EE708, Assignment 2 Submission

Dhruv Gupta, Roll no: 240354

Problem 1

Question

When the training set is small, the contribution of variance to error may be more than that of bias, and in such a case, we may prefer a simple model even though we know it is too simple for the task. Can you give an example?

Answer

- **Why would we prefer a simple model?**
  Since we have a small training set, there is less data to learn the true pattern that the model is trying to predict. Complex models will be likely to overfit because they will memorize the small training set, hence they will have high variance. A simple model might have some bias, but it is still better because the underfitting won't be as much of an issue as the overfitting when we compare the two.
- **An example of this:**
  Let's consider the example in a later part of this assignment where we are asked to do polynomial fitting using linear regression. It is a simple model, and we don't get the best fit, especially on higher degree curves, but if we were to use a complex model for this task instead (lets say a deep neural network with many layers), since the dataset is just of 100 samples, the model will just fit the data exactly as it is, this leads to very high variance and will not work on any other unknown testing set.
- **Bias-Variance Trade-off:**
  A simple model will have high bias, (underfitting) but low variance, which is our priority right now given the small size of the training set.
  A complex model will have high variance (overfitting), since it will fit the training data exactly but won't be able to fit the testing data.

Problem 2

Question

What is the effect of changing λ on bias and variance?

$$E = \sum_t \left[r^t - g(x^t|w)\right]^2 + \lambda \sum_i w_i^2$$

Answer

- **What does this represent?**
This function E represents the loss function for **Ridge Regression** (or L2 regularised regression)
The first term, $\sum_t \left[r^t - g(x^t|w)\right]^2$, represents the mean squared error (MSE) that measures the difference between the actual values r^t and the predicted values $g(x^t|w)$.
The second term, $\lambda \sum_i w_i^2$, is a regularization penalty that controls the magnitude of the weights w_i. λ is the regularization parameter that determines how much penalty we set on large weigts, ie a higher λ means that the weights will be closer to zero.
- **What happens when λ is increased?**
If λ is increased, the penalty term will dominate, and that will cause the weights given to be small, that flattens the model and can cause underfitting. The model will become too simple and the bias will be high, even though the variance might be low.
- **What happens when λ is decreased?**
When λ is decreased, the penalty term won't have that much of a contribution and higher weights will be allowed. This means that the model will have a tendency to overfit the training data and the variance will be high even though the bias is low.
- **Final Remarks:**
Increasing λ causes underfitting, high bias and low variance.
Decreasing λ causes overfitting, high variance and low bias.

Problem 3

Question

On average, do people gain weight as they age? Based on a dataset of 250 samples, some summary statistics for both age (x) and weight (y) are:

$$\sum_{i=1}^{n} x_i = 11211.00, \quad \sum_{i=1}^{n} y_i = 44520.80, \quad \sum_{i=1}^{n} x_i^2 = 543503.00$$

$$\sum_{i=1}^{n} y_i^2 = 8110405.02, \quad \sum_{i=1}^{n} x_i y_i = 1996904.15$$

Assume that the two variables are related according to the simple linear regression model.

(a) Calculate the least squares estimates of the slope and intercept.

(b) Use the equation of the fitted line to predict the weight that would be observed, on average, for a man who is 25 years old.

(c) Suppose that the observed weight of a 25-year-old man is 170 lbs. Find the residual for that observation.

(d) Was the prediction for the 25-year-old in part (c) an overestimate or underestimate? Explain briefly.

Answer

$$\sum_{i=1}^{n} x_i = 11211.00, \quad \sum_{i=1}^{n} y_i = 44520.80, \quad \sum_{i=1}^{n} x_i^2 = 543503.00$$

$$\sum_{i=1}^{n} y_i^2 = 8110405.02, \quad \sum_{i=1}^{n} x_i y_i = 1996904.15, \quad n = 250$$

Part (a): Calculate the Least Squares Estimates of the Slope and Intercept
The formulas for the slope (β_1) and intercept (β_0) are:

$$\beta_1 = \frac{S_{xy}}{S_{xx}}, \quad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

Where:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}, \quad S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{11211.00}{250} = 44.844, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{44520.80}{250} = 178.0832$$

$$S_{xy} = 1996904.15 - \frac{11211.00 \times 44520.80}{250} = 226.979$$

$$S_{xx} = 543503.00 - \frac{(11211.00)^2}{250} = 40776.116$$

Answer

The slope β_1 is:

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{226.979}{40776.116} = 0.00557$$

