CS 422 Assignment 4 Report

Name(s) of the student(s) completing the assignment:

Ivan Biacan

The data preprocessing steps you took (if any).

- Used SimpleImputer to fill NaN fields with the mean of the column

The dataset you used, its source and characteristics:

https://www.kaggle.com/dileep070/heart-disease-prediction-using-logistic-regression

Input Features of the dataset:

- Age, totChol, sysBP, diaBP, BMI, heartrate, glucose

Output Feature of the dataset:

- TenYearCHD (10-year risk of coronary heart disease)

The solution $\boldsymbol{w}$ (parameter vector).

Linear Regression Solution:

- Coefficients/Weights: [[0.06018171 0.00023588 0.01387904 0.00658462 -0.00869778 -0.00507179 0.00788667]]
- Intercept/Bias: [-7.36180474]

The learning rate(s) you used for gradient descent and how many iterations it took for gradient descent to converge.

Learning Rates/Parameters:

- max_iter = 10000
- All other parameters were left default

Iterations

- 95

Relevant evaluation metrics (accuracy, sensitivity, specificity, f1 score, log loss) for the training dataset.

- Accuracy: 0.8495575221238938
- Sensitivity: 0.9958188153310104529616724738676
- Specificity: 0.0423076923076923076923076923076923076923 0769
- F1 Score: 0.07942238267148015

Relevant evaluation metrics (accuracy, sensitivity, specificity, f1 score, log loss) for the test dataset with for both algorithms.

- Accuracy: 0.8537735849056604
- Sensitivity: 0.99447513812154696132596685082873
- Specificity: 0.03225806451612903225806451612903
- F1 Score: 0.06060606060606061