Report

ML Assignment 1

Shubham Lohan
2019275
CSAM

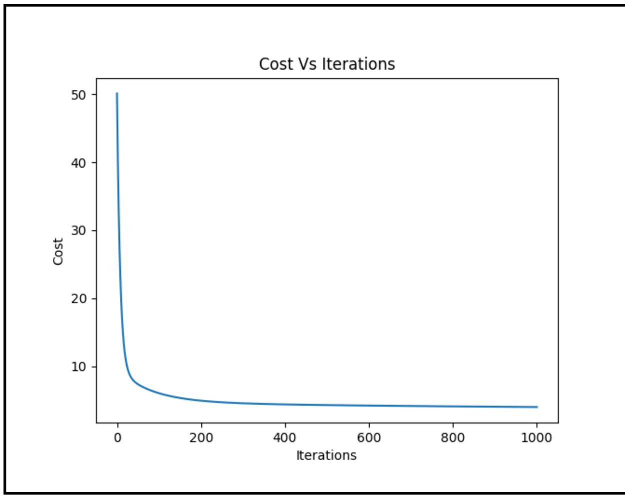
ML Assignment 1 **1**
Question 1: 1
Question 2: 2
Question 3: 5
Question 4: 6

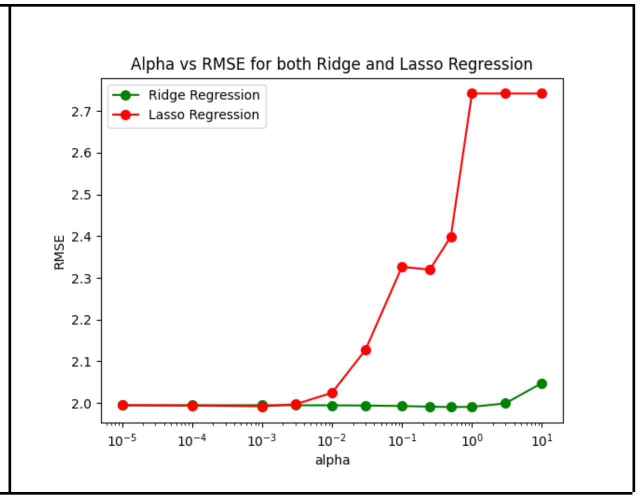
Question 1:

Pre-Processing strategy:

- Rows Shuffled
- Male is mapped to 1, Female is mapped to 2, and infant is mapped to 3
- Dataset is divided into 8:2 train: test with random seed to 0.

1.





2.

(a) Using Sklearn's Ridge and Lasso Implementation

```
RMSE at Training Dataset: 2.822772
RMSE at Testing Dataset: 2.432824
Ridge best alpha: 0.5
Ridge best coef: [4.30637521 - 0.38834862   1.24072869   9.51981219   8.16569944
                  7.90379554 - 18.90245419 - 8.07487403  10.46316726]
Lasso best alpha: 0.001
Lasso best coef: [4.26650567 - 0.38202805   0.           11.235931     8.32045844
                  8.5550278 - 19.78870552 - 8.65524114   9.78886033]
```

