Lab Assignment 4

Machine Learning (UML501)

Q 1	(**Based on Step-by-Step Implementation of Ridge Regression using Gradient Descent Optimization**) Generate a dataset with atleast seven highly correlated columns and a target variable. Implement Ridge Regression using Gradient Descent Optimization. Take different values of learning rate (such as 0.0001,0.001,0.01,0.1,1,10) and regularization parameter (10^{-15},10^{-10},10^{-5},10^{-3},0,1,10,20). Choose the best parameters for which ridge regression cost function is minimum and R2_score is maximum.
Q 2	Load the Hitters dataset from the following link https://drive.google.com/file/d/1qzCKF6JKKMB0p7ul_lLy8tdmRk3vE_bG/view?usp=sharing (a) Pre-process the data (null values, noise, categorical to numerical encoding) (b) Separate input and output features and perform scaling (c) Fit a Linear, Ridge (use regularization parameter as 0.5748), and LASSO (use regularization parameter as 0.5748) regression function on the dataset. (d) Evaluate the performance of each trained model on test set. Which model performs the best and Why?
Q 3	**Cross Validation for Ridge and Lasso Regression** Explore Ridge Cross Validation (RidgeCV) and Lasso Cross Validation (LassoCV) function of Python. Implement both on Boston House Prediction Dataset (load_boston dataset from sklearn.datasets).