The intercept β_0 is:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 178.0832 - (0.00557 \times 44.844) = 177.8333$$

Thus:

$$\boldsymbol{\beta_0 = 177.8333, \quad \beta_1 = 0.00557}$$

Part (b): Use the Equation of the Fitted Line to Predict the Weight of a 25-Year-Old Man
The prediction equation is:

$$\hat{y} = \beta_0 + \beta_1 x$$

For $x = 25$:

$$\hat{y} = 177.8333 + (0.00557 \times 25) = 177.97255$$

Hence, prediction for weight of a 25-year old man is $\boldsymbol{\hat{y} = 177.972}$
Part (c): Find the Residual for an Observed Weight of 170 lbs for a 25-Year-Old

$$\text{Residual} = \text{Observed} - \text{Predicted}$$

$$\text{Residual} = 170 - 177.97255 = \mathbf{-7.97255}$$

Part (d): Was the prediction for the 25-year-old in part (c) an overestimate or underestimate? Explain briefly.
The residual term we found in the previous part turned out to be negative, this means that the model actually predicted the value of the weight (177) greater than what the observed weight was (170), hence, the prediction was clearly an overestimate.

Problem 4

Question

An article in Concrete Research presented 14 data samples on compressive strength (x) and intrinsic permeability (y) of various concrete mixes and cures. Summary quantities are:

$$\sum_{i=1}^{n} y_i = 572, \quad \sum_{i=1}^{n} x_i = 43, \quad \sum_{i=1}^{n} x_i y_i = 1697.8$$

$$\sum_{i=1}^{n} y_i^2 = 23,530, \quad \sum_{i=1}^{n} x_i^2 = 157.42$$

Assume that the two variables are related according to the simple linear regression model.

(a) Calculate the least squares estimates of the slope and intercept. Estimate σ^2

(b) Use the equation of the fitted line to predict what permeability would be observed when the compressive strength is $(x) = 4.3$.

(c) Give a point estimate of the mean permeability when compressive strength is $(x) = 3.7$.

(d) Suppose that the observed value of permeability at $(x) = 3.7$ is $(y) = 46.1$. Calculate the value of the corresponding residual.

Answer

We have:

$$\sum_{i=1}^{n} y_i = 572, \quad \sum_{i=1}^{n} x_i = 43, \quad \sum_{i=1}^{n} x_i y_i = 1697.8$$

$$\sum_{i=1}^{n} y_i^2 = 23,530, \quad \sum_{i=1}^{n} x_i^2 = 157.42, \quad n = 14$$

Part (a): Calculate the Least Squares Estimates of the Slope and Intercept
The formulas for the slope (β_1) and intercept (β_0) are:

$$\beta_1 = \frac{S_{xy}}{S_{xx}}, \quad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

Where:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}, \quad S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

Proceeding in a similar manner as we did in the previous problem, we get the results:

$$S_{xy} = -59.0571, \quad S_{xx} = 25.34857$$

Answer

The slope β_1 is:

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{-59.0571}{25.34857} = -2.3298$$

The intercept β_0 is:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 40.8571 - (-2.3298 \times 3.0714) = 48.01284772$$

Thus:

$$\boldsymbol{\beta_0 = 48.01284772, \quad \beta_1 = -2.3298}$$

Estimating the variance σ^2

The estimated variance $\hat{\sigma}^2$ is calculated using the formula:

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(\sum_{i=1}^{n} y_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i y_i\right)$$

Substitute the given values into the equation:

$$\hat{\sigma}^2 = \frac{1}{12}(23,530 - 48.68 \times 572 - (-2.59) \times 1697.8)$$

First, calculate each term:

$$48.68 \times 572 = 27,852.16, \quad -2.59 \times 1697.8 = -4,394.38$$

Now, substitute these into the equation for $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{12}(23,530 - 27,852.16 + 4,394.38)$$

$$\hat{\sigma}^2 = \frac{1}{12}(23,530 - 23,457.78)$$

$$\hat{\sigma}^2 = \frac{1}{12} \times 72.22 \approx 6.02$$

Thus, **the estimate of σ^2 is approximately** 6.02.
Part (b): Use the equation of the fitted line to predict what permeability would be observed when the compressive strength is $x = 4.3$
The prediction equation is:

$$\hat{y} = \beta_0 + \beta_1 x$$

For $x = 4.3$:

$$\hat{y} = 48.01284772 + (-2.3298 \times 4.3) = 37.9947$$

Hence, prediction for observed permeability when compressive strength is 4.3 is:

$$\boldsymbol{\hat{y} = 37.9947}$$

Answer

Part (c): Give a point estimate of the mean permeability when compressive strength is $x = 3.7$

