

Ques1.

b) Using sklearn linear regression fit()

Fold	Training MSE (Calculated)	Training MSE (sklearn)	Validation MSE (Calculated)	Validation MSE (sklearn)
1	5.042	5.042	4.339	4.339
2	4.835	4.835	5.179	5.179
3	4.95	4.95	4.705	4.705
4	4.646	4.646	6.358	6.358
5	4.956	4.956	4.708	4.708
Mean	4.886	4.886	5.058	5.058

c) Using Normal Equations method

Fold	Training MSE (Calculated)	Training MSE (sklearn)	Validation MSE (Calculated)	Validation MSE (sklearn)
1	5.042	5.042	4.339	4.339
2	4.835	4.835	5.179	5.179
3	4.95	4.95	4.705	4.705
4	4.646	4.646	6.358	6.358
5	4.956	4.956	4.708	4.708
Mean	4.886	4.886	5.058	5.058

d) Using sklearn linear regression predictions()

Fold	Training MSE (Calculated)	Training MSE (sklearn)	Validation MSE (Calculated)	Validation MSE (sklearn)
1	5.042	5.042	4.339	4.339
2	4.835	4.835	5.179	5.179

3	4.95	4.95	4.705	4.705
4	4.646	4.646	6.358	6.358
5	4.956	4.956	4.708	4.708
Mean	4.886	4.886	5.058	5.058

Analysis of Results

There is no deviation in the calculated MSE and sklearn MSE values in any of the experiment results above.

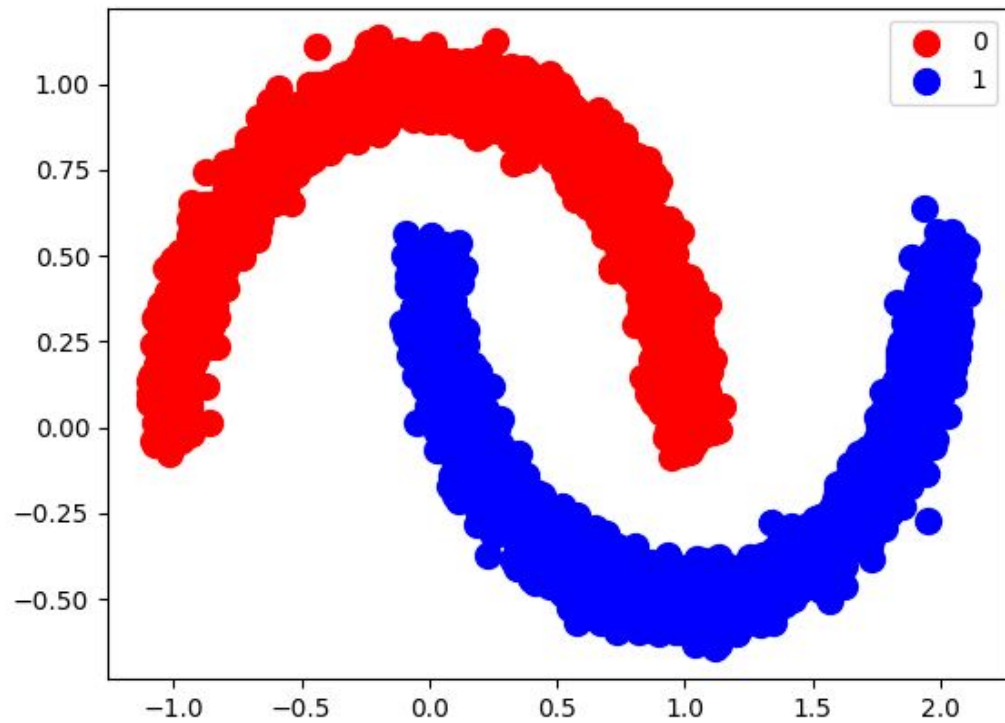
The training and validation MSE of the following three approaches are reported same:

- A) sklearn logistic regression fit() function used and predict() function implemented
- B) Calculating MSE using normal equations method
- C) Sklearn logistic regression fit() and predict() function used

The reason for the equal MSE in all three approaches is that sklearn Linear Regression model also uses Normal Equations approach to implement the regression internally.

