Tested running on flip1 server

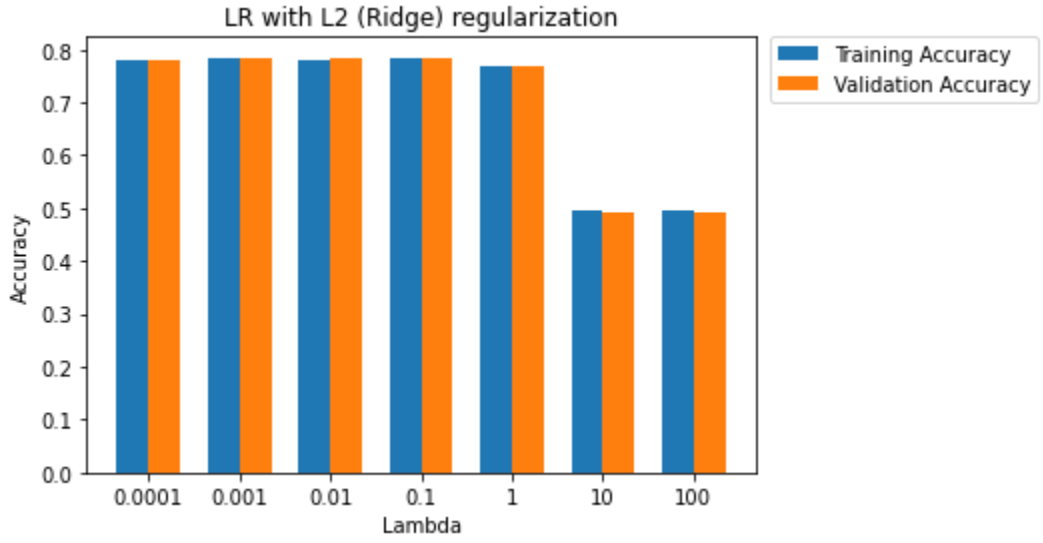
IA2 - Group 26

Cheng Zhen
Bharath Padmaraju
Bharghav Srikhakollu

Logistic Regression with L2 and L1 regularizations

Part - 1 : Logistic Regression with L2 (Ridge) regularization

(a) Plot the training accuracy and validation accuracy of the learned model as a function of i used for λ .
You can either plot both in the same figure or in separate figures. If separate, please align the figures in your report so that we can compare across.



Question: What trend do you observe for the training accuracy as we increase λ ? Why is this the case? What trend do you observe for the validation accuracy? What is the best λ value based on the validation accuracy?

Answer:

→ After multiple trials with different learning rates, "α" is set as "1"

→ As we increase the regularization parameter, from the plot we can see that the training accuracy first increases a bit, then slowly starts to decrease and at a certain point it becomes constant. Since we are increasing the regularization parameter it adds a penalty term which

minimizes the weights. It also reflects that we are giving more weightage to regularized terms than the data terms.
→ As we increase the regularization parameter, we see the similar trend for validation accuracy as well from the plot.
→ According to our experiments and plots the best "λ" value based on the validation accuracy is "0.1"
Training Accuracy is "0.7830"
Validation Accuracy is "0.7847".

(b) Consider the best λ* selected in (a), a value λ_ that is smaller than λ*, and a λ+ that is bigger than λ*. Report for each of three λ values, the resulting model's top 5 features with the largest weight magnitude |wj| (excluding w0).

→ Top 5 features with the largest weight magnitude for λ*(0.1)

Features	**Weights**
Vehicle_Damage	0.65910464
Previously_Insured	0.61251861
Policy_Sales_Channel_152	0.24531169
Vehicle_age_1	0.22116844
Vehicle_age_0	0.14282666

→ Top 5 features with the largest weight magnitude for λ+ (1)

Features	**Weights**
Vehicle_Damage	0.12227881

Previously_Insured	0.11298361
Vehicle_age_1	0.06332898
Policy_Sales_Channel_152	0.06190881
Vehicle_age_0	0.05048291

→ Top 5 features with the largest weight magnitude for λ- (0.01)

Features	**Weights**
Vehicle_Damage	1.25234572
Previously_Insured	1.15024044
Policy_Sales_Channel_2	0.71538501
Policy_Sales_Channel_134	0.71449817
Policy_Sales_Channel_110	0.6941016

