

Problem Statement

1. Apply the logistic regression method using the functions in the notebook «Logistic Regression as a Neural Network – BP alg.ipynb» to predict the biological response of a molecule (data: bioresponse.csv, description from Kaggle: *“The data is in the comma separated values (CSV) format. Each row in this data set represents a molecule. The first column contains experimental data describing a real biological response; the molecule was seen to elicit this response (1), or not (0). The remaining columns represent molecular descriptors (d1 through d1776), these are calculated properties that can capture some of the characteristics of the molecule - for example size, shape, or elemental constitution. The descriptor matrix has been normalized.”*). Use 75% of the sample to train the model, and the rest of the data to estimate its accuracy.
2. Modify optimize() function to implement the stochastic gradient descent (SGD) method. Apply it to solve the problem from p.1.
3. For two modifications of gradient descent (pp. 1 and 2), plot the learning curves (dependence of the value of the loss function on the iteration number), apply models with different values of the learning rate (at least 5 different learning rates). How does it affect the accuracy of the model?
4. *(not necessary) Implement the Adam optimization method using the numpy library and compare the accuracy of the model trained with it with the models trained by the classic GD and SGD algorithms.