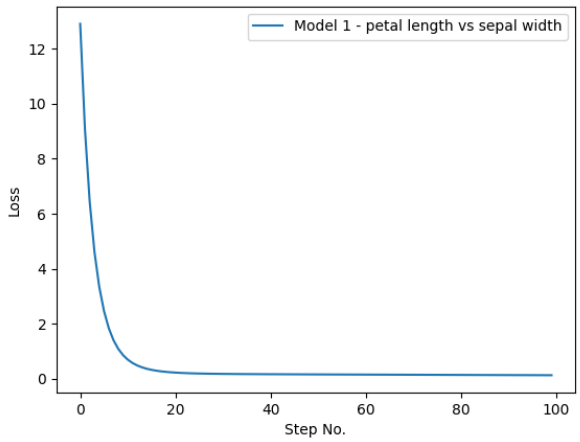
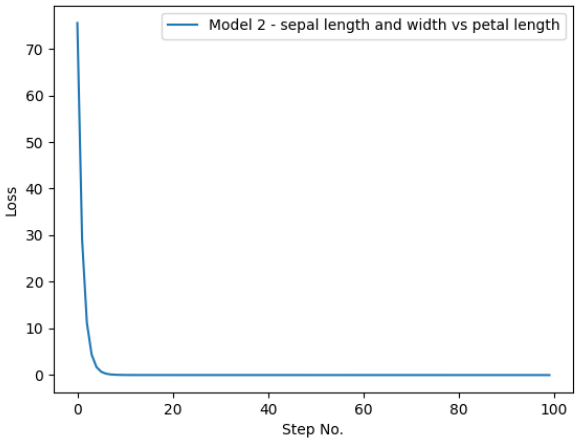
Assignment 1

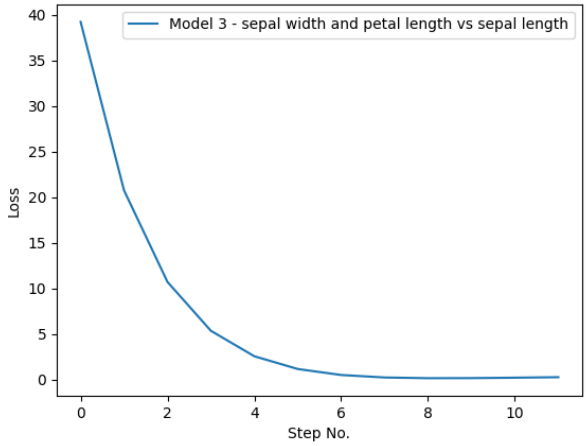
Linear Regression

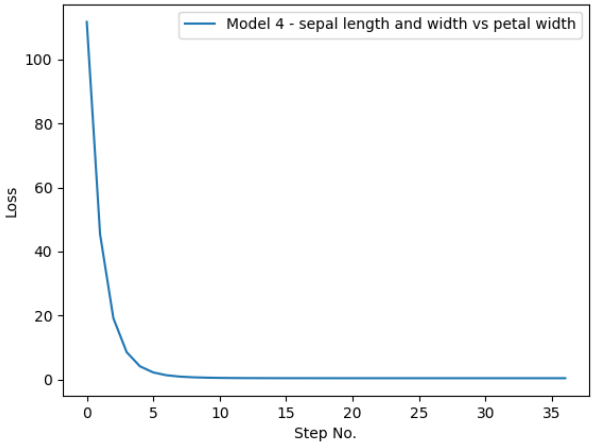
The Iris flower dataset (https://en.wikipedia.org/wiki/Iris_flower_data_set) was organized by Ronald Fisher in 1936. It is a commonly used dataset for introductory machine learning concepts. You will use this dataset for fitting and evaluating your regression model.