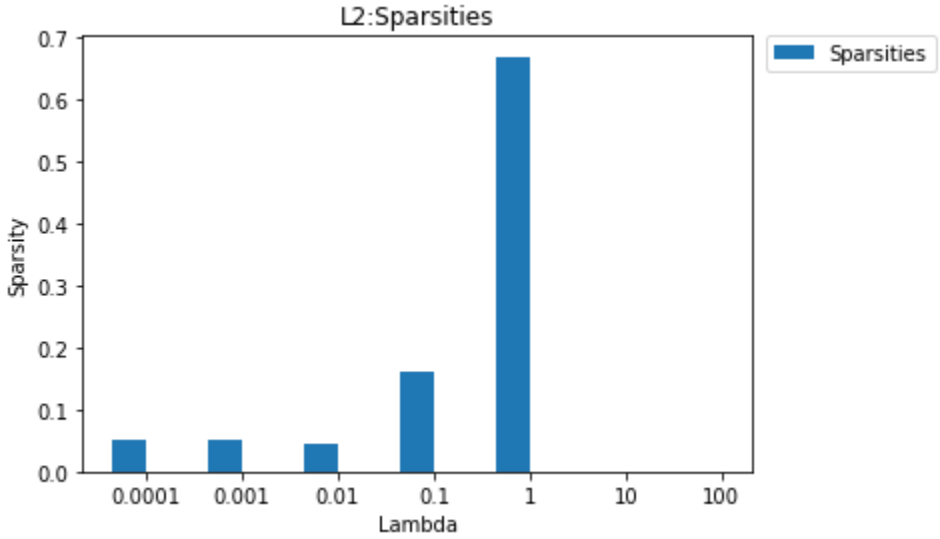
Question: Do you see differences in the selected top features with different λ values? What is your explanation for this behavior?

Answer:

→ For λ* and λ+ the top 5 features stayed the same and for λ- two out of the top 5 features in λ* are same and the remaining 3 features were new features for λ-

→ We observe that with an increase in λ, the magnitude of the top 5 features weights decreases. More λ value indicates added penalty term which inturn reduces the dependency of the model on the data. With larger values of λ, the weights almost get closer to zero for the features.

(c) For different values of λ, compute the sparsity of the model as the number of weights that approximately equal zero (<= 10^-6) and plot it as a function of i for λ.



Question: What trend do you observe for the sparsity of the model as we change λ ? If we further increase λ, what do you expect? Why?

Answer:

→As lambda increased, the sparsity increased. If we further increase lambda, the sparsity will not increase much. Instead, weights are moving close to 0 but not reaching it.

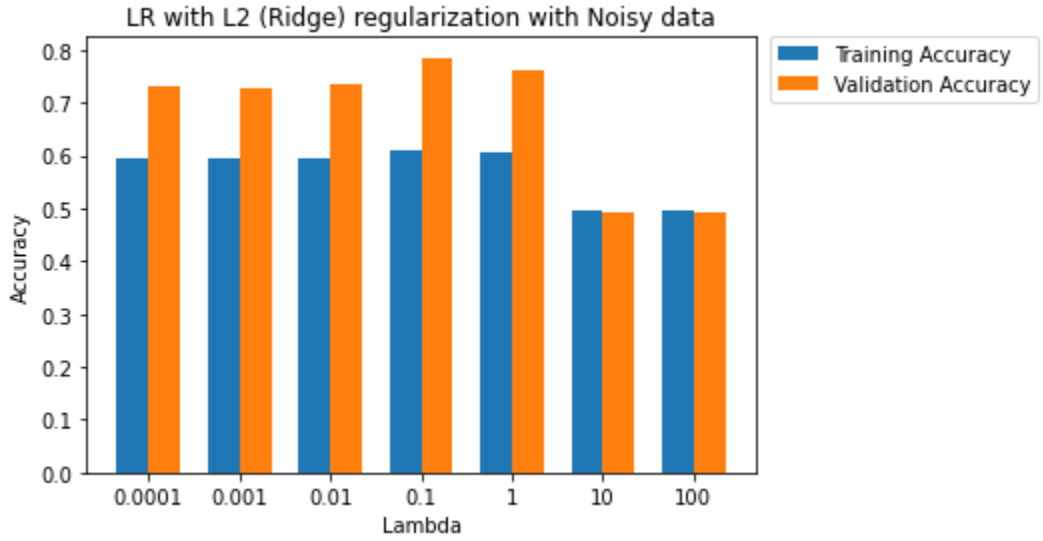
Part - 2 : Experiment with noisy training data

Question: What are some of the key differences you observe comparing the results obtained using noisy training data to those of part 1? What do you think is the effect of regularization on the model's robustness to noise in the training set? Why?

