Name: Dat Quoc Ngo

NET-ID: dqn170000

Course #: CS6364.002

CS 6364 Homework 4

September 28, 2021

Deadline for the first submission: Oct-11-2021.  
All assignments MUST have your name, student ID, course name/number at the beginning of your documents. Your homework MUST be submitted via Blackboard with file format and name convention as follows:

1. Q1  (Linear Regression): Use the python library (sklearn.linear model) to train a linear regression model for the Boston housing dataset:

Text

Description automatically generated

1. Q2  Implement the following five algorithms to train a linear regression model for the Boston housing data set https://towardsdatascience.com/linear-regression-on-boston-housing-dataset-f409b7e4a155

Split the dataset to a training set (70% samples) and a testing set (30% samples). Report the root mean squared errors (RMSE) on the training and testing sets.

1. The gradient descent algorithm

Num\_iter = 50

Learning\_rate = 1e-7

Text

Description automatically generated  
2. The stochastic gradient descent (SGD) algorithm

Num\_iter = 50

Learning\_rate = 3e-5

Text

Description automatically generated

3. The SGD algorithm with momentum

Learning\_rate = 3e-5

Num\_iter = 50

Momentum = 0.9

Text

Description automatically generated  
4. The SGD algorithm with Nesterov momentum

Learning\_rate = 3e-5

Num\_iter = 50

Momentum = 0.9

Text

Description automatically generated

5. The AdaGrad algorithm

Learning\_rate = 1e-2

Num\_iter = 100

Text

Description automatically generated

1. Q3  (Logistic Regression): Use the python library (sklearn.linear model) to train a logistic regression model for the Titanic dataset:

https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed.

Split the dataset to a training set (80% samples) and a testing set (20% samples). Report the overall classification accuracies on the training and testing sets and report the precision, recall, and F-measure scores for each of the two classes on the training and testing sets.

Calendar

Description automatically generated

1. Q4  (Logistic Regression): Implement the following five algorithms to train a logistic regression model for the Titantic dataset:

https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed.

Split the dataset to a training set (80% samples) and a testing set (20% samples). Report the overall classification accuracies on the training and testing sets and report the precision, recall, and F-measure scores for each of the two classes on the training and testing sets.

1. The gradient descent algorithm

Learning\_rate = 1e-3

Num\_iter = 50

Calendar

Description automatically generated

Calendar

Description automatically generated  
2. The stochastic gradient descent (SGD) algorithm

Learning\_rate = 3e-5

Num\_iter = 50

Calendar

Description automatically generated

3. The SGD algorithm with momentum

Learning\_rate = 3e-5

Num\_iter = 50

Momentum = 0.9

Calendar

Description automatically generated

4. The SGD algorithm with Nesterov momentum

Learning\_rate = 3e-5

Num\_iter = 50

Momentum = 0.9

Calendar

Description automatically generated

1. The AdaGrad algorithm

Learning\_rate = 1e-2

Num\_iter = 100

Calendar

Description automatically generated

Q4 (Bonus Question): You will get an additional full point (1.0) if you can answer this bonus question correctly. That means, if you answer Q1-Q4 correctly, you get a full point (1.0) for this HW assignment. If you can answer Q1-Q5 correctly, you will get 2.0 points.

1. Implement the Adam algorithm to train a linear regression model for the Boston housing data set https://towardsdatascience.com/linear-regression-on-boston-housing-dataset-f409b7e4a155

Learning\_rate = 1e-2

Num\_iter = 50

Text

Description automatically generated

1. Implement the Adam algorithm to train a logistic regression model for the Titantic dataset: https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed. Split the dataset to a training set (80% samples) and a testing set (20% samples). Report the overall classification accuracies on the training and testing sets and report the precision, recall, and F-measure scores for each of the two classes on the training and testing sets.

Learning\_rate = 1e-2

Num\_iter = 50

A picture containing calendar

Description automatically generated