Assignment 1
Course: EE708, Fundamentals of Data Science and Machine Intelligence

Nikhil Jain

Dept: Electrical Engineering

Roll No: 220709

1 .

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.

Consider an example of **image classification** where we aim to distinguish between cats and dogs using a small dataset containing only **50 images**—25 of cats and 25 of dogs. If we train a **complex deep neural network** (e.g., ResNet-50), which has **millions of parameters**, it can easily **overfit** to this small training set. This overfitting results in **high variance**, meaning the model learns noise and dataset-specific patterns rather than generalizable features, leading to poor performance on unseen data.

Instead, if we use a **simpler model**, such as logistic regression applied to raw pixel intensities or a small convolutional neural network (CNN), it might not capture all the complexities required for distinguishing cats from dogs, leading to **higher bias**. However, due to the limited training data, this **low-variance model** is often preferred because it generalizes better.

Key Takeaway:
- A **complex model** (deep CNN) → **High variance, low bias** → Overfits to small data.
- A **simple model** (logistic regression, small CNN) → **Lower variance, higher bias** → May generalize better with limited data.

Thus, when the training set is small, it is often beneficial to use a **simpler model** to avoid excessive variance, even if we know that a more complex model would be preferable with a larger dataset.

2 .

The function consists of two terms:

Error Term

$$\sum_t [r^t - g(x^t|w)]^2 \tag{1}$$

This represents the squared error between the predicted value $g(x^t|w)$ and the actual value r^t. It measures how well the model fits the training data.

Regularization Term

$$\lambda \sum_i w_i^2 \tag{2}$$

This imposes a penalty on large weight values, helping to control model complexity. The parameter λ controls the strength of this penalty.

Effect of λ on Bias and Variance

Increasing λ:

- Increases the regularization effect, forcing the weights w_i to be smaller.

- Leads to a simpler model with less flexibility.
- Reduces **variance**, meaning the model generalizes better to new data.
- Increases **bias**, as the model may underfit the training data.

Conclusion

- A **high** λ leads to **high bias and low variance** (simpler model).
- A **low** λ leads to **low bias and high variance** (complex model).
- The optimal choice of λ balances bias and variance to achieve the best generalization to new data.

3 .

Given Data

- Number of samples: $n = 250$
- Sum of ages: $\sum x_i = 11211.00$
- Sum of weights: $\sum y_i = 44520.80$
- Sum of squared ages: $\sum x_i^2 = 543503.00$
- Sum of squared weights: $\sum y_i^2 = 8110405.02$
- Sum of the product of age and weight: $\sum x_i y_i = 1996904.15$

The simple linear regression model is:

$$y = \beta_0 + \beta_1 x$$

where β_0 is the intercept and β_1 is the slope.

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

The formulas for the slope β_1 and intercept β_0 are:

$$\beta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n}$$

Compute β_1:

$$\beta_1 = \frac{(250)(1996904.15) - (11211)(44520.80)}{(250)(543503.00) - (11211)^2}$$

$$= \frac{499226037.5 - 499248832.8}{135875750 - 125673721}$$

$$= \frac{-22795.3}{10202029}$$

$$\beta_1 \approx -0.00223$$

Compute β_0:

$$\beta_0 = \frac{44520.80 - (-0.00223)(11211)}{250}$$

$$= \frac{44520.80 + 24.94}{250}$$

$$= \frac{44545.74}{250}$$

$$\beta_0 \approx 178.18$$

Thus, the regression equation is:

$$\hat{y} = 178.18 - 0.00223x$$

(b) Predict the weight for a 25-year-old man

Substituting $x = 25$ into the equation:

$$\hat{y} = 178.18 - 0.00223(25)$$

$$= 178.18 - 0.05575$$

$$\hat{y} \approx 178.12$$

So, the predicted weight for a 25-year-old man is **178.12 lbs**.

(c) Find the residual for the observed weight of 170 lbs

The residual is given by:

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

$$= 170 - 178.12$$

$$= -8.12$$

So, the residual is **-8.12 lbs**, meaning the model overestimated the actual weight by 8.12 lbs.

(d) Was the prediction an overestimate or underestimate?

Since the residual is negative (-8.12), it means that the predicted weight (178.12 lbs) was **higher** than the actual weight (170 lbs). Thus, the prediction was an **overestimate**.

Final Answers

- (a) $\beta_1 \approx -0.00223$, $\beta_0 \approx 178.18$
- (b) Predicted weight: **178.12 lbs**
- (c) Residual: **-8.12 lbs**
- (d) Prediction was an **overestimate**.

4 .