Ques2.

a)



Logistic Regression Results

	LogRegression (Without Regularization)		LogRegression (With L2 Regularization)		Sklearn Logistic Regression	
Fold	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy
1	88.9	88.3	89.2	88.6	89.1	88.6
2	88.775	88.3	88.925	89.1	89.05	89.0
3	88.75	89.0	88.85	89.1	88.875	89.2
4	88.4	90.5	88.675	90.6	88.7	90.9
5	89.075	87.7	89.4	87.8	89.475	87.8
Mean Accuracy	88.78	88.76	89.01	89.04	89.04	89.1

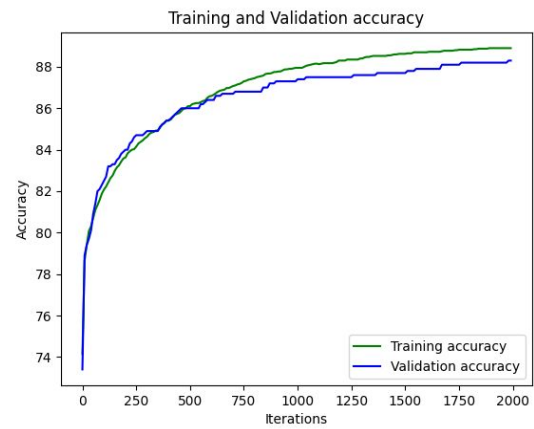
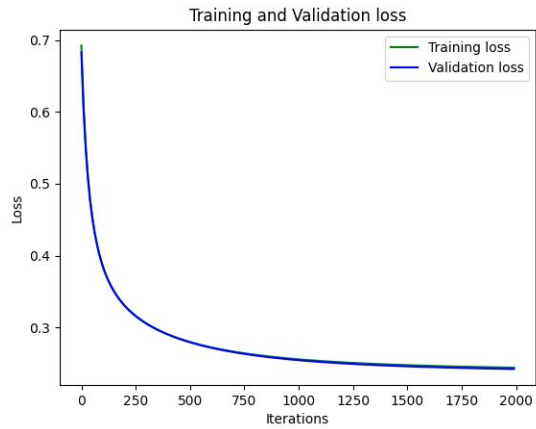
There is a difference of approx 0.5% in the training as well as validation accuracy of Logistic Regression without Regularization and Logistic Regression with L2 regularization. The training and validation accuracy of Logistic Regression with l2 regularization is very close to the training and validation accuracy of sklearn logistic regression implementation.

The optimal value of lambda after grid search came out be 0.01.

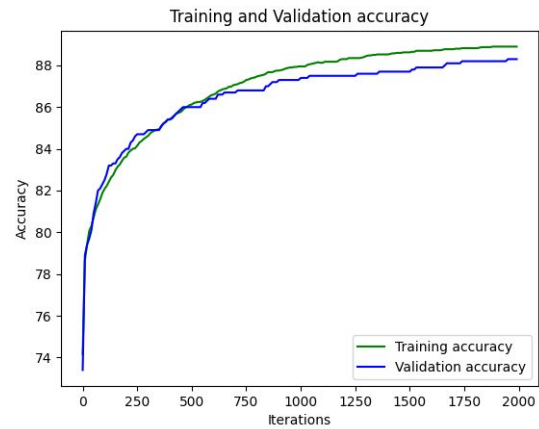
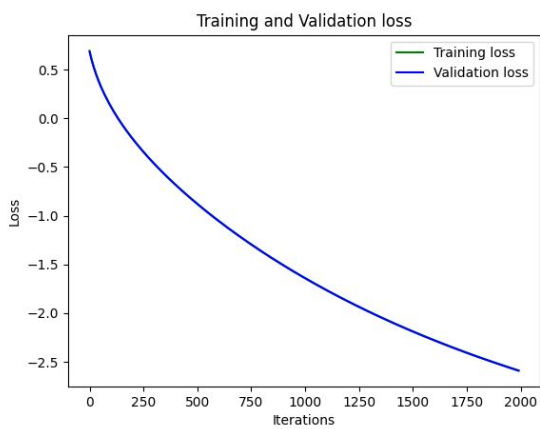
Since, the given dataset has only two features and 4000 samples in the training set, the model does not have high complexity therefore there is no chance of overfitting that's why there is no significant improvement in the performance of the model after applying L2 regularization.

