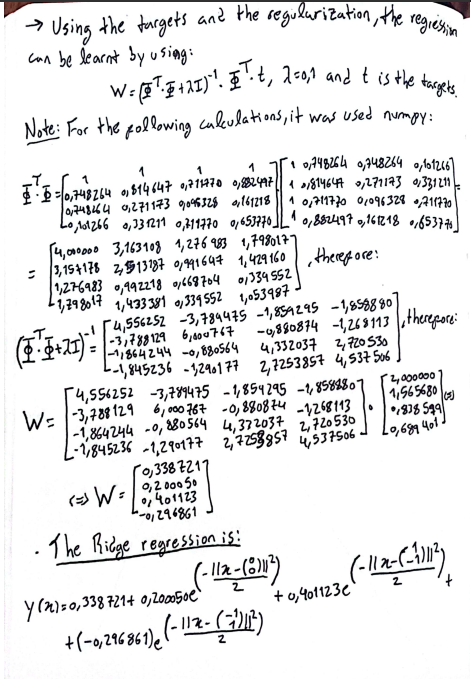
1. Consider the problem of learning a regression model from 4 bivariate observations



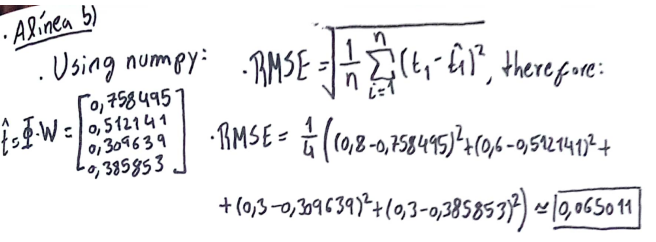


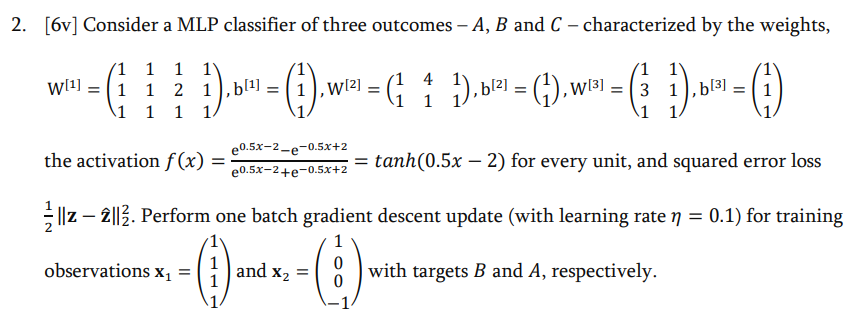
* 1. Uma imagem com Tipo de letra, file, escrita à mão, tipografia

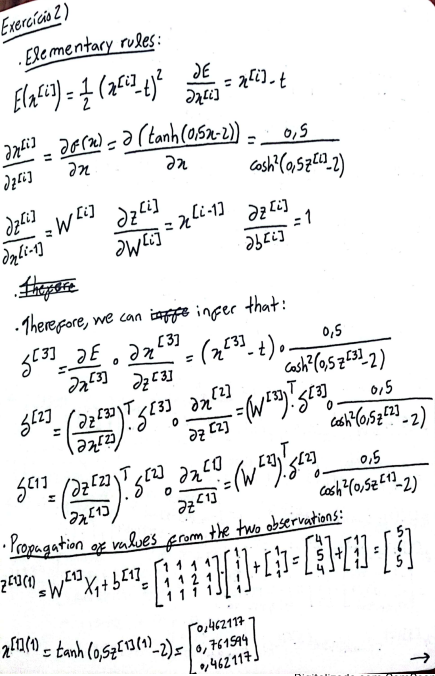
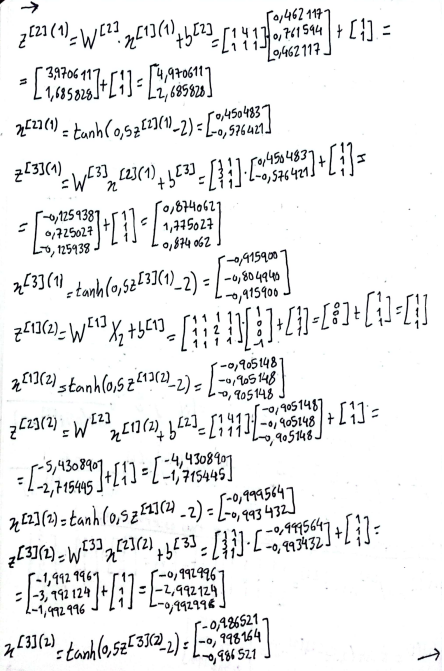
     Descrição gerada automaticamente[4v] Given the radial basis function, , that transforms the original space onto a new space characterized by the similarity of the original observations to the following data points,



