Machine Learning, 2018-19

Coursework Part II

Description

The final part of your coursework will be an implementation of logistic regression. This is based on the data provided by the Wikipedia example "Probability of passing an exam versus hours of study" (https://en.wikipedia.org/wiki/Logistic_regression).

Plagiarism: please make sure that the material you submit has been created by you. Any sources you use for code should be properly referenced. Your code will be checked for plagiarism using appropriate software.

Deliverables

The deliverables for your coursework should include:

1. A jupyter notebook, including the code you are submitting for the assignment.
2. A report (preferably PDF) that contains a brief write-up on your implementation, and answers for a set of free-form questions (more details below).

Data

This data is based on the example given in the Wikipedia entry.

```
X = np.asarray ([[0.50], [0.75], [1.00], [1.25], [1.50], [1.75],
[3.00], [4.75], [1.75], [3.25], [5.00], [2.00], [2.25], [2.50],
[2.75], [3.50], [4.00], [4.25], [4.50], [5.50]])

 y = np.asarray([0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1,0,1,0,1,1,1,1,1,1])
```

Tasks

The following details the list of tasks you need to complete for the assignment. Tasks with the [**code**] tag should be implemented in your notebook, and tasks with the [**question**] tag should be part of your report. If you are asked to plot a graph, include the graph in your report and the code in your notebook. You do not need to implement cross-validation in this coursework, simply fit the data.

1. [**code**] Implement logistic regression in python using batch gradient descent. Use your code to fit the data given above. Make sure you save the value of your loss function on each iteration in a data structure (e.g., list).

2. [**question**] After how many iterations, and for which learning rate (α) did your algorithm converge? Plot the loss function with respect to iterations to illustrate this point.

3. [**question**] What happens if α is too large? How does this affect the loss function? Plot the loss function with respect to iterations to illustrate this point.

4. [**question**] Assume that you are applying logistic regression to the iris (flower) dataset, as in the previous assignment. Answer the following questions:

   (a) How would your hypothesis function change in this case and why?

   (b) How would you utilize your implementation of logistic regression in order to perform (multi-class) classification on the iris dataset? Include some pseudocode while discussing your approach.

Note: Your code should run without issues on lab machines! Also, remember to include a short write-up of your implementation in your report