Colton Avila

AI534

15 October 2021

IA2 – Ridge and Lasso Logistic Regression

Pre-Processing

We modified the training and validation sets by normalizing the following features using z-score: ‘Age’, ‘Vehicle_Ago_0’, ‘Vehicle_Ago_1’, ‘Vehicle_Age_2’, ‘Annual Premium’, ‘Vintage’.

Logistic Regression with L2 (Ridge) regularization

1a)

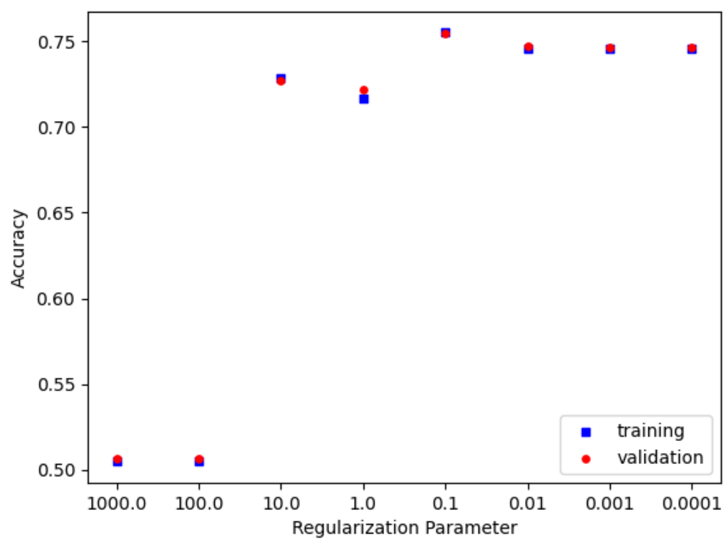


Figure 1.) Regularization parameter (λ) versus accuracy of prediction on the training and validation sets.

As we increase λ from the smallest value, 0.0001, we see an increase in accuracy until λ reaches 0.1. After reaching 0.1, the accuracy declines. 1000 and 100 had trouble converging so their accuracy was even lower than the lowest values. This occurs because at lower λ values our model overfits the data, while at larger λ values our model will ignore the training data. We see similar trends with the training data as we do with the validation data. A λ value of 0.1 resulted in the highest accuracy

1b)

We compared the L1 model’s top 5 features for three λ values, 1, 0.1 and 0.01.

Rank	λ = 1	λ = 0.1	λ = 0.01
1	Previously Insured	Previously Insured	Previously Insured
2	Vehicle_Age_1	Vehicle Age_1	Vehicle_Age_1
3	Dummy	Policy_Sales_Channel_152	Policy_Sales_Channel_152
4	Policy_Sales_Channel_152	Dummy	Dummy
5	Policy_Sales_Channel_160	Driving License	Driving License

Table 1. Top 5 features for the regression values near the optimal value.

The top 5 features for λ = 0.1 and λ = 0.01 are the same. However, when we increase the regularization parameter to 1, we see "Driving License" be replaced by "Policy_Sales_Channel_160". This happens because the model decides that with a larger regularization parameter, this Driving License is sensitive to our training data set and the Policy Sales Channel feature works better as a generalized predictor.

1c)