Fold 1

Without Regularization

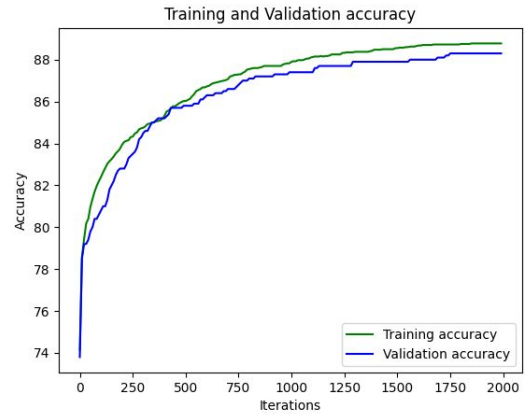
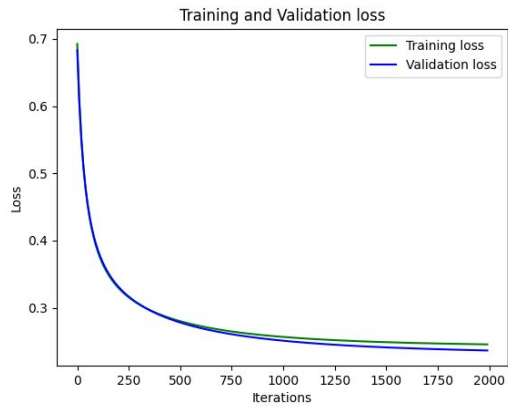


With Regularization

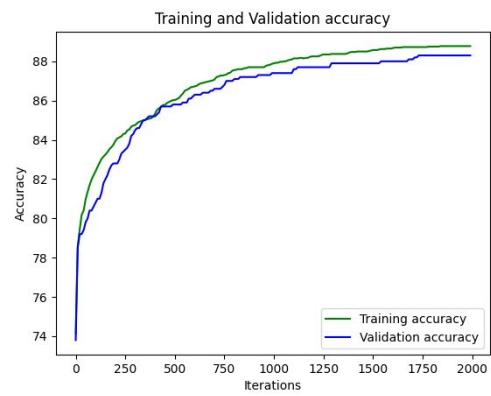
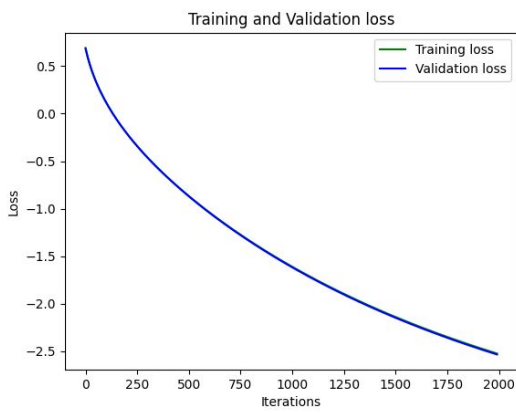


Fold 2

Without Regularization

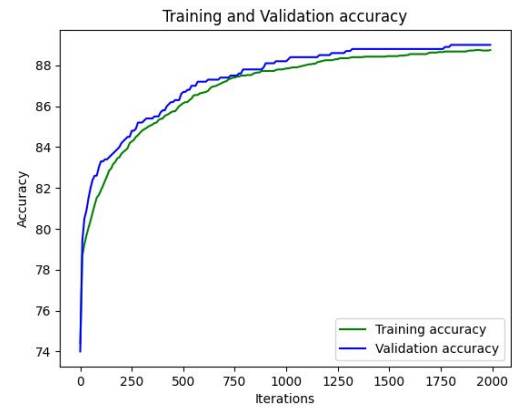
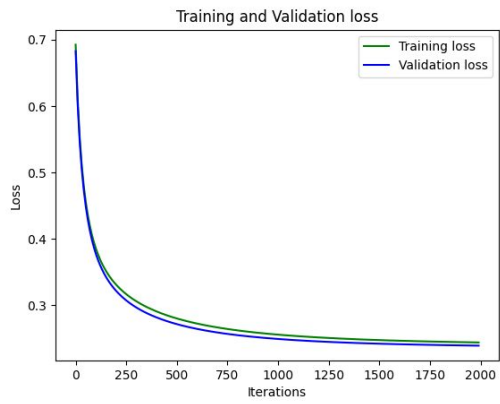


