BIL470/570 Machine Learning
HW - 1
Assigned: 25.1.2019
Due: 10.2.2019
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 optimization function. (Remember that you should maximize this function).

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

1. Take the first 20% of the data as the test data and the remaining as the training data.

2. Find β_i by mini-batch gradient descent. Note that, here you have to come up with a strategy to decide when to stop the optimization iterations. Also note that, you can not do this by checking the test error. (Why?)

3. Find the performance on test data in terms of accuracy for each iteration of learning.

4. Draw a curve where y-axis shows the test accuracy and x-axis shows the iteration. Add this to a one-page report that also includes how you decide on stopping the iterations, the mini-batch size, how you initialize the parameters, etc.

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

Tasks :

1. Dataset 1 : `https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+(Sonar,+Mines+vs.+Rocks)`

2. Dataset 2 : `https://archive.ics.uci.edu/ml/datasets/Ionosphere`

At the end, submit both the 1 page report and the code that you used to find the results. Use and follow piazza for questions, comments and updates.