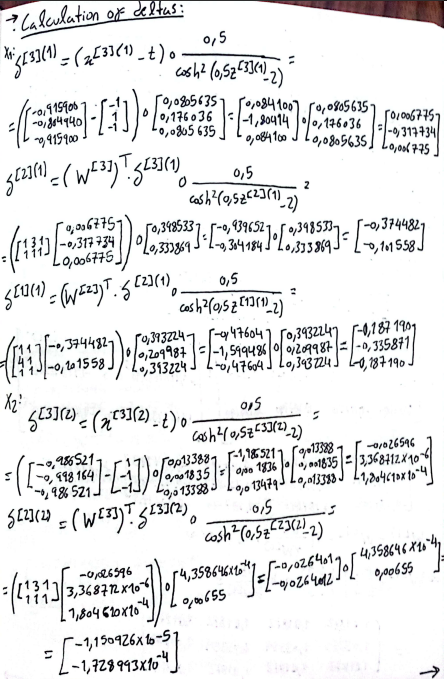
* 1. [2v] Compute the training RMSE for the learnt regression.

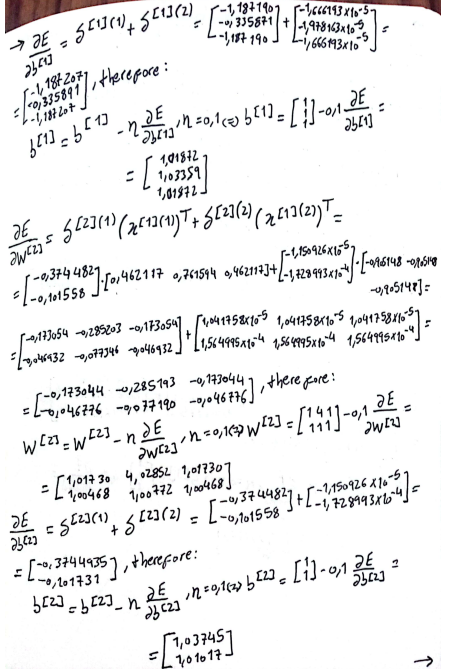
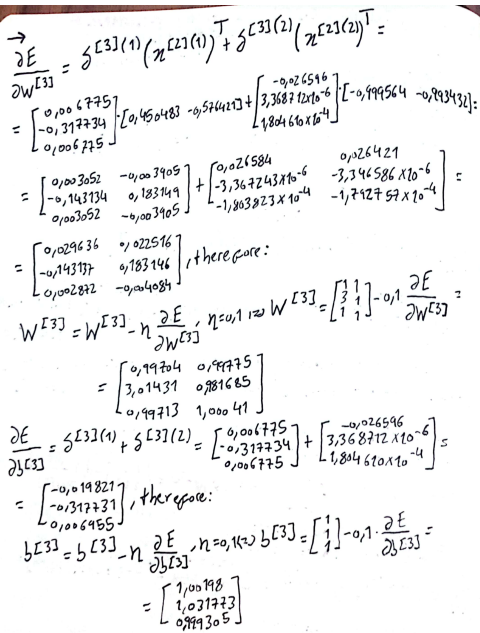