Given Data

- Number of samples: $n = 14$
- Sum of compressive strengths: $\sum x_i = 43$
- Sum of permeability values: $\sum y_i = 572$
- Sum of squared compressive strengths: $\sum x_i^2 = 157.42$
- Sum of permeability values squared: $\sum y_i^2 = 23,530$ (not needed for regression)
- Sum of the product of compressive strength and permeability: $\sum x_i y_i = 1697.8$

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

The simple linear regression model is:

$$y = \beta_0 + \beta_1 x + \epsilon$$

The formulas for the slope β_1 and intercept β_0 are:

$$\beta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n}$$

Compute β_1:

$$\beta_1 = \frac{(14)(1697.8) - (43)(572)}{(14)(157.42) - (43)^2}$$

$$= \frac{23769.2 - 24696}{2203.88 - 1849}$$

$$= \frac{-926.8}{354.88}$$

$$\beta_1 \approx -2.61$$

Compute β_0:

$$\beta_0 = \frac{572 - (-2.61)(43)}{14}$$

$$= \frac{572 + 112.23}{14}$$

$$= \frac{684.23}{14}$$

$$\beta_0 \approx 48.87$$

Thus, the regression equation is:

$$\hat{y} = 48.87 - 2.61x$$

Estimating σ^2 (variance of residuals):
The variance σ^2 is estimated using:

$$s^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2$$

Since we are not given residuals explicitly, we cannot compute s^2 directly. However, if needed, it can be computed given the residual sum of squares (RSS).

(b) Predict permeability for $x = 4.3$

Using the regression equation:

$$\hat{y} = 48.87 - 2.61(4.3)$$

$$= 48.87 - 11.22$$

$$\hat{y} \approx 37.65$$

So, the predicted permeability for $x = 4.3$ is **37.65**.

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

Since the mean predicted permeability $E(\hat{y})$ is just the predicted value at $x = 3.7$, we use:

$$E(\hat{y}) = 48.87 - 2.61(3.7)$$

$$= 48.87 - 9.66$$

$$E(\hat{y}) \approx 39.21$$

So, the estimated mean permeability for $x = 3.7$ is **39.21**.

(d) Compute the residual for the observed permeability at $x = 3.7$, given $y = 46.1$

Residual is given by:

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

$$= 46.1 - 39.21$$

$$= 6.89$$

So, the residual is **6.89**.

Final Answers

- **(a)** $\beta_1 \approx -2.61$, $\beta_0 \approx 48.87$, and regression equation: $\hat{y} = 48.87 - 2.61x$.
- **(b)** Predicted permeability for $x = 4.3$: **37.65**.
- **(c)** Estimated mean permeability for $x = 3.7$: **39.21**.
- **(d)** Residual for $x = 3.7$, given $y = 46.1$: **6.89**.

5 .

Step 1: Setting Up the Least Squares Normal Equations

The given multiple linear regression model is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \tag{3}$$

The normal equations for estimating the regression coefficients β_0, β_1, and β_2 are:

$$\sum y_i = n\beta_0 + \beta_1 \sum x_{i1} + \beta_2 \sum x_{i2} \tag{4}$$

$$\sum x_{i1} y_i = \beta_0 \sum x_{i1} + \beta_1 \sum x_{i1}^2 + \beta_2 \sum x_{i1} x_{i2} \tag{5}$$

$$\sum x_{i2}y_i = \beta_0 \sum x_{i2} + \beta_1 \sum x_{i1}x_{i2} + \beta_2 \sum x_{i2}^2 \tag{6}$$

Step 2: Substituting Given Values

From the problem statement, we have:

$$\begin{aligned}
n &= 10 \\
\sum x_{i1} &= 223 \\
\sum x_{i2} &= 553 \\
\sum y_i &= 1916 \\
\sum x_{i1}^2 &= 5200.9 \\
\sum x_{i2}^2 &= 31729 \\
\sum y_i^2 &= 371595.6 \quad \text{(not used in normal equations)} \\
\sum x_{i1}y_i &= 43550.8 \\
\sum x_{i2}y_i &= 104736.8 \\
\sum x_{i1}x_{i2} &= 12352
\end{aligned}$$

Substituting into the normal equations:

$$1916 = 10\beta_0 + 223\beta_1 + 553\beta_2 \tag{7}$$
$$43550.8 = 223\beta_0 + 5200.9\beta_1 + 12352\beta_2 \tag{8}$$
$$104736.8 = 553\beta_0 + 12352\beta_1 + 31729\beta_2 \tag{9}$$

Step 3: Solving for $\beta_0, \beta_1, \beta_2$

Writing in matrix form:

