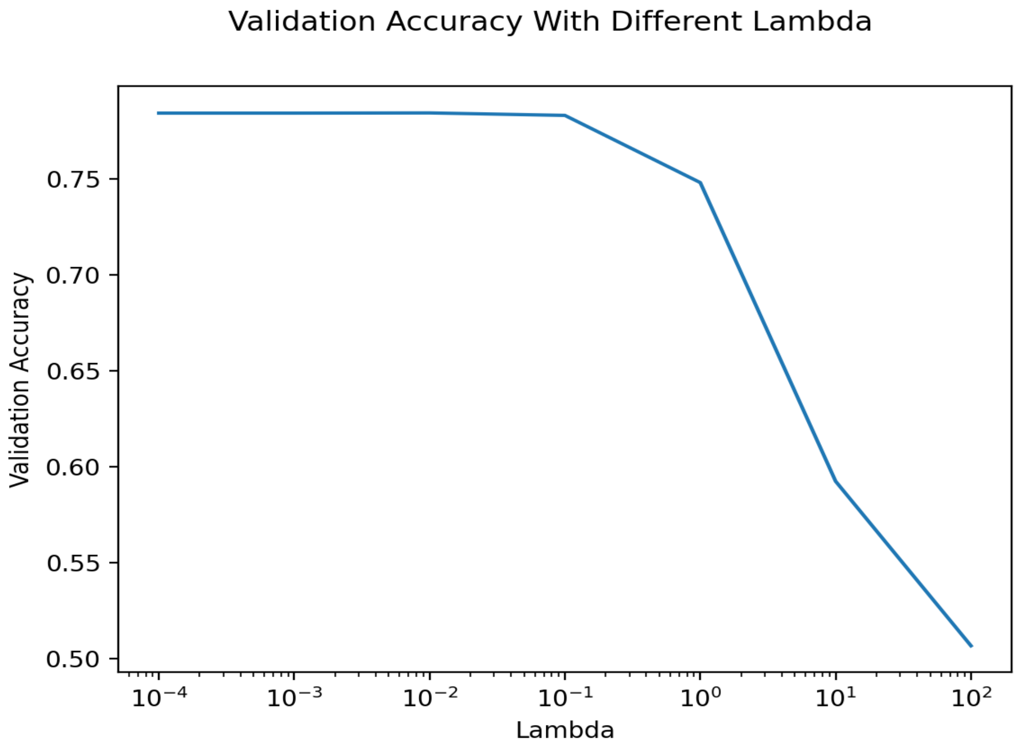
AI_534
MACHINE LEARNING

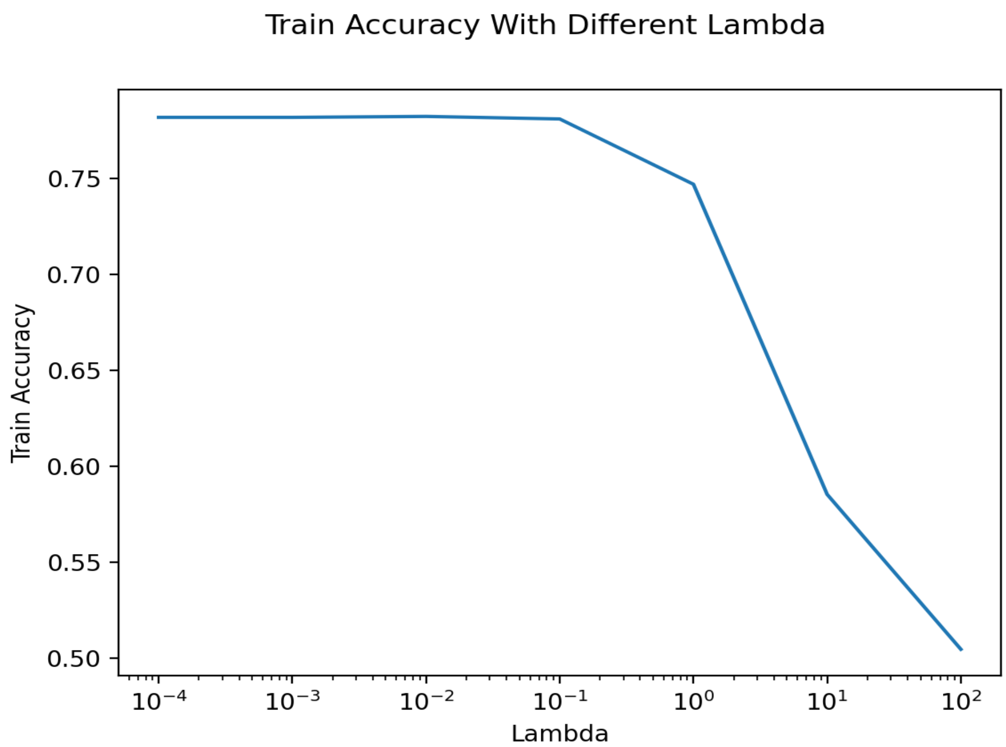
IA2 Report

CHI-CHIEH WENG
TSU-CHING LIN

Part 1

(a)





Both the results are based on a learning rate of 0.01. In our experience, we find that both accuracies (verification accuracy and training accuracy) decrease as the lambda increases. The main reason is that the bigger lambda value will cause a