The prediction equation is:

$$\hat{y} = \beta_0 + \beta_1 x$$

For $x = 3.7$:

$$\hat{y} = 48.01284772 + (-2.3298 \times 3.7) = 39.39259$$

Hence, prediction for mean permeability when compressive strength is 3.7 is:

$$\boldsymbol{\hat{y} = 39.39259}$$

Part (d): Suppose that the observed value of permeability at $(x) = 3.7$ is $(y) = 46.1$. Calculate the value of the corresponding residual.

$$\text{Residual} = \text{Observed} - \text{Predicted}$$

$$\text{Residual} = 46.1 - 39.39259 = \mathbf{6.70741}$$

Clearly the residual is positive, hence we predicted a value less than the observed value, so our model underestimated the value of permeability.

Problem 5

Question

A study was performed to investigate the shear strength of soil y as it relates to depth in feet x_1 and % moisture content x_2. Ten observations were collected, and the following summary quantities were obtained:

$$\sum x_{i1} = 223, \quad \sum x_{i2} = 553, \quad \sum y_i = 1916$$

$$\sum x_{i1}^2 = 5200.9, \quad \sum x_{i2}^2 = 31729, \quad \sum y_i^2 = 371595.6$$

$$\sum x_{i1} y_i = 43550.8, \quad \sum x_{i2} y_i = 104736.8, \quad \sum x_{i1} x_{i2} = 12352$$

(a) Set up the least squares normal equations for the model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

(b) Estimate the parameters in the model in part (a).

(c) What is the predicted strength when $x_1 = 18$ feet and $x_2 = 43\%$?

Answer

Part (a): Setting up the least squares normal equations
The normal equations for multiple regression are given by:

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}$$

Substituting given values:

$$\begin{bmatrix} 10 & 223 & 553 \\ 223 & 5200.9 & 12352 \\ 553 & 12352 & 31729 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1916 \\ 43550.8 \\ 104736.8 \end{bmatrix}$$

Part (b): Estimating the parameters
The normal equation system can be written as:

$$AB = C$$

where:

$$A = \begin{bmatrix} 10 & 223 & 553 \\ 223 & 5200.9 & 12352 \\ 553 & 12352 & 31729 \end{bmatrix}, B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, C = \begin{bmatrix} 1916 \\ 43550.8 \\ 104736.8 \end{bmatrix}$$

To solve for B, we compute:

$$B = A^{-1}C$$

First, find the inverse of A, denoted as A^{-1}, and multiply it by C. The computed values are:

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 4.063 \\ 1.218 \\ 2.743 \end{bmatrix}$$

Thus, the estimated regression equation is:

$$Y = 4.063 + 1.218x_1 + 2.743x_2$$

Part (c): Predicting the shear strength when $x_1 = 18$ and $x_2 = 43$
Using the estimated equation:

$$\hat{y} = 4.063 + 1.218(18) + 2.743(43)$$

$$\hat{y} = 4.063 + 21.924 + 118.949 = 144.936$$

Thus, the predicted shear strength is **144.936**.

Problem 6

Question

A regression model is described between the percent body fat (%BF) measured by immersion and BMI from a study on 250 male subjects. The researchers also measured 13 physical characteristics of each man, including his age (yrs), height (in), and waist size (in). Write out the regression model of the percent of body fat with both height and waist as predictors with the given information:

$$(X'X)^{-1} = \begin{bmatrix} 2.9705 & -4.0042E-2 & -4.1679E-2 \\ -0.4004 & 6.0774E-4 & -7.3875E-5 \\ -0.00417 & -7.3875E-5 & 2.5766E-4 \end{bmatrix}, \quad (X'y) = \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

Answer

We use the normal equation for multiple linear regression:

$$\mathbf{w} = (X'X)^{-1}(X'y)$$

Substituting the given values:

$$\mathbf{w} = \begin{bmatrix} 2.9705 & -4.0042E-2 & -4.1679E-2 \\ -0.4004 & 6.0774E-4 & -7.3875E-5 \\ -0.00417 & -7.3875E-5 & 2.5766E-4 \end{bmatrix} \cdot \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

This yields the estimated regression coefficients:

$$w_0 = -2737.53$$
$$w_1 = -1715.51$$
$$w_2 = 1.76$$

Thus, the estimated weight matrix is:

$$\mathbf{w} = \begin{bmatrix} -2737.53 \\ -1715.51 \\ 1.76 \end{bmatrix}$$

The regression equation is:

$$\hat{y} = -2737.53 + (-1715.51 \cdot x_1) + (1.76 \cdot x_2)$$

where x_1 represents height and x_2 represents waist size.