With Regularization

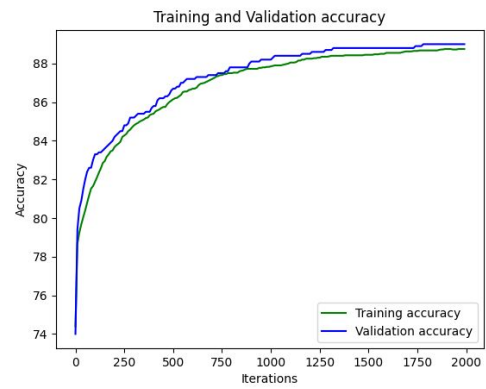
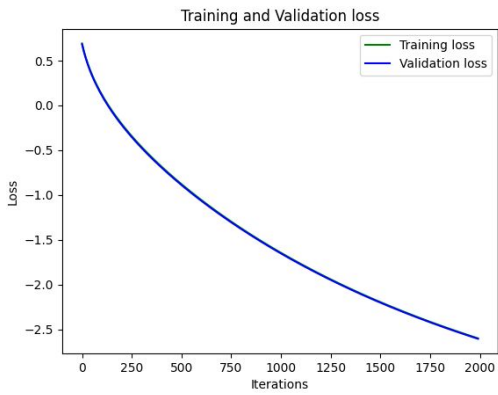


Fold 3

Without Regularization

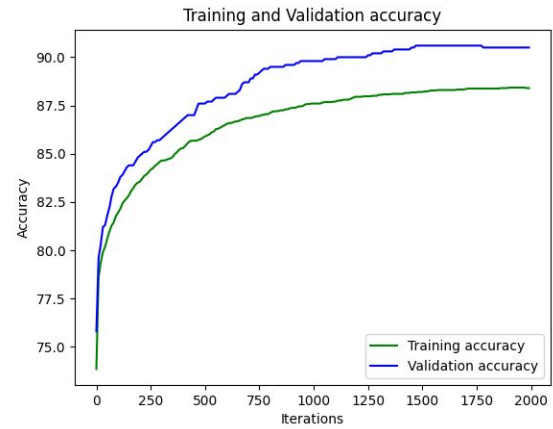
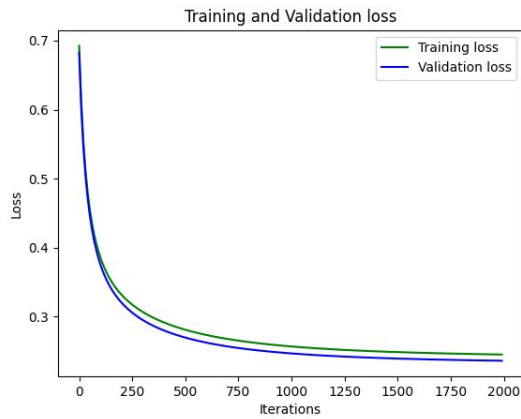


