

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:

HW#_Name_writeup.pdf (for writing part)

HW#_Name_code.zip (for coding part)

If you have any questions, please contact me.

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

<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.

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
2. The stochastic gradient descent (SGD) algorithm
3. The SGD algorithm with momentum
4. The SGD algorithm with Nesterov momentum
5. The AdaGrad algorithm

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.

Q4 (Logistic Regression): Implement the following five algorithms 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.

1. The gradient descent algorithm
2. The stochastic gradient descent (SGD) algorithm
3. The SGD algorithm with momentum
4. The SGD algorithm with Nesterov momentum

5. The AdaGrad algorithm

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>
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.
2. Implement the Adam algorithm 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.