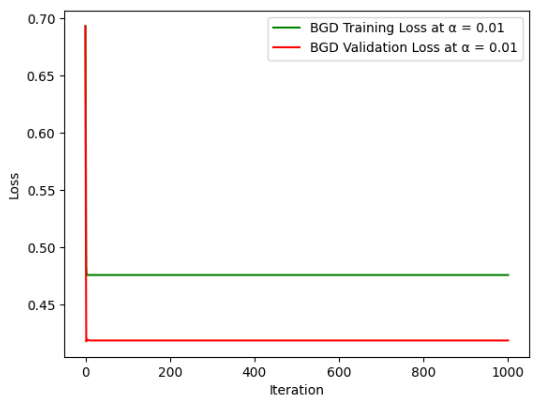
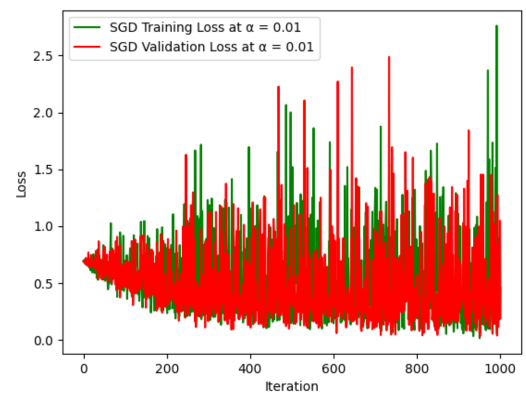
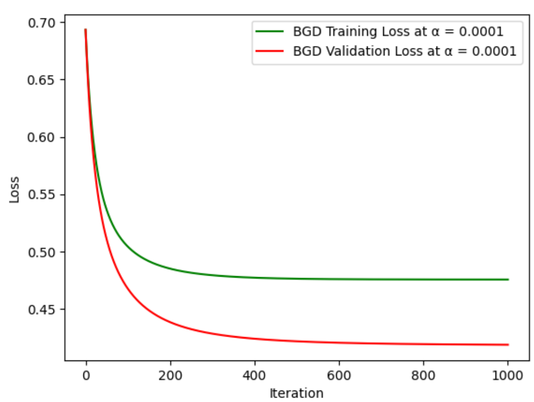
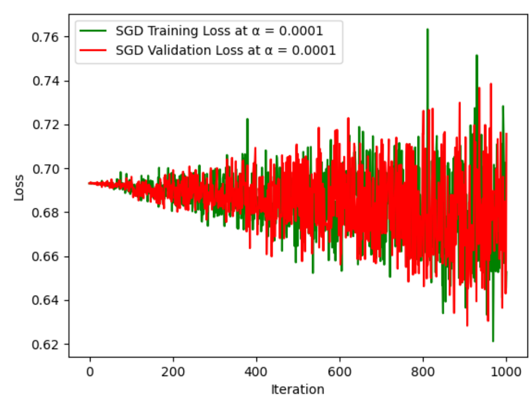
(b) Sklearn's Grid search function