more significant penalty of the w. In our experience, the best lambda value based on validation accuracy is 10^{-4} .

(b)

λ_ = 10^{-4}

1: Previously_Insured
2: Vehicle_Damage
3: Vehicle_Age_1
4: Policy_Sales_Channel_160
5: Driving_Lincense

λ* = 10^{-2}

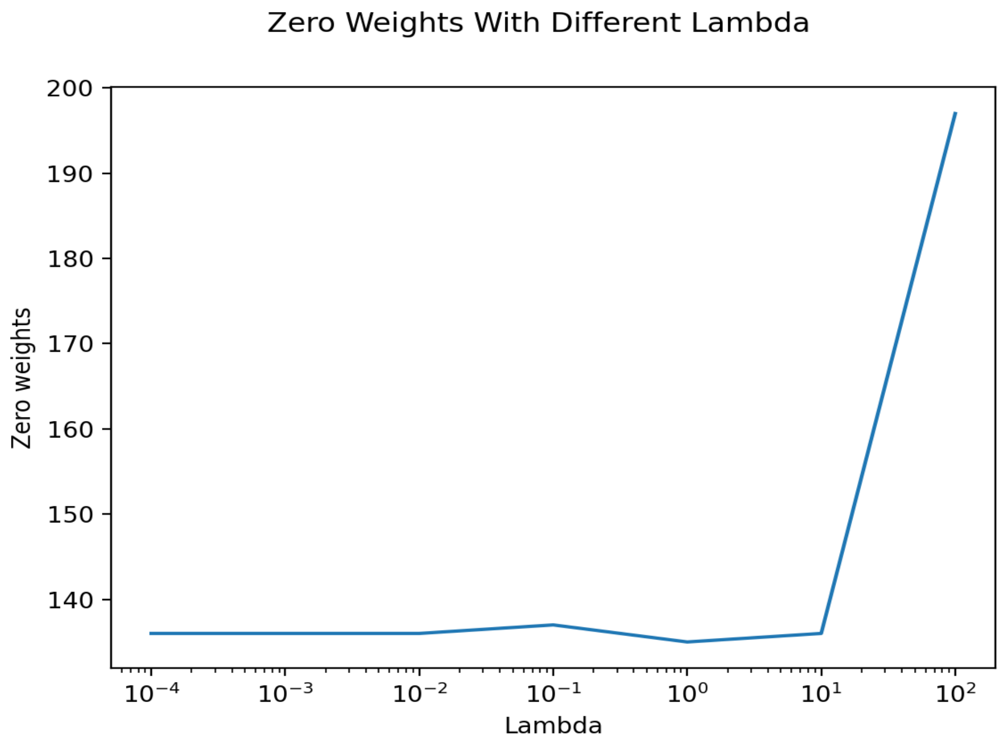
1: Previously_Insured
2: Vehicle_Damage
3: Policy_Sales_Channel_152
4: Policy_Sales_Channel_160
5: Vehicle_Age_1