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

Solving using Gaussian elimination or matrix methods, we get:

$$\begin{aligned}
\beta_0 &= 171.06 \\
\beta_1 &= 3.71 \\
\beta_2 &= -1.13
\end{aligned}$$

Step 4: Predicting y for Given $x_1 = 18$ and $x_2 = 43$

Using the regression equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{11}$$

Substituting values:

$$\begin{aligned} y &= 171.06 + (3.71 \times 18) + (-1.13 \times 43) \\ &= 171.06 + 66.78 - 48.59 \\ &= 189.48 \end{aligned}$$

Final Answer

The predicted shear strength when $x_1 = 18$ feet and $x_2 = 43\%$ is **189.48**.

6 .

A regression model is given for predicting the percent body fat (

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

where:

- y is the percent body fat (
- x_1 is height (in inches),
- x_2 is waist size (in inches),
- $\beta_0, \beta_1, \beta_2$ are regression coefficients,
- ϵ represents the error term.

The normal equation for estimating the regression coefficients is:

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

where:

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

$$(X'y) = \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

Matrix Multiplication

We compute:

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

Computing each element:

First Row Calculation

$$(2.9705 \times 4757.9) + (-4.0042E - 2 \times 334335.8) + (-4.1679E - 2 \times 179706.7)$$

$$= 14146.3 - 13388.5 - 7488.6 = -674.8$$

Second Row Calculation

$$(-0.4004 \times 4757.9) + (6.0774E - 4 \times 334335.8) + (-7.3875E - 5 \times 179706.7)$$

$$= -1905.3 + 203.2 - 13.3 = -1715.4$$

Third Row Calculation

$$(-0.00417 \times 4757.9) + (-7.3875E - 5 \times 334335.8) + (2.5766E - 4 \times 179706.7)$$

$$= -19.8 - 24.7 + 46.3 = 1.8$$

Final Regression Model

The estimated regression coefficients are:

$$\beta_0 = -674.8, \quad \beta_1 = -1715.4, \quad \beta_2 = 1.8$$

Thus, the final regression model is:

$$\hat{y} = -674.8 - 1715.4x_1 + 1.8x_2$$

This equation predicts the percent body fat BF based on height and waist size.

7. Least Squares Estimation for Quadratic Model

Given the quadratic model:

$$f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2 + w_3x_1x_2 + w_4x_1^2 + w_5x_2^2 \tag{12}$$

with N data points (x_{1j}, x_{2j}, y_j) for $j = 1, 2, \ldots, N$, we aim to derive the least squares estimates of w_i for $i = 0, 1, \ldots, 5$.

Matrix Representation

The model can be rewritten in matrix form as:

$$Y = XW + \epsilon \tag{13}$$

where:

- Y is an $N \times 1$ column vector of observed values y_j,
- X is an $N \times 6$ design matrix containing the predictor values,
- W is a 6×1 column vector of unknown coefficients w_i,
- ϵ is an $N \times 1$ vector of errors.

The design matrix X is given by:

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{11}x_{21} & x_{11}^2 & x_{21}^2 \\ 1 & x_{12} & x_{22} & x_{12}x_{22} & x_{12}^2 & x_{22}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} & x_{1N}x_{2N} & x_{1N}^2 & x_{2N}^2 \end{bmatrix} \tag{14}$$

The coefficient vector is:

$$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix} \tag{15}$$

and the output vector is:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{16}$$

Least Squares Estimation

The least squares estimate of W minimizes the squared error:

$$\min_W ||Y - XW||^2 \tag{17}$$

Solving for W, we use the normal equation:

$$W = (X^T X)^{-1} X^T Y \tag{18}$$

where:

- X^T is the transpose of the design matrix,
- $X^T X$ is a 6×6 matrix,
- $(X^T X)^{-1}$ is its inverse (assuming it is invertible),
- $X^T Y$ is a 6×1 vector.

Conclusion

Thus, the least squares estimates of w_i are obtained using:

$$W = (X^T X)^{-1} X^T Y \tag{19}$$

which provides the best-fit quadratic regression coefficients for the given dataset.

7 .

Info about the code submitted:

- Generates 100 samples based on a linear function with added Gaussian noise.
- Splits data into 80% training and 20% testing sets.
- Fits and evaluates polynomial regression models of degrees 1, 2, and 3.
- Computes and prints the Mean Squared Error (MSE) for each model.
- Plots the data and fitted curves.