With Regularization

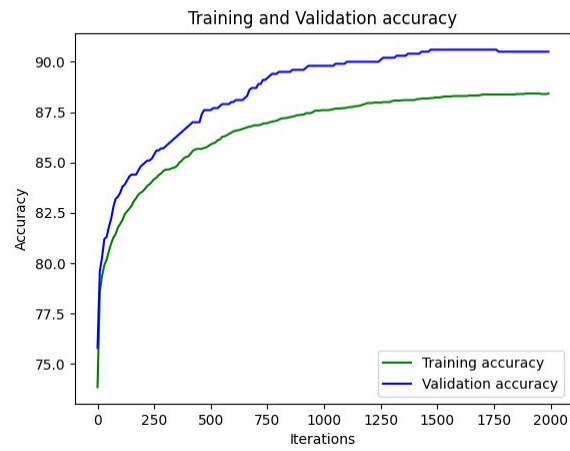
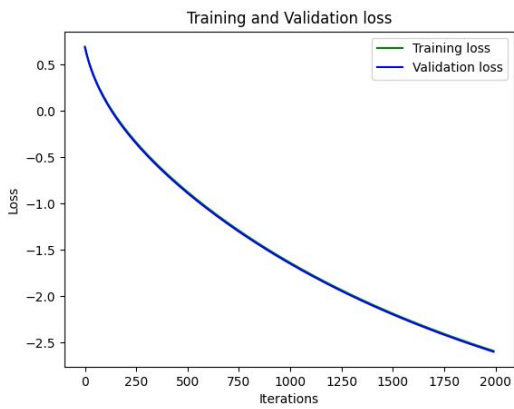


Fold 4

Without Regularization

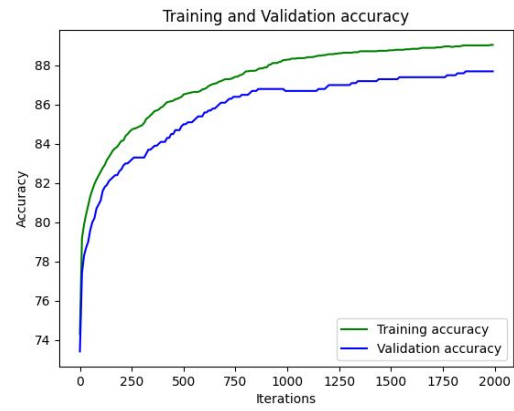
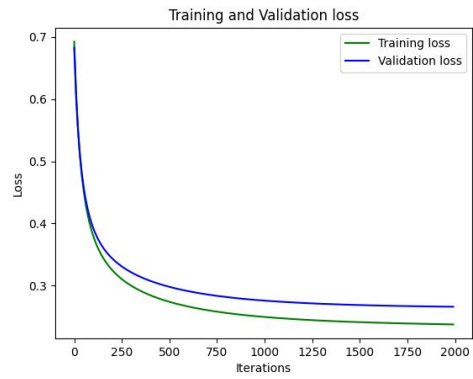


With Regularization

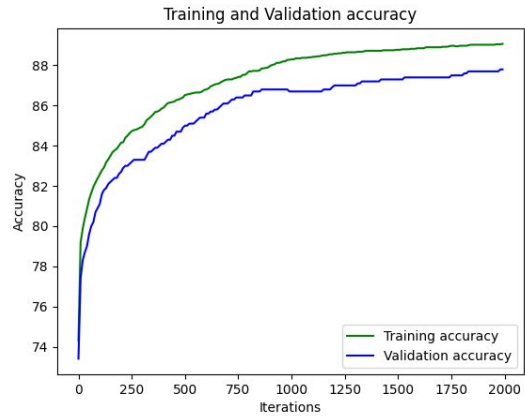
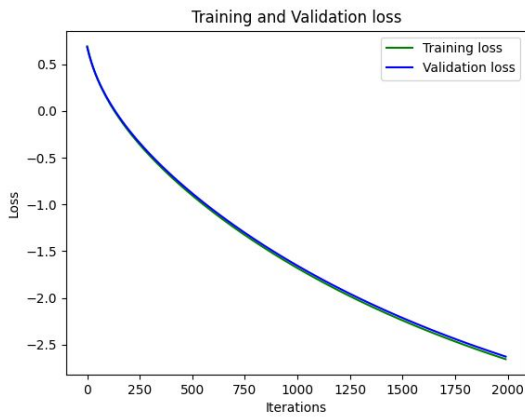