λ+ = 10^{0}

1: Previously_Insured
2: Vehicle_Damage
3: Vehicle_Age_1
4: Policy_Sales_Channel_152
5: Age

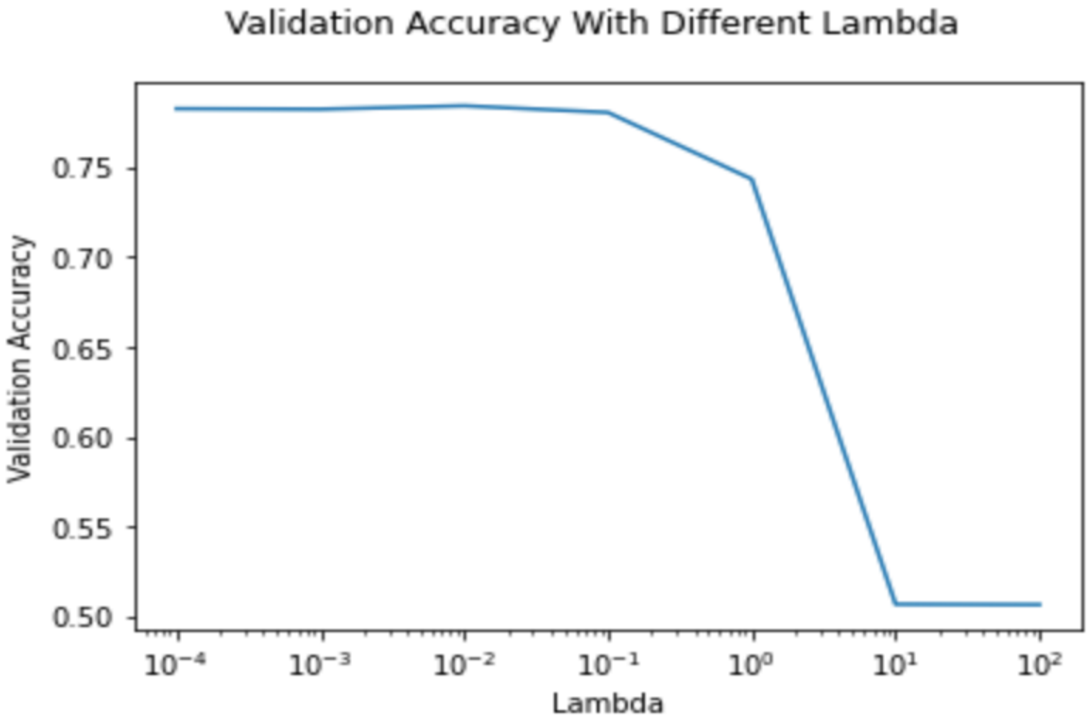
The first two features hardly change, and we believe that the top features tend to stop regulating between λ_ and λ*, because the penalty provided by the λ value is small and does not produce any change in the weights. As the λ value increases, the feature value starts to change because the penalty starts to increase, which has a significant effect on the weight. However, as the value of λ increases too much, the top feature will eventually remain unchanged and we believe that the regularization provides too heavy a penalty, resulting in no change in the weights.

