Assignment 1: Adaline and Logistic Regression

Purposes

- Understanding Adaline and logistic regression algorithms.
- Investigating the performance difference of the two algorithms.
- Understanding the linear classification problem.
- Understanding mini-batch Stochastic Gradient Descent (SGD).

Implementation Tasks

1. (3 points) Modify the classes `AdalineGD` and `LogisticRegressionGD` in the textbook such that the bias data field `b_` is absorbed by the weight vector `w_`. Your program is required to be compatible with the training programs in the textbook.
2. (2 points) Compare the performance of Adaline and logistic regression (bias absorbed versions) on the Iris and Wine datasets that can be obtained from the UCI machine learning repository. You may use the Python program given in our textbook (Page 117) to import the datasets.
   - Iris dataset - You may consider the samples with the labels *setosa*, *versicolor* to form a training set for binary classification.
   - Wine dataset - You may consider the samples with in the first two classes (1 and 2) to form a training set for binary classification.

   The comparisons should be done based on the convergence of the *loss*. In order to make apple-to-apple comparisons, you should use the same hyperparameters and number of epochs for both learning algorithms.
3. (2 points) Adaline and perceptron learning can only be used for binary classification, however, the Iris dataset has 3 classes: *setosa*, *versicolor* and *virginica*. If you are only allowed to use perceptrons but the number is not limited, how would you like to perform a multiclass classification for the whole Iris data set? Please write a program for this task.
4. (3 points) *Stochastic gradient descent (SGD)* and *mini-batch gradient descent* are described in our textbook (Page 45). Their combination is called mini-batch SGD which has been intensively used to handle large-scale datasets in machine learning. Please implement a new function `fit_mini_batch_SGD` in the logistic regression class for mini-batch SGD. Choose a small batch size, e.g., 32, and compare the performance of GD, SGD and mini-batch SGD in the aspects of time cost and loss convergence speed. The comparison should be made based on the same hyperparameters. You may use the Wine dataset.

Please do not use `scikit-learn` in this assignment.

Report

You are required to write a report for this assignment. It should include the following parts:

- For Task 1, explain how the bias is transformed to an extra weight and why the translated model is equivalent to the original one.
- For Task 2, explain the comparisons using figures.
- For Task 3, explain the method you developed for multiclass classification and its correctness.
- For Task 4, explain the comparisons using figures.