

CSCI555.20 Data Mining and Machine Learning

Winter 2025, Assignment 1

Part I: In this part, you will implement the gradient-descent algorithm in stochastic learning mode that would train a linear regression model on data that you would generate with the “make_regression” utility of the “scikit-learn” software. You would write a one-page report that would show the convergence graph of the algorithm for different learning rates. For the “make_regression” utility, see the link below. (weightage: 5 points)

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_regression.html#sklearn.datasets.make_regression

Deadline: February 12

Part II: In this part, you will implement the gradient descent algorithm in both stochastic and batch learning modes to train a logistic regression model for the “college placement data” (uploaded to Moodle in the general section). You would compare the performance of the algorithm in both learning modes. In addition, you would see how the model generalizes to the test data. You would summarize your findings in a one-page report. (weightage: 5 points)

Deadline: February 24

Part III: Fit linear and logistic regression models to the “diabetes” and “breast cancer” data in scikit-learn (See the links below for the diabetes and breast cancer data).

Deadline: February 1

Diabetes

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html#sklearn.datasets.load_diabetes

Breast Cancer

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html#sklearn.datasets.load_breast_cancer