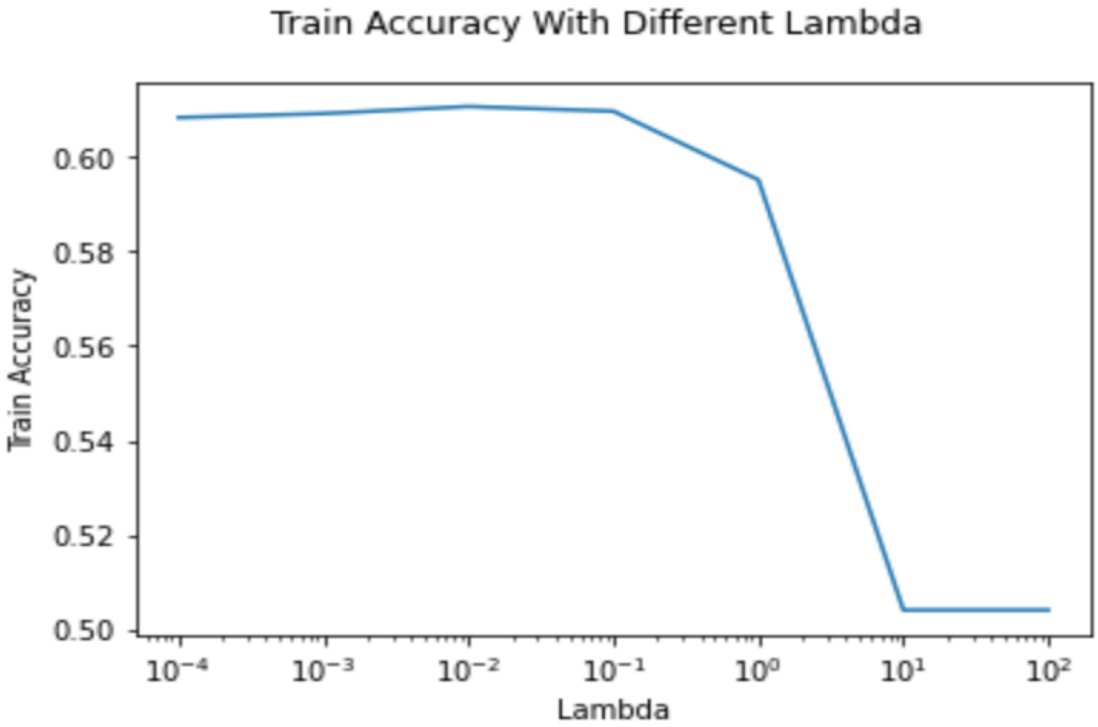
(c)



As the result, the weights become more and more sparse as the λ value increases. We believe that the weights should become more and more sparse if the λ value is further increased. The reason for this may be that larger regularization terms lead to excessive penalties during the training process.

