

Assignment 1 - Linear Regression & Logistic Regression

Part A:

The attached dataset **"house_data.csv"** contains 21613 records of house sale prices. It includes homes sold between May 2014 and May 2015.

- 1- Apply Simple Linear regression with gradient descent to predict the **price** based on **sqft_living** (Square footage of the apartments interior living space).

Given the hypothesis function: $Y = C_1 + C_2 X$

Y (target variable) = Price, X (predictor) = sqft_living, C_1 and C_2 are the parameters of the function.

- 2- Apply Multiple Linear regression with gradient descent to predict **price** based on **5 predictors (grade, bathrooms, lat, sqft_living, view)**.

Given the hypothesis function: $Y = C_1 + C_2 X_2 + C_3 X_3 + C_4 X_4 + C_5 X_5 + C_6 X_6$

Y (target variable) = Price, X (predictor) = (grade, bathrooms, lat, sqft_living, view),

C_1, C_2, C_3, C_4, C_5 and C_6 are the parameters of the function.

- a) Implement the gradient descent function to optimize parameters of the function.
- b) Calculate error function to see how the error of the hypothesis function changes with every iteration of gradient descent (hint: you will need to calculate error in every iteration) .

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

MSE = mean squared error
n = number of data points
 Y_i = observed values
 \hat{Y}_i = predicted values

- c) Use optimized hypothesis function to make predictions on new data.
- d) Try different values of learning rate and see how this changes the accuracy of the model.

Part B:

The attached dataset **"heart.csv"** contain 303 records of patients have heart disease or not according to features in it. You are required to build **Logistic Regression** model using **gradient descent** to predict whether patient have heart disease or not (**target**) **based on 4 predictors (trestbps, chol, thalach, oldpeak)**.

- a) Implement the gradient descent function to optimize parameters of the function.
- b) Calculate error function to see how the error of the hypothesis function changes with every iteration of gradient descent(hint: you will need to calculate error in every iteration).
- c) Use optimized hypothesis function to make predictions on new data.
- d) Try different values of learning rate and see how this changes the accuracy of the model.

Important Notes:

- You can only use “pandas”, “numpy” and “matplotlib” libraries. (***Don’t use “sklearn”***)
- The maximum number of students in a team is 3 and the minimum is 2.
- No late submission is allowed.
- Cheating students will take negative grades and no excuses will be accepted.