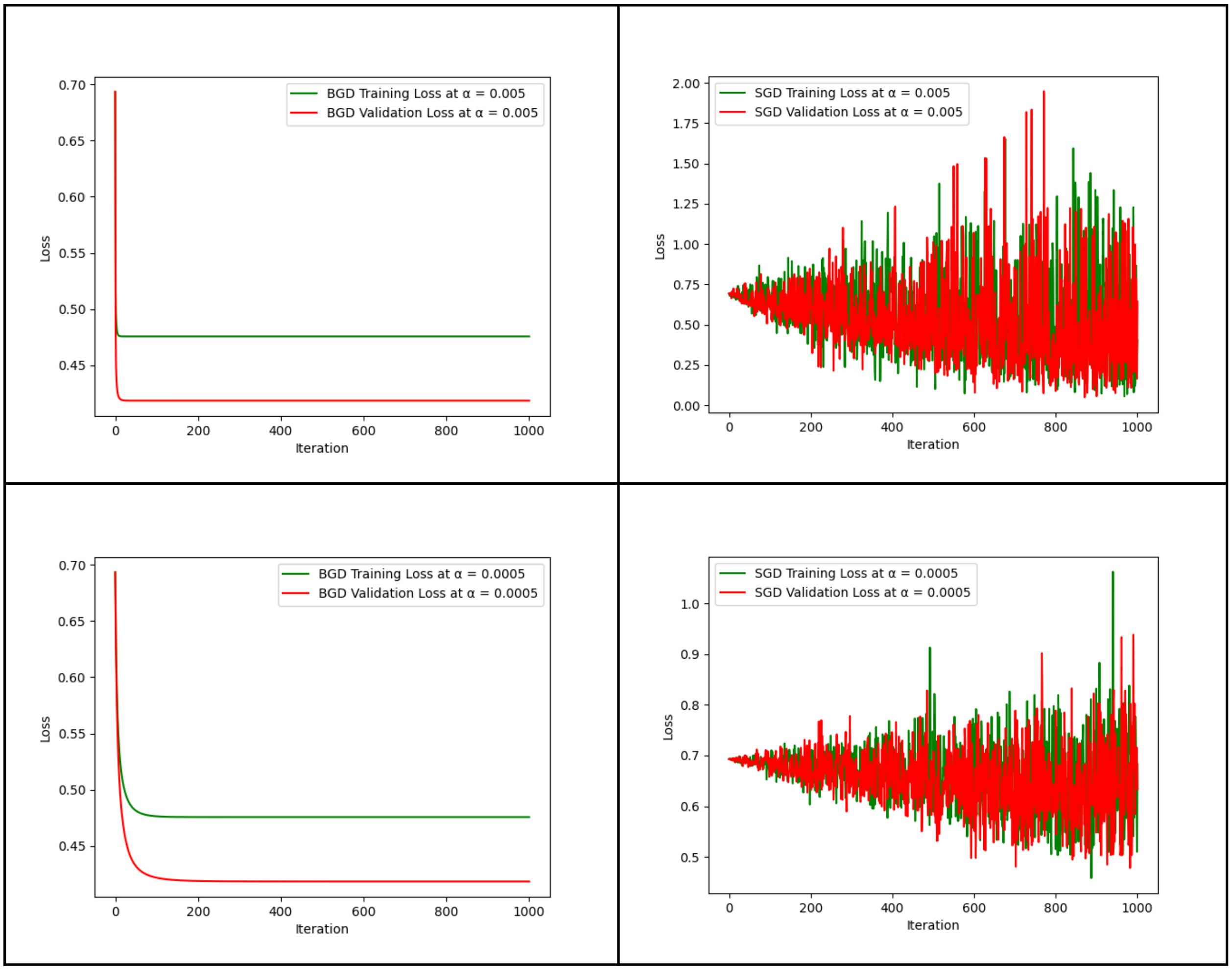
```
Ridge best alpha: 1
Ridge best coef: [4.42465431 - 0.40405838   2.43139224   7.58801632   8.26251436
                  7.14700751 - 17.82383629 - 6.83075109  10.43930842]
Lasso best alpha: 0.003
Lasso best coef: [4.80528132 - 0.4000195    0.            9.76733603   6.6716346
                  6.7179772 - 17.76907111 - 4.35490265  10.88861322]
```

Question 2:

Pre-Processing strategy:

- Rows Shuffled
- Absurd values (like Zero) is replaced with the Median of that category values
- Normalize the data by subtracting mean to each value and dividing with the standard deviation of each column of data
- Dataset is divided into 7:2:1 train:val:test with random seed to 0.

<table>
<tr><th>Training and Validation Loss
Vs
Iteration Loss
(For BGD)</th><th>Training and Validation Loss
Vs
Iteration Loss
(For SGD)</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>



1.b

We can observe that for a smaller value of learning rate the cost function takes time to converge and for a higher learning rate our cost function overshoots.
The learning rate is directly proportional to speed the minimization of working of the cost function.(but sometime overshoots there we should picks the best learning rate by trying the different value)

1.c

```
Using BGD
Confusion Matrix[[43  3]
                 [12 19]]
Accuracy: 0.8051948051948052
Precision: 0.8636363636363636
Recall: 0.6129032258064516
f1: 0.7169811320754716
Using SGD
Confusion Matrix[[42  4]
                 [12 19]]
```