Part 2

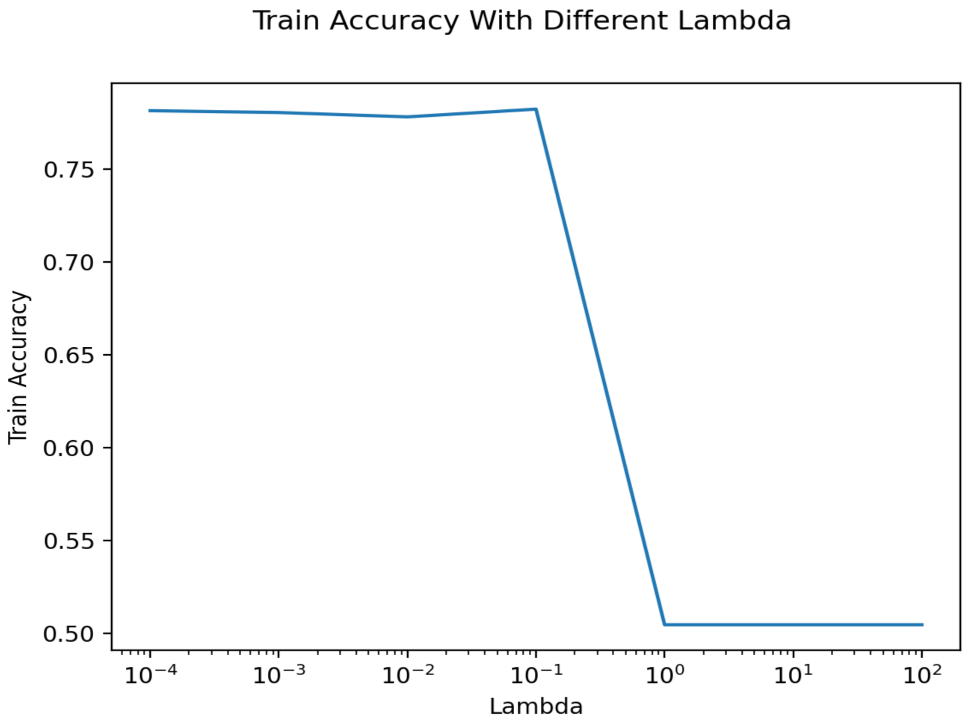


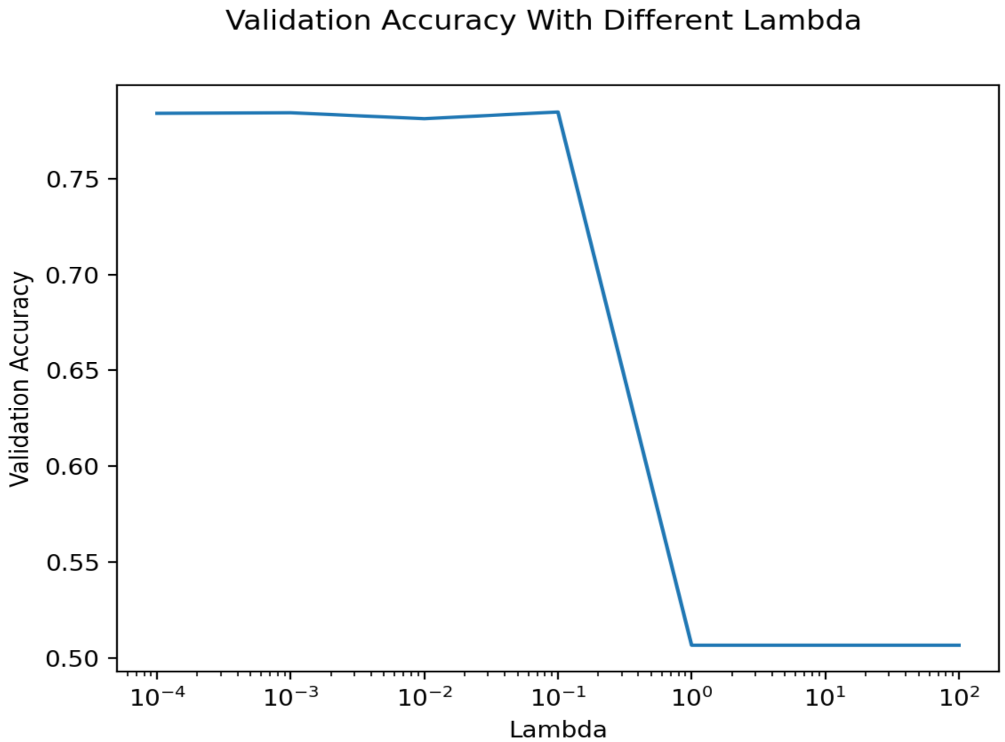


The accuracy becomes lower in our experience. The main reason is that there are more noisy data would cause less penalty. And as the lambda value increase, the accuracy is slower to decrease compared to the non-noisy data. The reason is that the data contain some useless data, so the penalty would also become less significant to the result.

Part 3

(a)





Both the results are based on a learning rate of 0.01, and similar to the L2 regularization, the increase in λ does not vary much for smaller λ, but increases slightly when λ is 10^{-2} and 10^{-1} , after which the accuracy decreases. the best lambda value based on validation accuracy is 10^{-1} .

(b)

λ_ = 10^{-4}

1: Previously_Insured
2: Vehicle_Damage
3: Vehicle_Age_1
4: Policy_Sales_Channel_152
5: Driving_Lincense

λ* = 10^{-2}

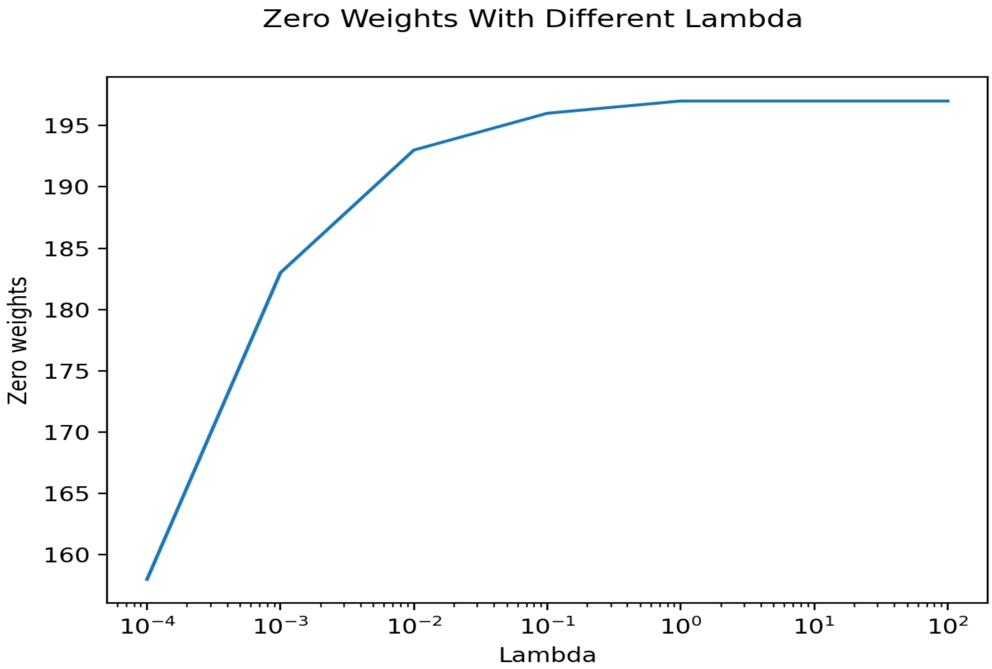
1: Previously_Insured
2: Vehicle_Damage
3: Policy_Sales_Channel_152
4: Vehicle_Age_1
5: Age

λ+ = 10^{0}

1: Policy_Sales_Channel_81
2: Policy_Sales_Channel_88
3: Policy_Sales_Channel_87
4: Policy_Sales_Channel_86
5: Policy_Sales_Channel_89

Similar to the L2 regularization in part I(b), the top feature tends to stop regulating between λ_ and λ* because the λ value provides a small penalty and does not produce any change in the weight. As the value of λ increases, the features start to change because the penalty starts to increase. However, unlike the L2 regularization in part I(b), there is a drastic change in the feature value at the time of λ+. We believe that this is due to the different algorithms between L1 and L2, which causes the non-important features to become 0, thus leading to a drastic change in this stage.

(c)



Similar to part 1 (c), the weights become more and more sparse as the value of λ increases. If the value of λ is further increased, the weights are expected to become increasingly sparse. However, in the L1 regularization, the sparsity increases rapidly when λ is 10^{-4} or larger, which is faster than the L2 regularization. In addition, we believe that as the value of λ increases to infinity, the penalty on the weights becomes very heavy, and then most of the weights will become zero.

(d)

The main difference between the L1 and L2 regularization methods is the L1 regularization would cause the result to become more sparse. In our experience, the result would have more zero in the L1 regularization method. This is because the computing method is different.

The L2 regularization method achieves the best validation accuracy.
The L1 regularization method is more sensitive to the data parameter, and it also would cause the feature weights to become sparser.

The advantages and disadvantages of L1 and L2 regulations are below:
Advantages:
L1: The cost of outliers in the data is lower than L2 regularization. Based on the L1 compute method, the cost would just increase linearly when the data increase.
L2: The cost of weights is lower than L1 regularization. Because L2 regularization takes the square of the weights, it can solve in terms of matrix math.

Disadvantages:
L1: The cost of weights is higher. Because the L1 regularization method involves taking the absolute values of the weights, which means the solution can not use a piecewise function to solve. That causes the computing cost to increase.

L2: There is a higher cost of outliers in the data, because the method use square to compute the weights. Therefore, as the data set increases, the cost of outliers would increase exponentially.