Homework # 2

August 10, 2019

1 Optimization I: Linear Regression

1.1 Single Variable Regression

Analyze the data set "**single_variate_optimization.csv**" and understand the relationship between the target variable Y and the predictor X

(1) Calculate the variance, standard deviation of X and Y

(2) Calculate the covariance, correlation coefficient of X and Y

(3) Using the LinearRegression from sklearn package in python and build a linear regression model $Y \sim X$, what are β_0 and β_1?

(4) Can you replicate β_1 using the calculation from (1) and (2)? How about β_0?

1.2 Multivariate Regression

Analyze the data set "**multi_variate_optimization.csv**" and understand the relationship between the target variable Y and the predictor Xs

(1) Using the LinearRegression from sklearn package in python and build a linear regression model $Y \sim X$, what are the βs?

(2) Replicate the βs using matrix operation. hint: $\hat{\beta} = (X^T X)^{-1} X^T Y$. Make sure you replicate β_0.

2 Optimization II: Logistic Regression

Analyze the data set "**logistic_regression_optimization.csv**" and understand the relationship between the target variable Y and the predictor Xs

(1) Using the LogisticRegression from sklearn package in python and build a linear regression model $Y \sim X$, what are the βs? Hint: make sure that you set the regularization parameter to be large e.g. $C = 1e4$. See the code below for your reference

```
from sklearn.linear_model import LogisticRegression
logisticRegr = LogisticRegression(C=1e4)
```

(2) Replicate the βs using Newton-Raphson method. Hint: initialize βs with some random values e.g. 0.5, update βs using the formula below until converge. What is your convergence condition? How are your βs compare to problem (1)

$$\vec{\beta}^{(m+1)} \leftarrow \vec{\beta}^{(m)} + (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} - \hat{\mathbf{Y}})$$

Here,

$$\mathbf{W} = \begin{bmatrix} \hat{y}^1(1-\hat{y}^1) & & \\ & \ddots & \\ & & \hat{y}^n(1-\hat{y}^n) \end{bmatrix}$$