Assignment 2 - Logistic Regression and SVM

Implement two different classification models

The attached dataset **"heart.csv"** contain 303 records of patients have heart disease or not according to features in it.

Part A – Logistic Regression:

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) Split dataset into training and testing sets.
b) Implement the gradient descent function to optimize parameters of the function.
c) 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 and choose the min one at the end).
d) Use optimized hypothesis function to make predictions on new data.
e) Try different values of learning rate and see how this changes the accuracy of the model.
f) Plot cost function against number of iterations using different learning rate values.

Part B - SVM:

You are required to build **SVM** model to predict whether patient have heart disease or not **(target).**

Cost Function:

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases}$$

Update Weights Equations:
   1- **If Point is correctly classified.**

$$w = w - \alpha \cdot (2\lambda w)$$

   2- **If Point is not correctly classified.**
$$w = w + \alpha \cdot (y_i \cdot x_i - 2\lambda w)$$

a) Split dataset into training and testing sets.
b) Try different set of features and choose the best features.
c) Try different values of learning rate and see how this changes the accuracy of the model.
d) Implement accuracy function (correctly predicted values / test set size).
Note: The final grade is proportional to the accuracy of your results.

Important Notes:

- You can only use "pandas", "numpy" and "matplotlib" libraries. ***(Don't use "sklearn")***
- The maximum number of students in a team is 4 and minimum 3.
- No late submission is allowed.
- Cheating students will take negative grades and no excuses will be accepted.
- **Deadline is 2/1/2021 at 11:59 PM.**