Problem 7

Question

Let us say we have two variables x_1 and x_2 and we want to make a quadratic fit using them, namely

$$f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2 + w_3x_1x_2 + w_4x_1^2 + w_5x_2^2$$

Derive the least square estimates of w_i, $i = 0, 1, \ldots, 5$, given N data samples.

Answer

The given quadratic model is:

$$f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2 + w_3x_1x_2 + w_4x_1^2 + w_5x_2^2$$

We have to fit x_1 and x_2, but effectively we have to fit a polynomial $f(x_1, x_2)$, so we can treat it as performing linear regression on 5 features, as shown in the polynomial, namely: $x_1, x_2, x_1\dot{x}_2, x_1^2, x_2^2$. This means we can construct our design matrix as follows:

For N data samples, the design matrix X is given by:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} & x_{11}^2 & x_{12}^2 \\ 1 & x_{21} & x_{22} & x_{21}x_{22} & x_{21}^2 & x_{22}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & x_{N1}x_{N2} & x_{N1}^2 & x_{N2}^2 \end{bmatrix}$$

where each row corresponds to a data point, and each column represents a basis function in the model.

Formulating the Normal Equation

The least squares estimate of the parameter vector w is given by:

$$w = (X^TX)^{-1}X^Ty$$

where: - X^TX is a 6×6 matrix, - X^Ty is a 6×1 vector, - y is the vector of target values.

Estimating w

After computing the inverse of X^TX and performing the necessary matrix multiplications, we obtain the estimated values: $w_0, w_1, w_2, w_3, w_4, w_5$
which together form the parameter vector:

$$w = [w_0 \quad w_1 \quad w_2 \quad w_3 \quad w_4 \quad w_5]^T$$

Thus, we have derived the least squares estimates for the quadratic regression model.

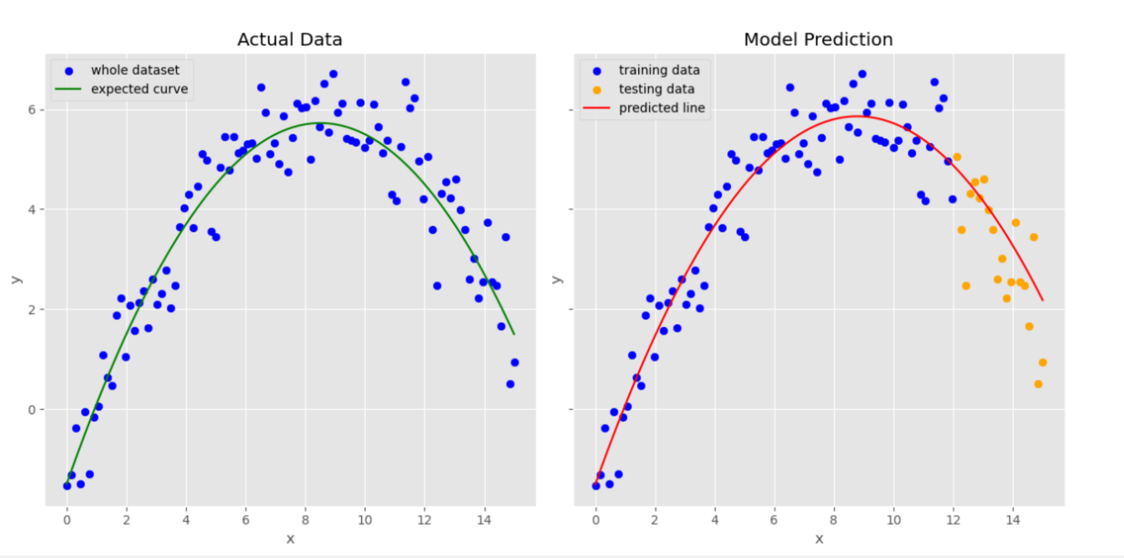
Problem 8

Question

Assume a linear model and add 0-mean Gaussian noise to generate 100 samples.
a. Divide your sample into training and testing sets (80:20).
b. Use linear regression for the training half. Compute the mean squared error (MSE) on the testing set.
c. Plot the fitted model along with the data.
d. Repeat the same for polynomials of degrees 2 and 3 as well