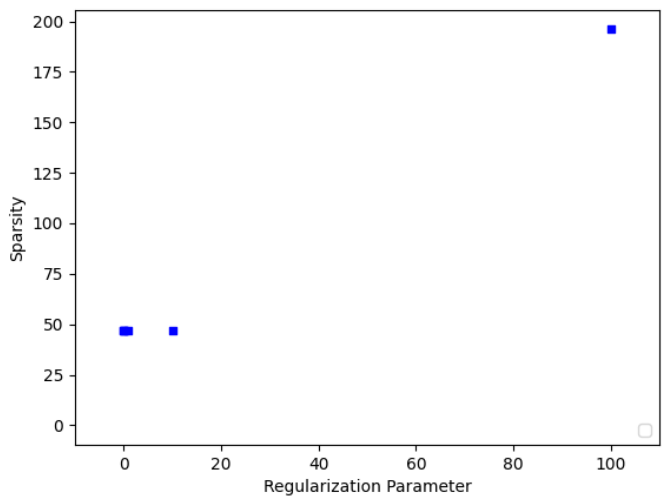
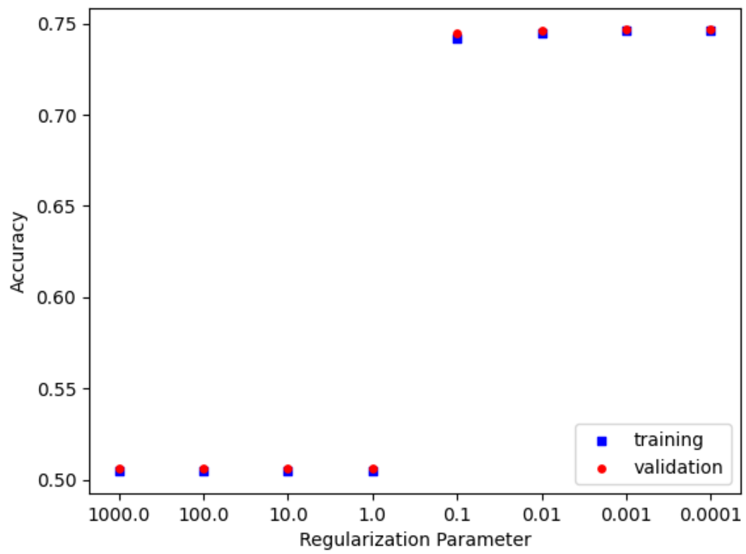


Figure 2. Sparsity (amount of removed parameters) against the regularization parameter.

We calculated sparsity as the sum of feature weights that equaled zero. In L2 regularization we do not eliminate any features from the model, so we do not see a sparse model emerge with an increase in λ. In Figure 2 we see sparseness emerge from values that are all zero. We also see a large value for λ = 100, but that is due to lack of convergence.

Logistic Regression with L1 (Lasso) Regularization

2a)



We see that as we decrease λ, our accuracy for the training and validation set increase. We see a large jump from 1.0 to 0.1 due to the increase in features that were not eliminated. If we continued to lower the λ, we would see the training and validation points diverge. This is due to overfitting on the training data. Our best λ value for the values we tested were 0.001

2b)

We compared the model's top 5 features for three λ values, 0.01, 0.001 and 0.0001.

Rank	λ = 0.01	λ = 0.001	λ = 0.0001
1	Previously Insured	Previously Insured	Previously Insured
2	Vehicle Age 1	Vehicle Age 1	Vehicle Age 1
3	Policy Sales Channel 152	Policy Sales Channel 152	Policy Sales Channel 152
4	Dummy	Dummy	Dummy
5	Driving License	Driving License	Driving License

Table 1. Top 5 features for the regression values near the optimal value.

The top 5 features all λ values are the same. For these λ's, we do not eliminate these values thus we keep the best performing features.

2c)

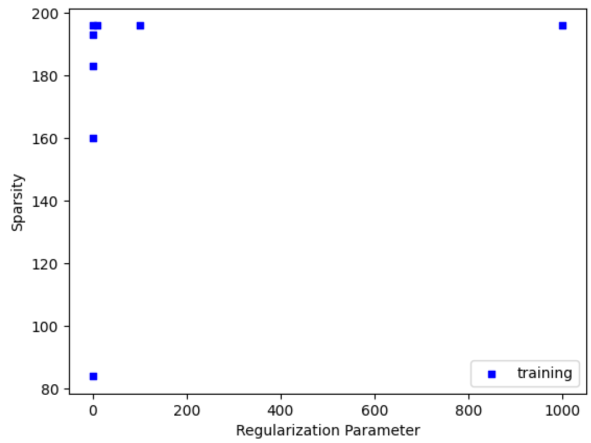


Figure 2. Sparsity of L1 Regularization Parameters

We see that as we raise λ the sparsity increases. This occurs because L1 Regularization eliminates coefficients. As we increase λ we increase bias making these models rely less on the training data.

3)

To better increase our accuracy for the Kaggle Competition (team name: Colton Avila), we tried to introduce second order features. We tried combinations of the 'Annual Premium', 'Vintage', and 'Age' features. We also tried squaring these values as well. We also tried removing outliers outside of the 5^{th} and 95^{th} quantile. We used L2 Regularization and ended up getting an accuracy of 75.6%. This value was low due to the lack of data and domain knowledge.