Assignment 1: Studies of Perceptron Learning and Adaline

Purposes

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

Implementation Tasks

1. (6 points) Use the Python programs for perceptron learning and Adaline given in the textbook and do the comparisons on the *loss*, *number of updates*, *margin* of the resulting separation hyperplane, and the *convergence*. What are the conclusions you can get? Can you provide any mathematical explanations to support your conclusions? You may use the Iris dataset (the first 2 classes with all features) in the test and are required to give **at least 3 mathematically different conclusions**. In order to make apple-to-apple comparisons, you should use the same initial parameter values, training data, learning rate and number of epochs for both perceptron learning and Adaline.
2. (4 points) Modify the class Perceptron given 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 program in the textbook.
3. (6 points) A perceptron 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 Iris data set (all features)? Please write a program (demo) for this task using the new perceptron class developed in the previous task.
4. (4 points) In the textbook, both SGD and mini-batch GD are explained. Let us write a new training function `fit_mini_batch_SGD` in the class `AdalineSGD` such that the new function combines SGD and mini-batch GD in training. In every epoch, the mini-batch SGD method updates the learning parameters based on a randomly selected subset of the training data. The size (number of samples) of the subset is called the *batch size* which is a hyperparameter decided by the user.

Report

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

- For Task 1, you need to explicitly give the conclusions and their explanations.
- For Task 2, explain how the bias is transformed to an extra weight and why the translated model is equivalent to the original one.
- For Task 3, explain the method you developed for multiclass classification and its correctness.
- For Task 4, compare the performance of mini-batch SGD to SGD and GD based on the same initial parameter values, training data, learning rate and number of epochs.