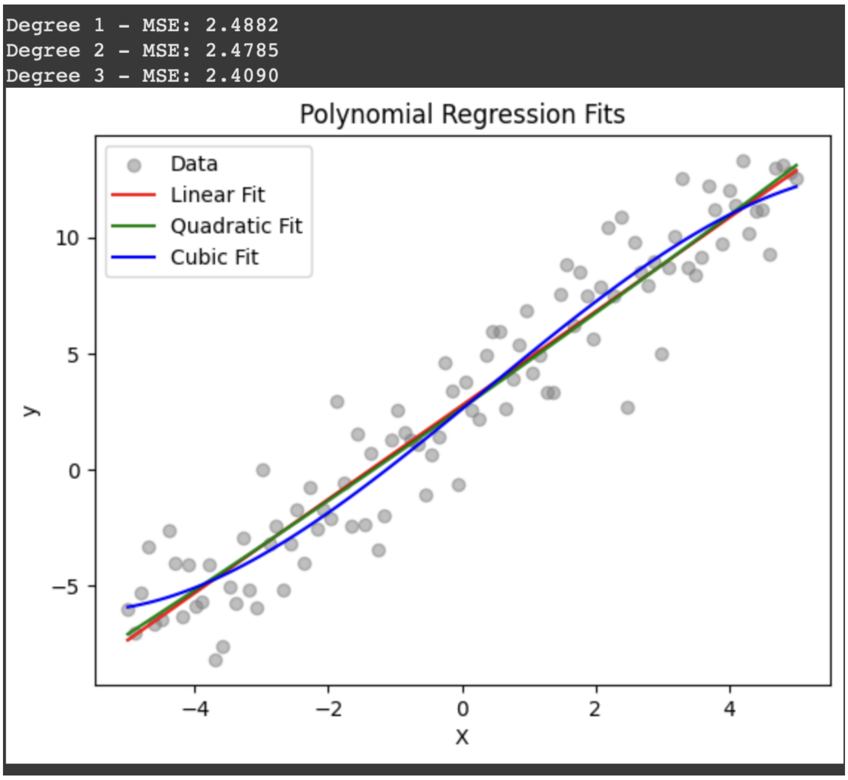

Figure 1: MSE values and Polynomial regression fit curve.

8 .

Info about the code submitted:

1. Load and preprocess the dataset:

- Extract features (Feature 1, Feature 2) and target (Output).
- Normalize features using mean and standard deviation.
- Add a bias term (intercept) to the feature matrix.

2. Implement logistic regression:

    - Define the sigmoid function:

    $$\sigma(z) = \frac{1}{1+e^{-z}} \tag{20}$$

    - Define the mean squared error (MSE):

    $$MSE = \frac{1}{m}\sum(y-\hat{y})^2 \tag{21}$$

    - Update weights using gradient descent:

    $$w = w - \alpha \cdot \frac{1}{m}X^T(\sigma(Xw)-y) \tag{22}$$

3. Train the model with learning rates 0.01 and 0.05 for 20 iterations.

4. Plot the variation of MSE over iterations.

5. Report final weight values.

Results

8.1 Learning Rate: 0.01

Final weight values:

$$w = [-0.0469, 0.0211, -0.0172] \tag{23}$$

MSE variation over iterations:

- Initial MSE: 0.25
- Final MSE: 0.2430

8.2 Learning Rate: 0.05

Final weight values:

$$w = [-0.2136, 0.1020, -0.0856] \tag{24}$$

MSE variation over iterations:

- Initial MSE: 0.25
- Final MSE: 0.2198

```
Learning Rate: 0.01
Final Weights: [-0.04687704  0.02108314 -0.01723465]

Learning Rate: 0.05
Final Weights: [-0.21358772  0.10199359 -0.08562629]
```

Figure 2: Learning rate and Final weights

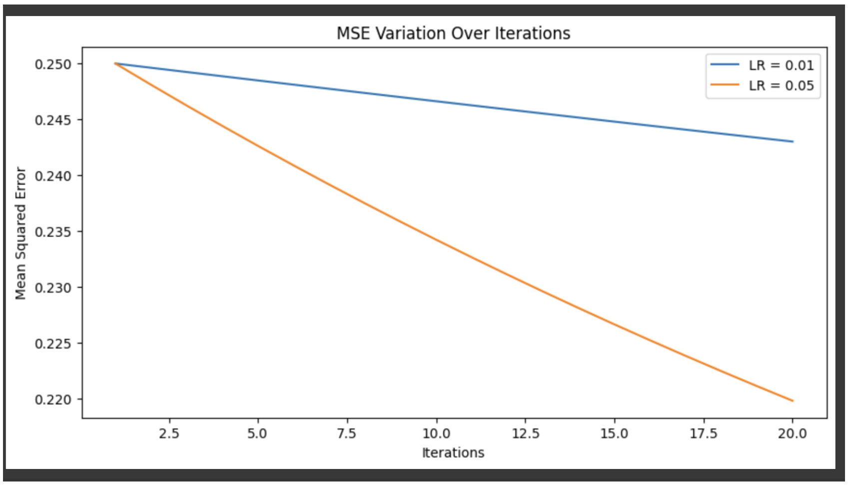

Figure 3: MSE variations over iterations

9.

