**Assignment-1**

Perform **Multivariate linear regression** using gradient descent on the Air Quality dataset and report Mean Squared Error (MSE) for training and testing dataset.

*One student from a group will submit the work. I will create the submission portal via Canvas. Please note that I will not mark any email attachments. For any submission, I will mark only whatever I will get in Canvas. I am delaying the assignment post in Canvas because the group formation is not finalized yet.*

* Use the top 75% of data for training and the rest 25% for testing.
* Apply feature scaling and feature selection to maximize the test performance.
* Do not use any machine learning library.
* Submit **a report, Jupyter notebook, and processed dataset files**. The report should explain the algorithm, codes, and related figures and tables. Try to report how you have experimented (trial and error) to achieve the best performance.
* Mark distribution is 70% on code and 30% on reporting.
* Use Canvas's "discussions" tab for any query so that everyone will see the ongoing discussion. Do not write emails to me about this.
* All members within a group will collaborate, but two groups must **not** cooperate.

Dataset: <https://archive.ics.uci.edu/dataset/360/air+quality>

MSE: <https://statisticsbyjim.com/regression/mean-squared-error-mse/>

Report template: <https://www.overleaf.com/read/wjqngsgtdcpv#604696>

**Assignment-2**

Perform **regularization**on Multivariate linear regression and logistic regression using gradient descent. Use the Air Quality dataset for linear regression and report Mean Squared Error (MSE) for the training and testing dataset. Similarly, use the [Smarket.csv](https://northsouth.instructure.com/courses/30924/files/2419650?wrap=1) (CSE445\code\data\Smarket.csv) dataset for the binary classification problem and report accuracy for the training and testing dataset.

One student from a group (of three people max) will submit the work.

* Use the top 75% of data for training and the rest 25% for testing.
* Apply feature scaling and feature selection to maximize the test performance.
* Do not use any machine learning library.
* Submit a report, Jupyter notebook, and processed dataset files. The report should explain the algorithm, codes, and related figures and tables. Try to report how you have experimented (trial and error) to achieve the best performance.
* Mark distribution is 70% on code and 30% on reporting.
* Use the "discussions" tab of Canvas for any query so that everyone will see the ongoing discussion. Do not write emails to me about this.
* All members within a group will collaborate, but two groups should not cooperate with each other.

Dataset: https://archive.ics.uci.edu/dataset/360/air+quality

MSE: https://statisticsbyjim.com/regression/mean-squared-error-mse/

Report template: <https://www.overleaf.com/read/wjqngsgtdcpv#604696>