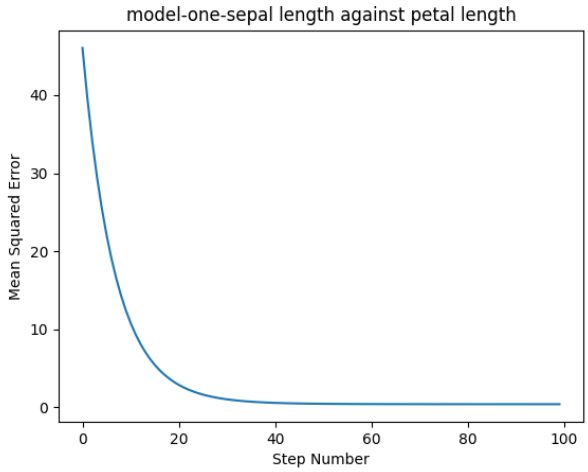
Assignment 1

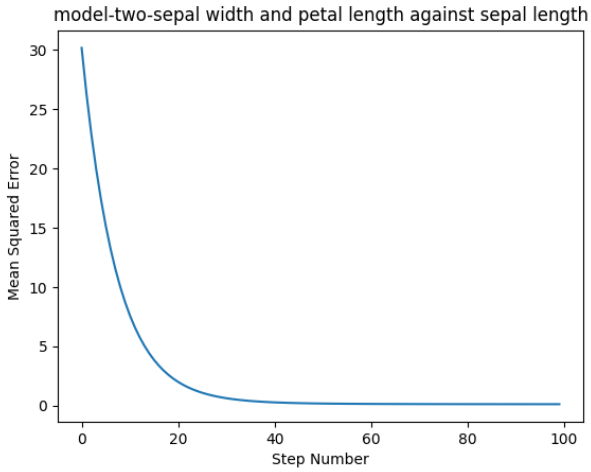
Linear Regression

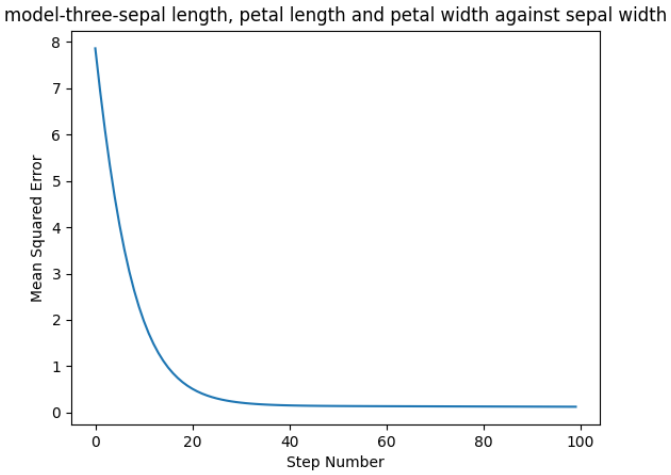
The ***Iris* flower data set** or **Fisher's *Iris* data set** is a multivariate data set used and made famous by the British statistician and biologist Ronald Fisher in his 1936 paper *The use of multiple measurements in taxonomic problems* as an example of linear discriminant analysis.[1] It is sometimes called **Anderson's *Iris* data set** because Edgar Anderson collected the data to quantify the morphologic variation of *Iris* flowers of three related species.[2] Two of the three species were collected in the Gaspé Peninsula "all from the same pasture, and picked on the same day and measured at the same time by the same person with the same apparatus".[3]

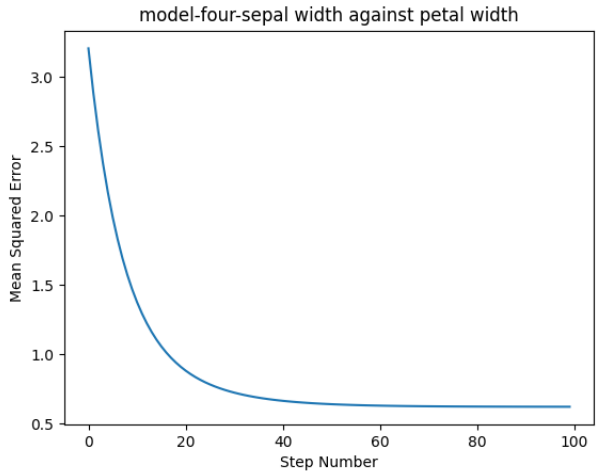
The data set consists of 50 samples from each of three species of *Iris* (*Iris setosa*, *Iris virginica* and *Iris versicolor*). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other. Fisher's paper was published in the *Annals of Eugenics* (today the *Annals of Human Genetics*).[1]

In below figures, we have a plot a graph of loss given by MSE against the step number :