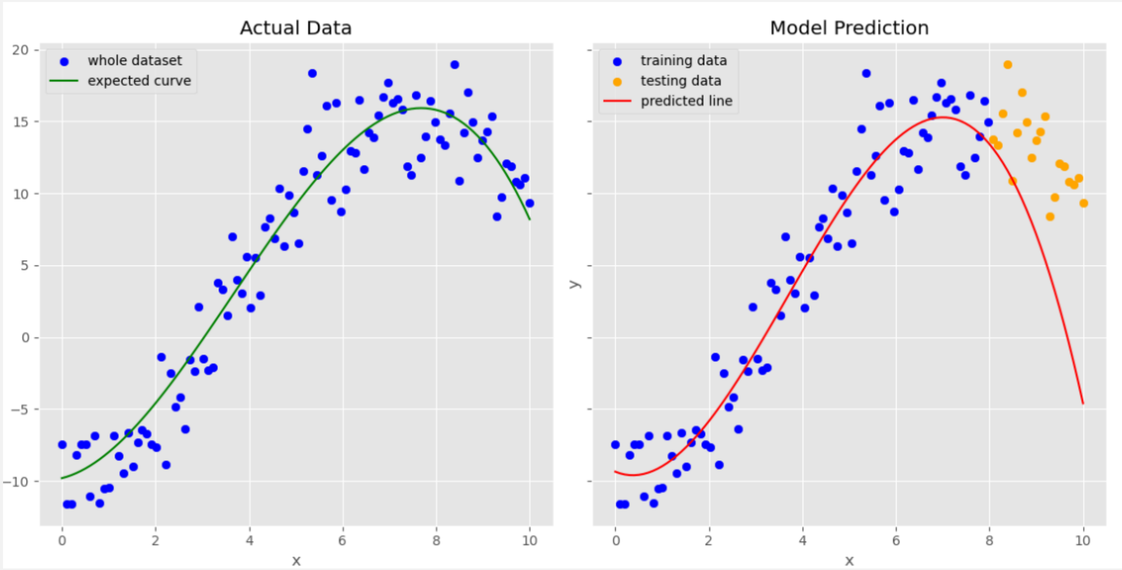
Output

Simple Linear Regression (degree 1)
Mean Squared Error on the testing set is: 8.228128879606325



Polynomial Linear Regression (degree 2)
Mean Squared Error of the model for testing data is: 1417.882487932521

Output



Polynomial Linear Regression (degree 3)
Mean Squared Error for the model during testing is: 67.7496760314958

Problem 9

Question

Implement logistic regression using dataset A2-P2.csv.
Write a code for gradient descent with learning rates of 0.01 and 0.05. For each learning rate:
a. Plot variation of mean squared error for 20 iterations.
b. Specify the final weight value.

Output

To implement logistic regression and gradient descent in this dataset, since there are only 200 values in the dataset, we can use **Batch Gradient Descent**, since BGD will give us better accuracy and won't be that hard to do since theres only 200 samples.

There are two methods to perform logistic regression in this case:

- Minimize the mean squared error function and perform gradient descent to reach the best possible weights.
- Maximize the log likelihood function and perform gradient ascent to reach the best possible weights.

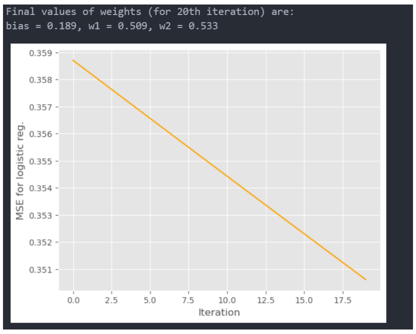
I will be using both these methods and then comparing the results:

Output

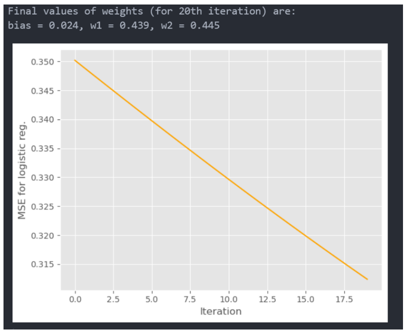
Minimizing the mean squared error function for gradient descent in logistic regression

$$\text{MSE}(w) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(w^T x_i)\right)^2$$

$$\nabla_w \text{MSE}(w) = -\frac{2}{n}\sum_{i=1}^{n}\left(y_i - \sigma(w^T x_i)\right)\cdot \sigma(w^T x_i)\cdot (1-\sigma(w^T x_i))\cdot x_i$$