Uma imagem com texto, escrita à mão, Tipo de letra, tinta

Descrição gerada automaticamente



# Programming and critical analysis [8v]

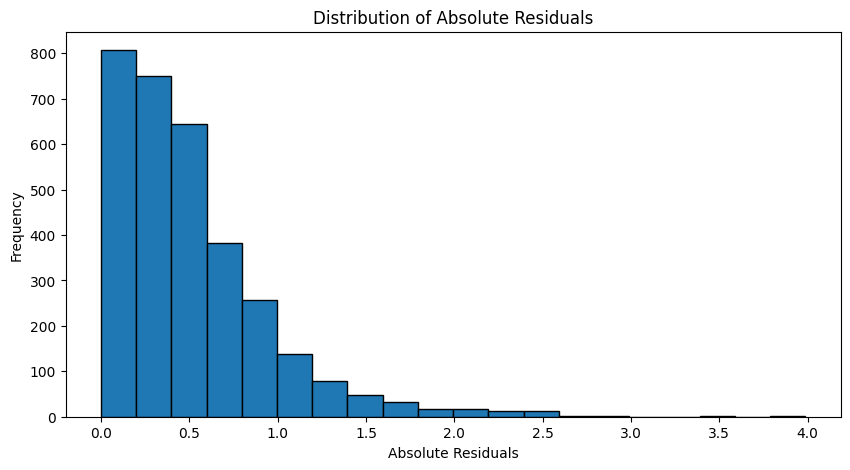
Consider the winequality-red.csv dataset (available at the webpage) where the goal is to estimate the quality (sensory appreciation) of a wine based on physicochemical inputs.

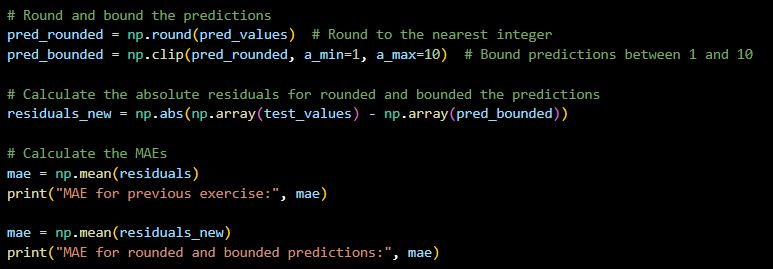
Using a 80-20 training-test split with a fixed seed (random\_state=0), you are asked to learn MLP regressors to answer the following questions.

Given their stochastic behavior, average the performance of each MLP from 10 runs (for reproducibility consider seeding the MLPs with random\_state ∈ {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.