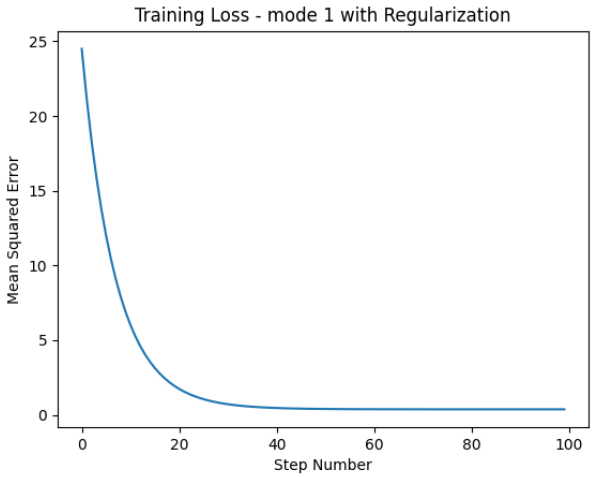


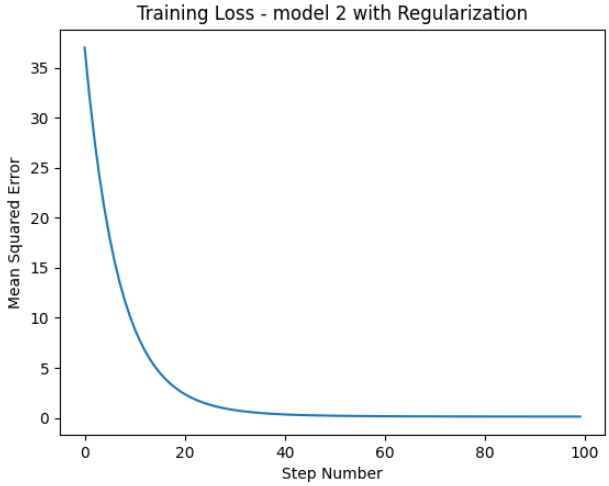


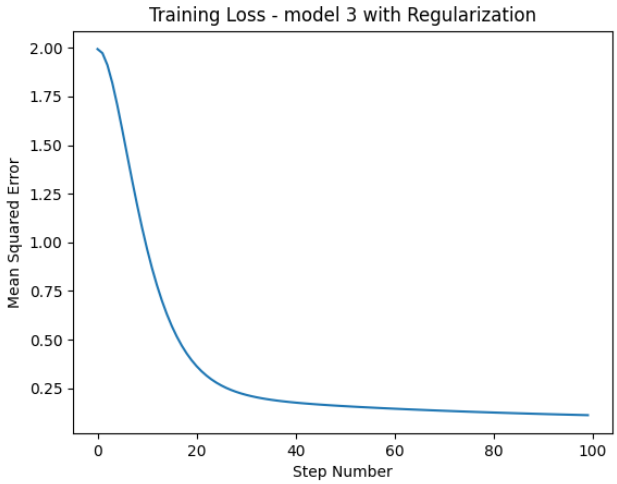


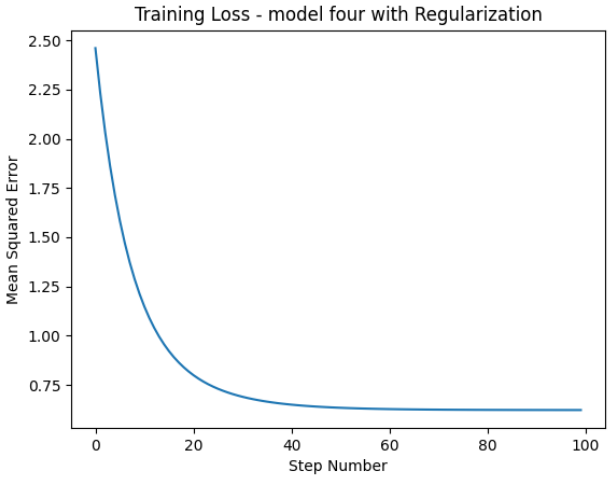


In below figures, the loss as MSE plotted against the step for same 4 models given above with L2 regularization:









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

Model Name	Weight difference	Bias difference
Model one – sepal length against petal length	0.0002901724177241105	0.000642405394899459
Model two - sepal width, petal length against sepal length	0.010604391210132291	0.00038035539204805957
Model three - sepal length, petal length, petal width against sepal width	0.5928149615062459	0.0008307678900418836
Model four - sepal width against petal width	0.0006593598948632096	0.00012424612231543009

Below table shows the evaluation on regressions for MSE of the test data :

Model Name	MSE of the Test data
Model one – sepal length against petal length	5.048444922904025
Model two - sepal width, petal length against sepal length	1.1436039533507383
Model three - sepal length, petal length, petal width against sepal width	0.4180748440715257
Model four - sepal width against petal width	1.3830959502566988

Logistic Regression

In Logistic Regression, model 1 is taking petal length and width as input features to predict if it's iris-setosa. Similarly, model 2 is taking sepal width and length as input features to predict if it's iris-setosa. Model 3 is taking into consideration all features and predicting if it's iris-setosa.

Below, graphs of logistic regression model 1 and 2 is given: