[Lab 1] Non-regularized regression

Jae Yun JUN KIM*

September 10, 2020

Due: Before the next lab session.
Evaluation: Code and explanation about the code (in groups of two or three people)
Remark:

- Only groups of two or three people accepted. Forbidden groups of larger number of people.
- No late homework will be accepted.
- No plagiarism. If plagiarism happens, both the "lender" and the "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

Note: During the lab session, show 1) and 2) to the professor. They will be graded.

1) Read the dataset given in the provided file `microsoft.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 stock prices 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. What are the optimal values of the parameters?

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

5) Fit the univariate linear regression parameters to the dataset using the closed-form method. 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 choosing yourself some new input data. Plot also these results.

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

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