Answer:

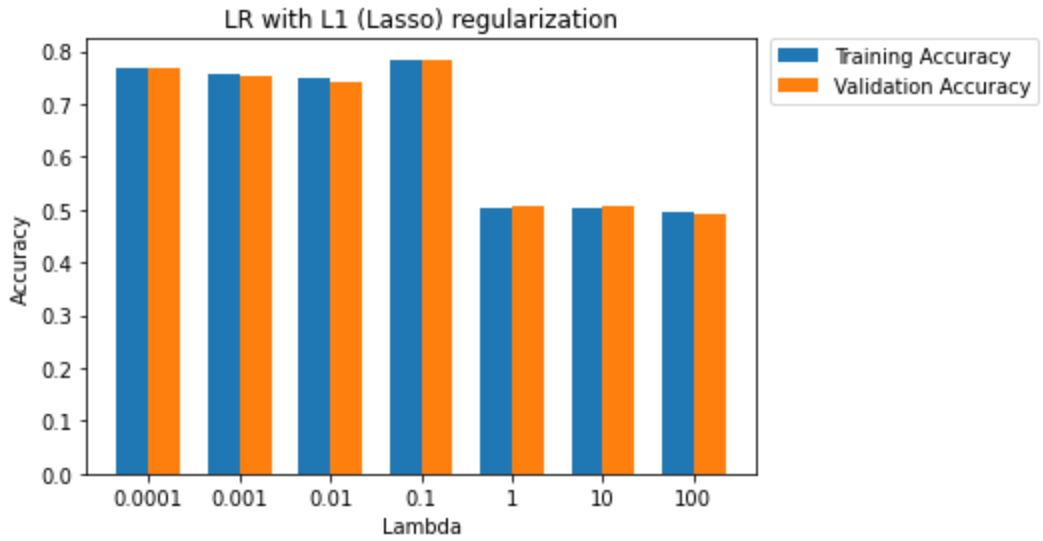
→ With noisy training data we observe that training accuracy is very less in comparison with the training accuracy with normal train data.

Also, we can observe from the plot that the training accuracy with noisy data is very less than the validation accuracy.
→ The regularization helps increase the robustness of the model.
→ Despite the noisy training data, we still got a decent model who reached a good accuracy for validation data over 70%. This means the quality of the model was not affected by the noise much.



Part - 3 : Logistic Regression with L1(Lasso) regularization

(a) Plot the training accuracy and validation accuracy of the learned model as a function of i used for the λ value.



Question: What trend do you observe for the training accuracy as we increase λ ? Why is this the case? What trend do you observe for the validation accuracy? What is the best λ value based on the validation accuracy?

Answer:

→ After multiple trials with different learning rates, "α" is set as "0.01"

→ As we increase the regularization parameter, from the plot we can see that the training accuracy first increases a bit, then slowly starts to decrease and at a certain point it becomes constant. With Lasso, we observe that when λ value increases, some features which are not of much importance tend to have weights to 0. This is the expected behavior with Lasso as it will help in reduction in features and by resulting in a sparser solution.

→ As we increase the regularization parameter, we see the similar trend for validation accuracy as well from the plot.

→ According to our experiments and plots the best "λ" value based on the validation accuracy is "0.1"

Training Accuracy is "0.7821"

Validation Accuracy is "0.7842".

(b) Consider the best λ* selected in 2(a), a value λ_ that is smaller than λ*, and a λ+ that is bigger than λ*. Report for each of three λ values, the resulting model's top 5 features with the largest weight magnitude |wj| (excluding w0).

→ Top 5 features with the largest weight magnitude for λ* (0.1)

Features	**Weights**
Vehicle_Damage	0.28598237
Region_Code_22	0.18606907
Policy_Sales_Channel_54	0.18020487
Policy_Sales_Channel_134	0.17933412
Policy_Sales_Channel_35	0.17260754

→ Top 5 features with the largest weight magnitude for λ+ (1)

Features	**Weights**
Policy_Sales_Channel_1	0.00000000
Policy_Sales_Channel_2	0.00000000
Vehicle_Age_2	0.00000000
Policy_Sales_Channel_3	0.00000000
Policy_Sales_Channel_163	0.00000000

→ Top 5 features with the largest weight magnitude for λ- (0.01)

Features	Weights
Policy_Sales_Channel_73	0.91524275
Policy_Sales_Channel_54	0.91381185
Policy_Sales_Channel_109	0.91377687
Policy_Sales_Channel_32	0.91067659
Policy_Sales_Channel_118	0.9060509

Question: Do you see differences in the selected top features with different λ values? What is your explanation for this behavior?