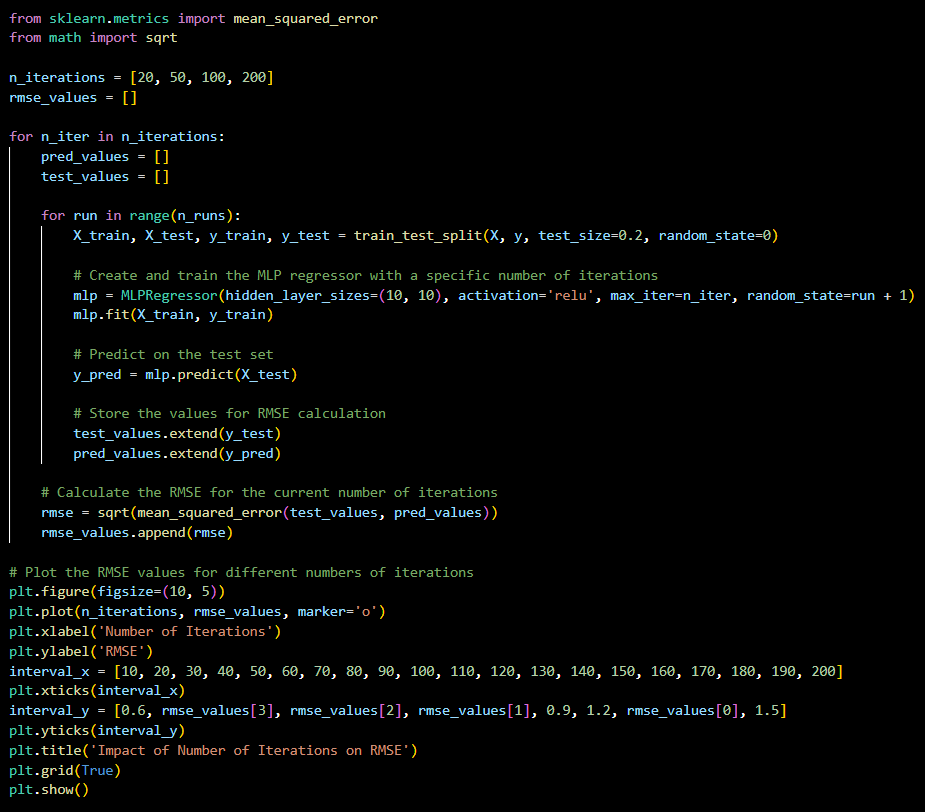
1. [1.5v] Since we are in the presence of a *integer regression* task, a recommended trick is to round and bound estimates. Assess the impact of these operations on the MAE of the MLP learnt in previous question.

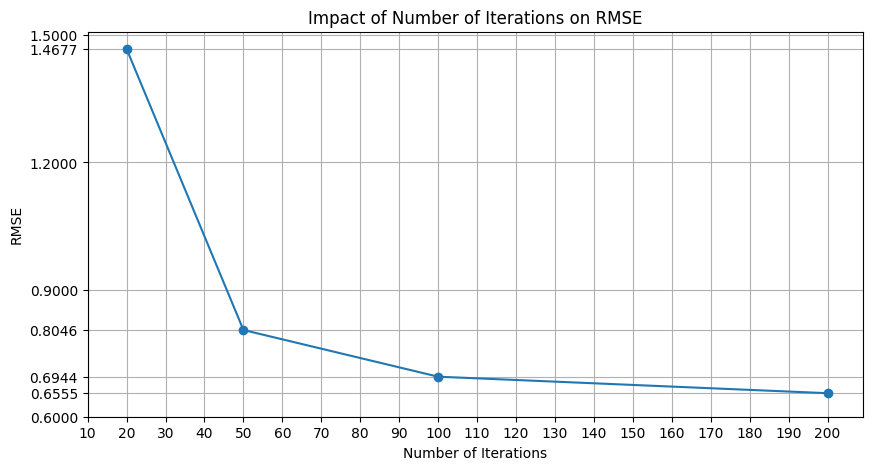
MAE for previous exercise: 0.5097171955009514

MAE for rounded and bounded predictions: 0.43875

**Comment about results:**

By rounding and bounding the estimates, it is observed that the MAE (Mean Absolute Error) reduces in relation to the previous exercise, which indicates that there was a better performance of this new model.

1. [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).



1. [1.5v] Critically comment the results obtained in previous question, hypothesizing at least one reason why early stopping favors and/or worsens performance.

When the number of iterations is low (20 iterations), the model doesn't have a chance to adjust the weights to completely fit the training data, which leads to a relatively high RMSE (≈1.47). As the number of iterations increases, the model has more opportunities to adjust the weights and reduce the error in the training data, thus decreasing the RMSE. However, if training continues for too many iterations, the model may start to adjust too much to the training data. As it is possible to observe from the exercise above, for a number of iterations equal to 200 a relatively low RMSE (≈0.66) was obtained. When this happens, there may be an increase in error in the test data, thus being in the presence of overfitting.

With this, we conclude that early stopping, on the one hand, is an effective technique to avoid overfitting. However, when we apply this technique, the model may not be able to converge to its optimal performance, due to its training being stopped too early.

# END