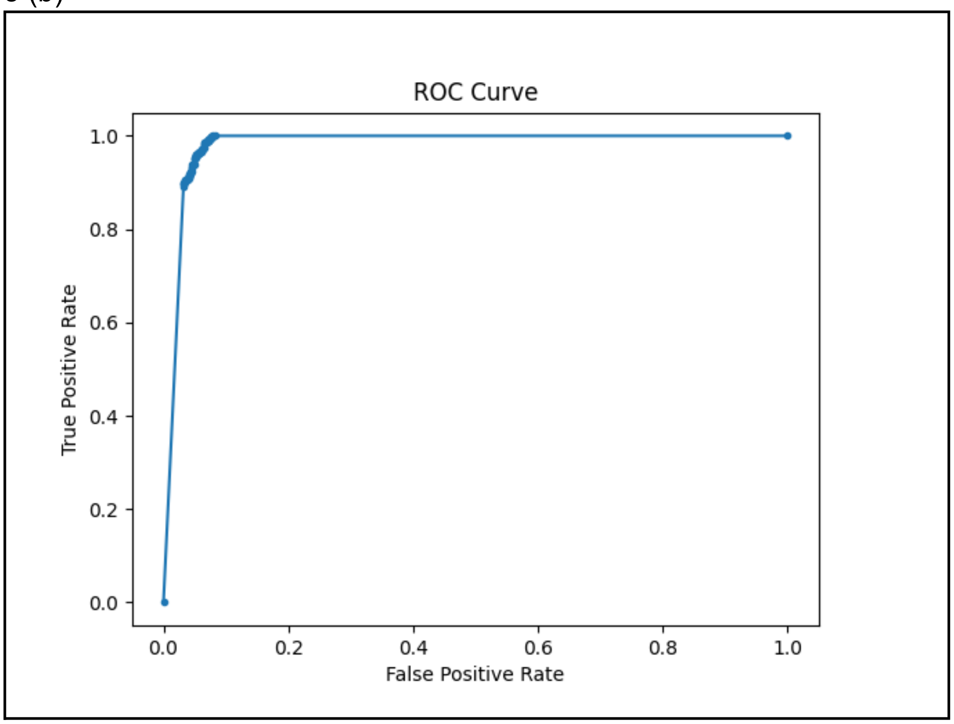
```
Accuracy: 0.7922077922077922
Precision: 0.8260869565217391
Recall: 0.6129032258064516
f1: 0.7037037037037037
Using Sklearn's Logistic Regression
Confusion Matrix[[43  3]
                [12 19]]
Accuracy: 0.8051948051948052
Precision: 0.8636363636363636
Recall: 0.6129032258064516
f1: 0.7169811320754716
```

Question 3:

Pre-Processing strategy:

- Extractions of Required Dataset for Trouser and Pullover
- Binarization of the extracted dataset

3 (b)



3 c

```
Using Scratch code:
Accuracy: 0.9295
Sklearn's Implementation:
0.93
Precision 0.8996282527881041
Recall 0.968
Confusion Matrix:  [[968  32]
                   [108 892]]
```

Question 4:

1.

a. Adding new parameter after segregating the dataset.and run the model for both dataset if the final outcome for the both model give different result then it is suspicious.
b. Adding new parameter after segregating the dataset.and run the model for both dataset if the final outcome for the both model give different result then it is suspicious.
c. Adding new parameter after segregating the dataset.and run the model for both dataset if the final outcome for the both model give different result then it is suspicious.

   Else in our model we can add a new feature X_2, making equation as $W = B_0 + x_1 * B_1 + x_2 * B_2 + u$. Now coefficients can be compared for insights.

2. Because of the internal implementation of the cost function in L2 regularization, we get smaller model coefficients.
3.

Date

(d) $x = a_1c_1 + a_2c_2 + \cdots + a_nc_n$

$\Rightarrow x =$

$\alpha \rightarrow$

$$\beta_{\ell_2} = \arg\min_{\beta}\left(\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{i,p})\right)^2 + \lambda \sum_{j=0}^{p} |\beta_j|^2\right)$$

Likelyhood, $L(\beta|y) := P(y|\beta)$

$$= \prod_{i=1}^{n} P_y(y_i|\beta, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{i,p}))^2}{2\sigma^2}}$$

Max. Likelyhood, $\beta = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}))^2$

$$= \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Posterior

$$P(\theta|y) = \frac{P(y|\theta)\, P(\theta)}{P(y)}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

Spiral

Teacher's Sign