Fold 5

Without Regularization

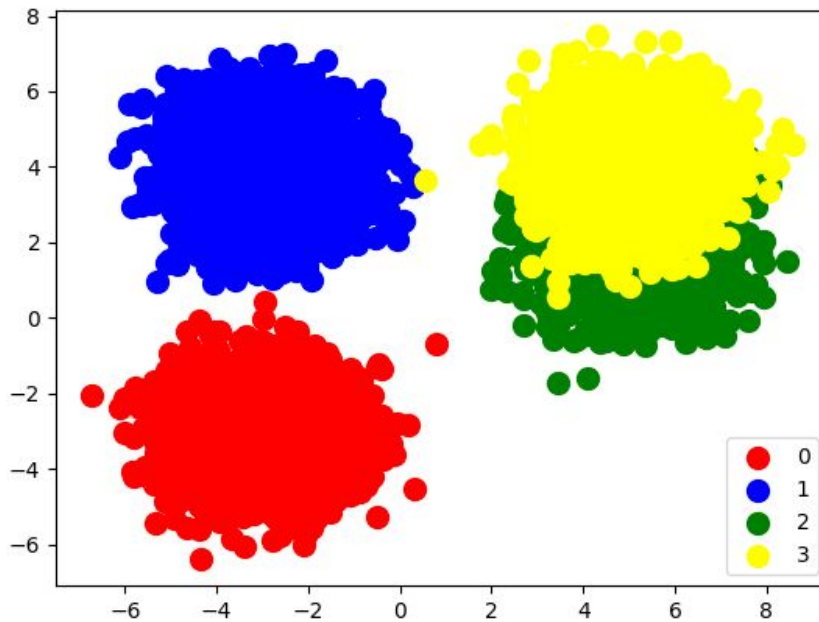


With Regularization



Ques. 3.

a) Scatter Plot



b)

MutiClass Logistic Regression (One Vs One Approach) Results

Fold	LogRegression (Without Regularization)		LogRegression (With L2 Regularization)		Sklearn Logistic Regression (OVO)	
	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy
1	91.9625	92.25	91.9625	92.25	92.2375	93.25
2	91.975	92.4	91.975	92.4	92.375	93.3
3	92.0125	92.0	92.0125	92.0	92.525	92.15
4	92.2	91.1	92.2	91.1	92.625	91.85
5	91.9875	92.15	91.9875	92.15	92.6125	92.0
Mean Accuracy	92.0275	91.98	92.0275	91.98	92.475	92.51

Class Wise Accuracy

Fold	Class 1 Accuracy	Class 2 Accuracy	Class 3 Accuracy	Class 4 Accuracy
1	99.81	100	83.78	85.07
2	100	100	81.06	89.08
3	99.78	100	83.95	85.44
4	100	100	82.99	81.7
5	100	100	81.3	86.08

c)

MutiClass Logistic Regression (One Vs All/One Vs Rest Approach) Results

	LogRegression (Without Regularization)		LogRegression (With L2 Regularization)		Sklearn Logistic Regression (OVR)	
Fold	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy
1	91.76	92.15	91.74	92.15	92.175	92.95
2	91.75	92.25	91.725	92.25	92.25	93.2
3	91.825	91.9	91.8	91.9	92.3875	92.0
4	92.1	90.85	92.075	90.85	92.5375	91.45
5	91.7875	92.0	91.7875	91.9	92.425	92.2
Mean Accuracy	91.845	91.83	91.825	91.81	92.355	92.36

Class Wise Accuracy

Fold	Class 1 Accuracy	Class 2 Accuracy	Class 3 Accuracy	Class 4 Accuracy
1	100	100	82.74	85.46
2	100	100	80.49	89.07
3	100	100	82.97	85.82
4	100	100	81.56	82.08
5	100	100	80.69	86.08

Analysis of Results

Sklearn implementation of OVO and OVR approach performs little better (~0.5%) than the implemented version of OVO and OVR approaches respectively.

From class wise accuracy, it can be inferred that the class 1 and class 2 are well separated while class 3 and class 4 are having some overlapping region which can be verified from the scatter plot.

In both the approaches OVO and OVR, l2 regularization does not improve the performance as the model without regularization is simple and performs well already.