Homework # 2

October 6, 2020

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) Caculate the βs using the algorithm that we derived in class. 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 logistic 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) Calculate βs using Newton-Raphson method. Hint: initialize βs with some random values e.g. 0.1. Update βs using the formula below until $\vec{\beta}$ converges, i.e. stop updating when the max value of $|\Delta\beta_j|$ of all j is less than $\delta = 1e^{-15}$. How are your βs compare to what you get in question (1)?

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

Here,

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

(3) How many iteration is needed before $\vec{\beta}$ converges?

(4) Change the convergence condition by setting δ to be 0.5 and 0.1. What are the $\vec{\beta}$s, for the two δs that you choose? How many iterations are needed before they converge?