BGD with $\alpha = 0.01$



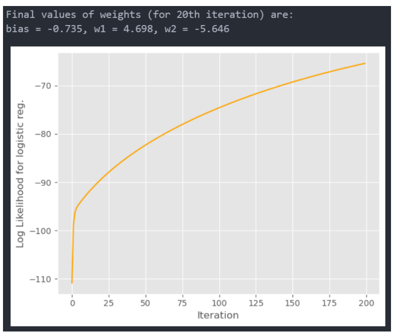
BGD with $\alpha = 0.05$

Output

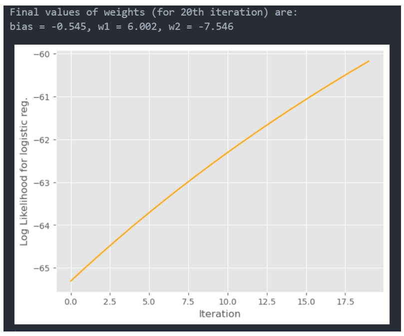
Maximizing the log likelihood function for gradient ascent during logistic regression

$$\log L(w) = l(w) = \sum_{i=1}^{n}\left[y_i \log\sigma(w^T x_i) + (1-y_i)\log(1-\sigma(w^T x_i))\right]$$

$$\nabla_w \text{l}(w) = \sum_{i=1}^{n}\left(y_i - \sigma(w^T x_i)\right)x_i$$



BGA with $\alpha = 0.01$



BGA with $\alpha = 0.05$

Problem 10

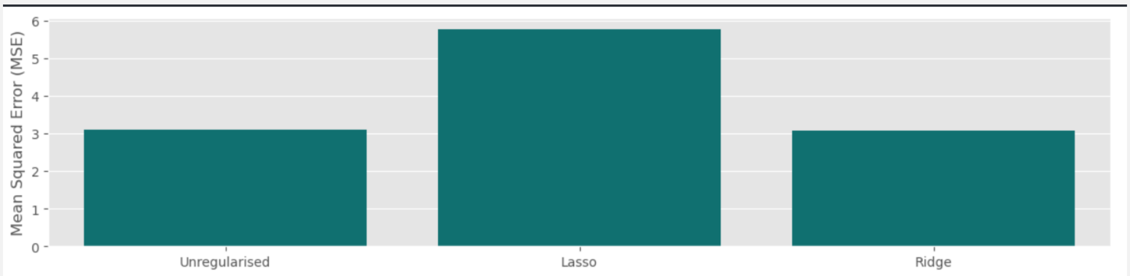
Question

Write a code to implement regression models using dataset A2-P3.csv. Divide the dataset into training and testing sets (80:20). Implement the following models using the training dataset and compute MSE on the test dataset:

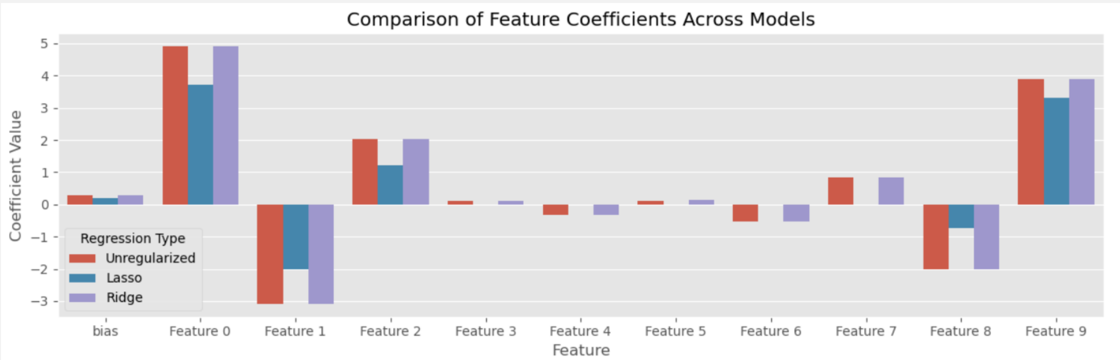
a. Linear regression.
b. Linear regression with LASSO regularization $\left(\frac{\lambda}{2} = 1\right)$
c. Linear regression with ridge regularization $\left(\frac{\lambda}{2} = 0.1\right)$

Use bar plots to compare MSE and feature coefficients (weights) for the three methods.

Output



Comparison between MSE in unregularised, lasso and ridge regressions



Comparison of feature coefficients (weights) for the three types of regression