BIL470/570 Machine Learning
HW - 1
Assigned: 18.10.2021
Due: 01.11.2021
Submission: Will be announced on piazza.
Regulations: Late submissions are not allowed. Plagiarism is strictly forbidden, all that take part will be punished according to the regulations of the university.

In this assignment you are going to model the two data sets given below by using logistic regression:

$$\hat{f}(x) = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_p x_p)$$

where x is a training sample vector of size p and σ is the sigmoid function. Use the log-likelihood as the cost function. (Remember that you should maximize this function).

The following are the steps to complete the assignment for each task given below:

1. Find β_i by all gradient ascent variants we discussed in class, separately: batch, stochastic and mini-batch. Note that, here you have to come up with a strategy to decide when to stop the optimization iterations and the value of the learning rate.
2. Find the performance on test data in terms of accuracy for each epoch of learning, where an epoch is one pass of the entire training data.
3. Draw curves where y-axis shows the test accuracy and x-axis shows the epochs. Add these to a report that also includes how you decide on stopping the iterations, the mini-batch size, how you initialize the parameters, learning rate, etc. Use Latex for the report (`http://overleaf.com`).

For all the coding use Python as the programming language. No libraries other than numpy and matplotlib are allowed.

Dataset :

`https://www.kaggle.com/c/tabular-playground-series-oct-2021`[1]

At the end, submit both the report and the code that you used to find the results. Use and follow piazza for questions, comments and updates.

[1]This is an ongoing Kaggle competition dataset. You can submit your results to the competition to see how you perform. But it is optional.