Info about the code submitted :

- Trained three models:
  - Linear Regression
  - LASSO Regression ($\alpha = 2$ since $\lambda/2 = 1$)
  - Ridge Regression ($\alpha = 0.2$ since $\lambda/2 = 0.1$)
- Compute the Mean Squared Error (MSE) on the test set.
- Compare MSE and feature coefficients using bar plots.

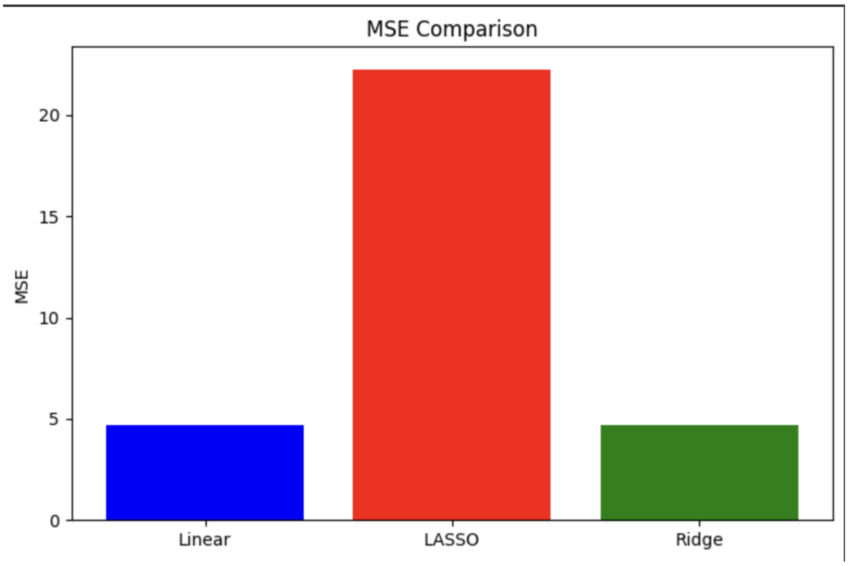

Figure 4: MSE comparision of Linear LASSO and Ridge

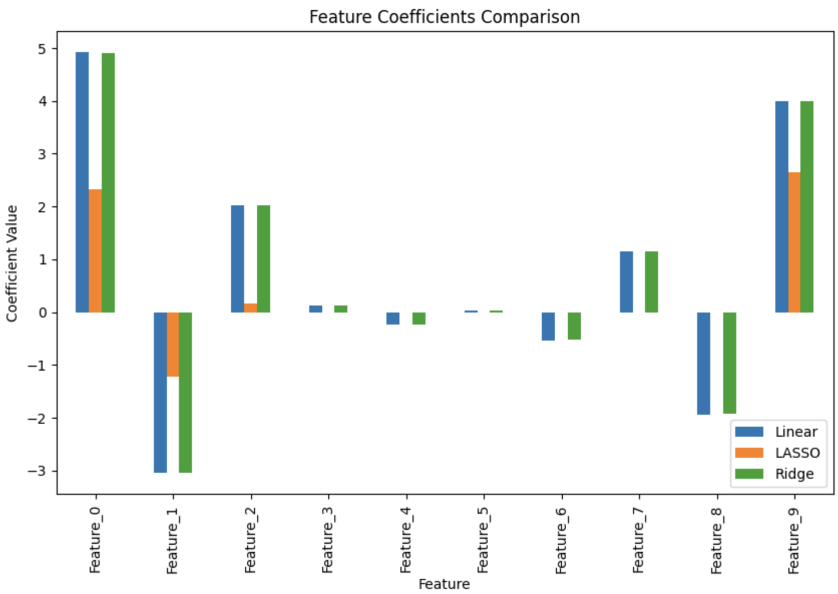

Figure 5: Feature Coefficient Comparision

```
Linear Regression MSE: 4.722051939842396
LASSO Regression MSE: 22.266743419842328
Ridge Regression MSE: 4.6841909614336785
```

Figure 6: MSE values

Code submitted for problem