[Lab/Homework] Non-regularized regression

Jae Yun JUN KIM*

November 2, 2020

Due: Before Monday November 9, 2020, 13h (before the next lab session).
Evaluation: Submit code, explanation about the code, and the answers to the questions
Remark:

- Only groups of two/three people accepted. Forbidden groups of one or larger number of people.
- No late homework will be accepted.
- No plagiarism. If plagiarism happens, both the "lender" and "borrower" will have a zero.
- Code yourself from scratch. No homework will be considered if you solve the problem using any ML library.
- Do thoroughly all the demanded tasks.
- Study the theory for the questions.

1 Tasks

1) Read the dataset given in the provided file `data.csv` and plot the output value as a function of the input data.

2) Suppose that we would like to design a non-regularized regressor for a single-period forecast of the 1D position of a robot as function of time by training the regressor with the provided data. For this, construct first the training input and output matrices ($X_{\text{train}} \in \mathbb{R}^{I\times N}$ and $Y_{\text{train}} \in \mathbb{R}^{I\times J}$) and the test input and output matrices ($X_{\text{test}\in\mathbb{R}^{3\times N}}$ and $Y_{\text{test}} \in \mathbb{R}^{3\times J}$), where I is the number of training examples, N is the number of features (or the input-variable dimension), and J is the output-variable dimension. Let for instance N be 150 and J be 1.

3) Fit the univariate linear regression parameters to the dataset using batch gradient descent (BGD). What are the optimal values of the parameters?

4) Fit the univariate linear regression parameters to the dataset using stochastic gradient descent (SGD). What are the optimal values of the parameters?

5) Fit the univariate linear regression parameters to the dataset using the closed-form solution (CFS). What are the optimal values of the parameters?

6) Plot the linear regressors obtained in 3), 4) and 5) over the original dataset.

7) Test your model using the test data using the three different methods (BGD, SGD, and CFS). Plot also these results.

8) Repeat the tasks 2), 5), 6)(only with CFS), and 7) with $J = 30$ for a multi-period forecast.

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr