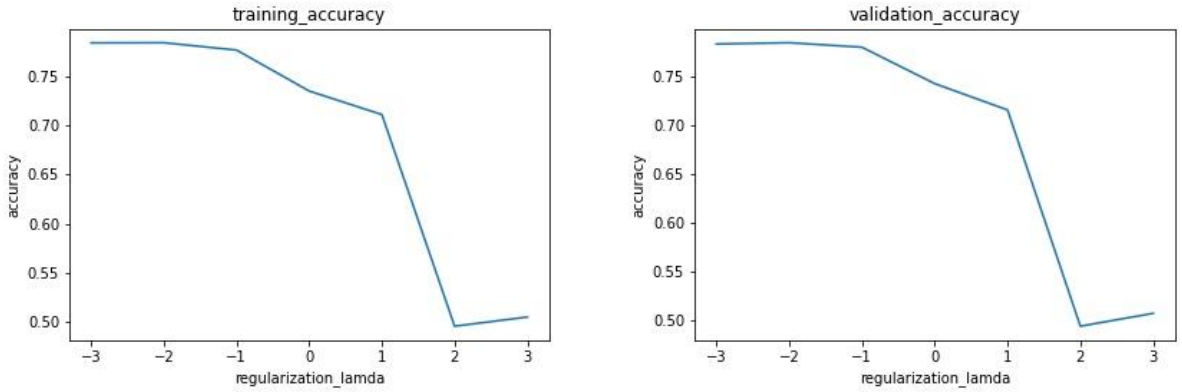
IA2 report
Name: Yijie Ren

General Introduction:
This report illustrates the results of both Logistic Regression with L1 and L2 regularization. It is surprising for me that as λ value increases, the weights tend to be zeros. The main experience I learned from the following experiments is that the regularization terms are useful but cannot be too large. Also, the regularization terms do not always improve the results.

Part I:
(a)



Both the above figures are based on learning rate 10^{-2}. As λ value increases, the accuracy tends to decrease in the training accuracy figure. The reason of this trend is mainly because the larger λ value will give too "heavy" penalties during the update process of w. The trend of validation accuracy remains the same as the training accuracy. The best value λ value based on validation accuracy is 10^{-3}.

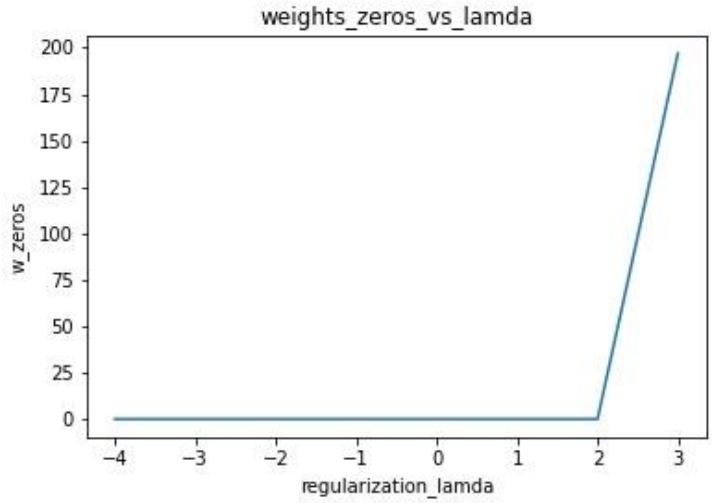
(b)
λ_{-} = 0.0001, top 5 features are 'Previously_Insured', 'Vehicle_Damage', 'dummy', 'Driving_License', 'Policy_Sales_Channel_26'

λ^{*} = 0.01, top 5 features are 'Previously_Insured', 'Vehicle_Damage', 'dummy', 'Driving_License', 'Policy_Sales_Channel_26'

λ_{+} = 1.0, top 5 features are 'Previously_Insured', 'Vehicle_Damage', 'Vehicle_Age_1', 'Policy_Sales_Channel_152', 'Age'

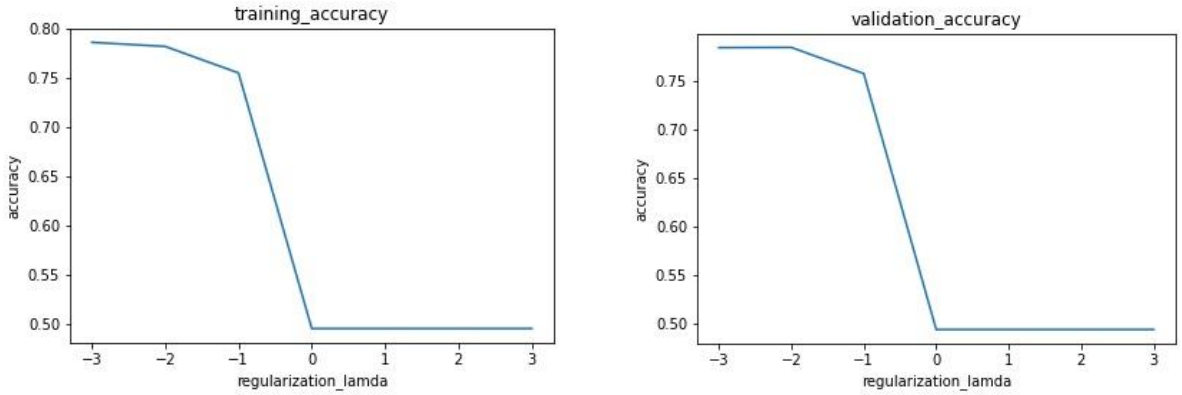
As λ value decreases, the top features tend to stop regulating, which remains the same as λ^{*}, since the penalty provided by λ value is too small, which does not make any changes on weights. As λ value increases, the top features start to change, since the penalties start to increase, which can make a lot of differences to the weights. However, as λ value increases too much, the top features will remain the same eventually, since regularization provides too heavy penalties, which leads to non-variance of the weights.

(c)



The above figure is based on learning rate = 10^{-3}, which is clearer. As λ value increases, the weights become sparser. If further increase λ value, I expect the weights array becomes sparser and sparser. The reason can be that larger regularization terms will lead to "too heavy penalties" during the training process.

Part II:
(a)



Both the above figures are based on learning rate 10^{-2}. The results are very similar to L2 regularization in Part I (a). As λ value increases, the accuracy tends to decrease in the training accuracy figure. The reason of this trend is mainly because the larger λ value will give too "heavy" penalties during the update process of w. The trend of validation accuracy remains the same as the training accuracy. The best value λ value based on validation accuracy is 10^{-3}.

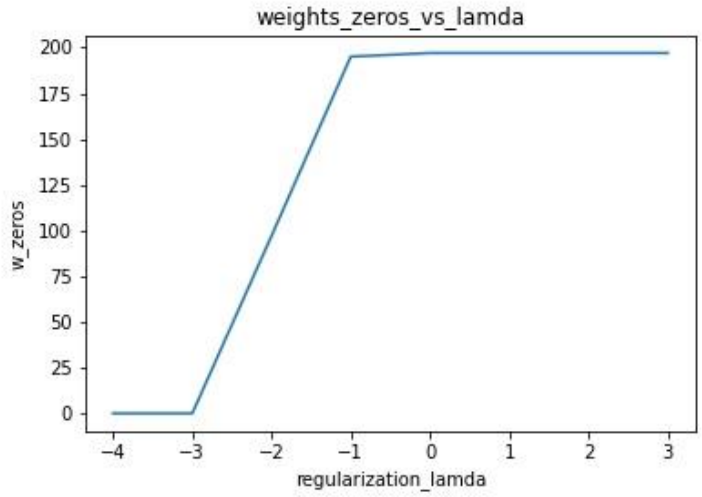
(b)
λ_= 0.0001, top 5 features are 'Previously_Insured', 'Vehicle_Damage', 'dummy', 'Driving_License', 'Policy_Sales_Channel_26'

λ^* = 0.01, top 5 features are 'Previously_Insured', 'Vehicle_Damage', 'dummy', 'Driving_License', 'Policy_Sales_Channel_26'

λ_+ = 1.0, top 5 features are 'Policy_Sales_Channel_163', 'Policy_Sales_Channel_11', 'Policy_Sales_Channel_9', 'Policy_Sales_Channel_8', 'Policy_Sales_Channel_7'

Similar to L2 regularization in Part I (b), as λ value decreases, the top features tend to stop regulating, which remains the same as λ^*, since the penalty provided by λ value is too small, which does not make any changes on weights. As λ value increases, the top features start to change, since the penalties start to increase, which can make a lot of differences to the weights. However, as λ value increases too much, the top features will remain the same eventually, since regularization provides too heavy penalties, which leads to non-variance of the weights.

(c)



Similar to Part I (c), as λ value increases, the weights become sparser. If further increase λ value, I expect the weights array becomes sparser and sparser. The reason can be that larger regularization terms will lead to "too heavy penalties" during the training process. The trend is the same as I observed in 1(c). Since both L1 and L2 are regularization terms, as λ value increases towards infinity, both penalties for weights become too heavy, then most of the weights will become 0.

Part III:
I used L2 regularization and combined both training and validation datasets for training purpose. I also turned the remaining 3 features (Age, Vintage, Annual_Premium) into categories roughly based on the frequencies of different ranges, and then conducted one hot encoding to all 3 of them.

I think the limitation comes from the dataset itself, as well as the L1 & L2 regularization.