Answer:

→ For λ*, λ+ and λ- the top 5 features changed and we can observe that with the increment of λ the magnitude of weights of the features decreases.

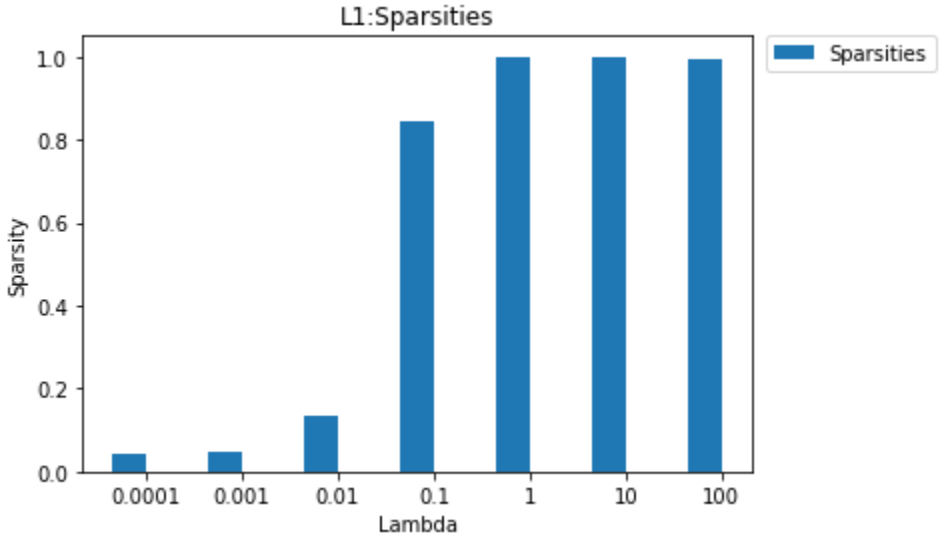
→ From the formula given:

w_j <- sign(w_j) max(|w_j | - αλ, 0)

As we fixed one particular learning rate, the change depends on the λ. When the λ is small we can say that the above expression will result in some constant value but with the increment of λ value this will become 0 value which is the weight coefficient. Because of this we can see that the magnitude of weights of features move closer to absolute zero values.

→ Lasso regularization generates sparser solutions. As we increase the λ value, we are getting 0 for weight coefficients. If we further increase the λ value, the sparsity of the model will also increase and there is a possibility of underfitting the data.

(c) For different values of λ, compute the 'sparsity' of the model as the number of weights that equal zero and plot it against λ.



Question: What trend do you observe for the sparsity of the model as we change λ ? If we further increase λ, what do you expect? Is this trend different from what you observed in 1(c)? Provide your explanation for your observation.

Answer:

→ As lambda increased, we saw an increase in sparsity. If we further increase lambda, the sparsity will keep increasing. No this is a different trend compared to that of 1(c). Our experimental results showed that sparsity kept going up all the way to ~100%. This is because Lasso regularizer pushes the weights going to 0.

(d) Finally,please compare the results acquired in part 3 with that of part 1.

Question: What are the key differences between the two regularization methods observed on this dataset? Specifically, which method achieves the best validation accuracy? Which method is more sensitive to the choice of the regularization parameter for this data? Which method produced sparser feature weights? What are the advantages and disadvantages of each method in general?

Answer:

→ Lasso regularization is more sensitive to learning rate than L2 regularization on this data set. The sparsity of models learned with L1 is higher than that from L2 with a large lambda value on this data set.

→ We got very similar validation accuracies from the two regularization methods. The difference in between is negligible.

→ Lasso regularization is more sensitive to the regularization parameter

→ Lasso regularization method produced sparser feature weights.

→ Advantages of Ridge Regularization:

- Useful for solving multi collinearity problems.
- Useful when the size of the data is less (less features)

→ Disadvantages of Ridge Regularization:

- Not good for feature reduction as it minimizes the coefficients but never brings them to absolute zero.

→ Advantages of Lasso Regularization:

- Useful for feature reduction as it tends to make coefficients to absolute zero
- Reduces the complexity of the model for better performance

→ Disadvantages of Lasso Regularization:

- Not suitable when there is multicollinearity in the data.