The MSE at each step is represented below for 4 models:

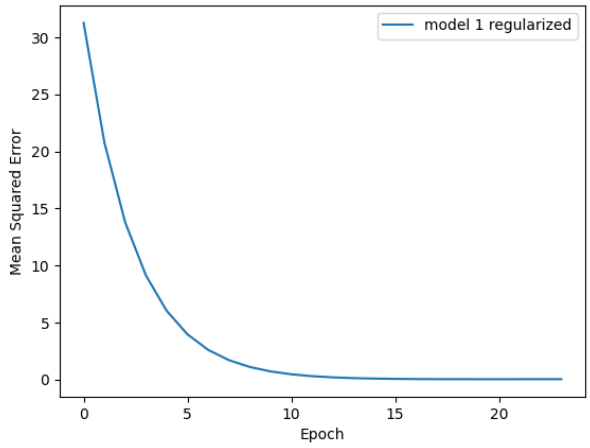


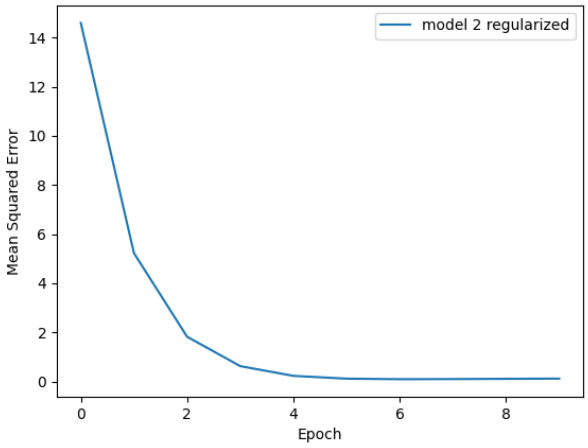


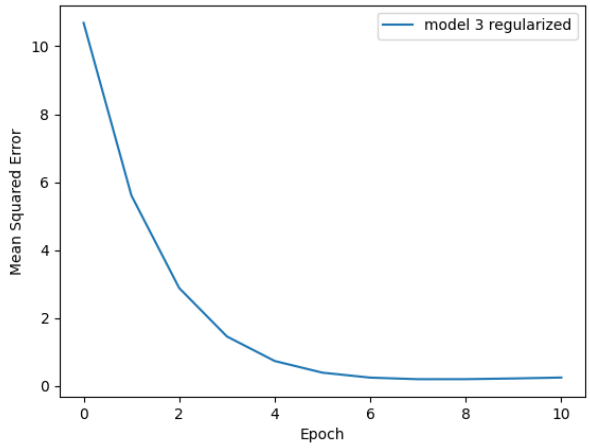


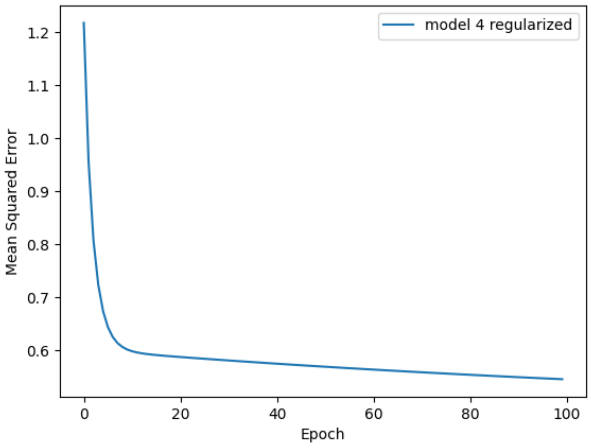


The MSE is shown against the step for same 4 models with L2 regularization:









Below table is showing the difference between the weights and bias for the models trained with and without L2 regularization:

Model Name	Weight difference	Bias difference
Model 1 - petal length vs sepal width	0.30077092059297206	1.3444268277793094
Model 2 - sepal length and width vs sepal width	1.1918409899531262	0.32324543096066827
Model 3 - sepal width and petal length vs sepal length	0.9179324741143361	0.5440236123390891
Model 4 - sepal length and width vs petal width	0.975844990123635	0.5953416324779834

Below table shows the output of evaluation regression scripts for MSE of the test data :

Model Name	MSE of the Test data
Model 1 - petal length vs sepal width	0.131856638655959
Model 2 - sepal length and width vs sepal width	0.008418933330571811
Model 3 - sepal width and petal length vs sepal length	0.1446515524507178
Model 4 - sepal length and width vs petal width	0.44532362319354807

Logistic Regression

Below are the plots of model 1 which has petal length and width as input features to predict if plant is iris virginica, model 2 has sepal length and width as input to predict if plant is iris-setosa.