M1 Master's Degree in Cognitive Science 2024–2025 | December 19^{th}
TA: Hugo Ninou (hugo.ninou@ens.fr)

MATHEMATICAL TOOLS FOR DATA SCIENCE, MACHINE LEARNING AND STATISTICAL MODELLING

TD3: LINEAR REGRESSION

Click here to access the notebook for this tutorial.

You should copy the notebook from Google colab (go to file->save a copy in drive) in order to be able to save your changes.

1 Decoding a Monkey's Intent: Predicting Movements Before They Happen

In this tutorial, we will work with the same dataset as in HW2. Previously, our objective was to perform a principal component analysis (PCA) on this dataset to visualize neural activity during movement preparation. To recap, the data was collected from a monkey performing a center-out arm-reaching task. The monkey was holding a controller and was first required to keep the controller in a central position. Then, a cue was presented, indicating one of eight possible targets arranged in a circular pattern around the center. During this "CUE" period, the monkey was instructed to maintain the controller in the central position. When the "GO" signal was given, the monkey moved the controller toward the cued target.
Neural recordings were obtained during this task (courtesy of Yifat Prut's lab), resulting in data from a total of 777 recorded neurons. For each neuron, the data is represented as a matrix of size (#trials × #time steps), where each entry is 1 if the neuron was firing at that time step and 0 otherwise.

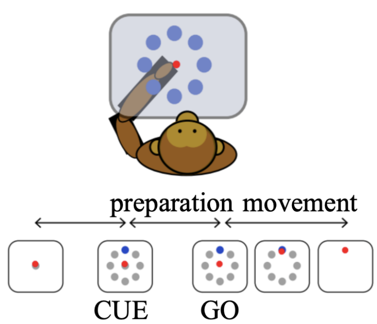

Figure 1: Center-out reaching task

The goal of this tutorial is to train a model that can predict the target the monkey is aiming for based solely on its preparatory neural activity. In other words, the model will predict the direction of the monkey's movement before the movement begins.

1. (a) Perform a PCA on the neural states recorded 300 ms before the GO signal for neurons with at least 20 trials per target.
   (b) Plot the neural states along the 3 first principal components (PCs).
   (c) How are the neural states arranged in this representation ?
   (d) What is the variance explained by the 3 first PCs ?

1.1 Linear Regression

We will start by using linear regression as the model.

2. (a) Define the targets and predictors for this linear regression model ?
   (b) What equation must the model coefficients verify ?
3. (a) Solve this equation assuming that $X_{\text{train}}^{\intercal} X_{\text{train}}$ has linearly independent columns.
   (b) Split the data into a training and a testing set using an 80%-20% ratio.
4. (a) Show the matrix "$(X_{\text{train}}^{\intercal} X_{\text{train}})^{-1}(X_{\text{train}}^{\intercal} X_{\text{train}})$".
   (b) What issue arises here ?
5. What method would you suggest to compute the coefficients in this case ?
6. Alternatively, how could you modify the matrix $X_{\text{train}}^{\intercal} X_{\text{train}}$ to ensure it has linearly independent columns?
7. (a) Implement linear regression and plot the results of the model on both the training and testing sets in the same graph.
   (b) Comment on the results.

1.2 Ridge Regression

Ridge regression modifies linear regression by adding a L2 regularization term to the coefficients. It minimizes the cost function $\|X - \beta Y\|_2^2 + \alpha \|\beta\|_2^2$ where $\|.\|_2$ denotes the L2 norm.

8. Does increasing the α parameter increase or decrease the complexity of the model ?
9. Perform 100 cross-validation steps to compute the scores for linear regression.
10. Repeat the cross-validation for Ridge regression, with 50 logarithmically spaced values of α ranging from 10^{-6} to 10^{-1}.
11. Plot the cross-validated scores for Ridge regression and linear regression on the same graph.
12. (a) What is the optimal α as for the score ?
    (b) Plot the results for one iteration of the optimal ridge regression model on both the training and testing sets in the same graph.
    (c) Comment on the results.

1.3 Lasso Regression

Lasso regression replaces Ridge regression's L2 regularization with an L1 regularization term. It minimizes the cost function $\|X - \beta Y\|_2^2 + \alpha \|\beta\|_1$ where $\|.\|_1$ denotes the L1 norm.

13. (a) Implement the Lasso regression model with parameter $\alpha = 10^{-4}$ and print the coefficients.
    (b) What major difference do you observe in the coefficients compared to Ridge regression ?

14. (a) Perform 100 cross-validation steps for Lasso regression with 50 logarithmically spaced values of α ranging from 10^{-5} to 10^{-3}.
    (b) What is the optimal α for Lasso regression ?

15. Compare the cross-validation results of Lasso regression with Ridge regression and standard linear regression.