

Q1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

A1. In Ridge, the alpha is 7 and in Lasso, alpha is 0.001.

If we double or increase the value of alpha, there will be decrease in both Models output. That is the reason we first look for optimal value to alpha before making predictions.

We should not be looking for predictors post change because overall significance would be already reduced. Yes, we can do it to find out variable importance if change of alpha is applied.

Q2. You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

A2. My preference would be the optimal value of Alpha in Ridge because when we compared the both model output, statistics of Ridge been on higher side.

Also, with Lasso because of its behavior, it zeroes more features compared to Ridge which reduces the opportunity areas for the business.

Q3. After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

A3. After removal of top 5 predictors, below are the new predictors

- Built In Garages (GarageType_BuiltIn)
- Garages Detached from home (GarageType_Detachd)
- Homes without Garages (GarageType_NotAvail)
- Neighborhood Area Somerst (Neighborhood_Somerst)
- Basement -1 floor completed (BsmtFinSF1)

Q4. How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

A4. In my understanding, if a model structure is built on correct basic formulas instead of focusing on some specific feature\variable (except target variable), the robustness stays in it and generalizable would be visible when some of features are changes in incoming data and model continue to perform well without making changes in its code structure.

To maintain the robustness & generalizability of a model, we have to make it simple which will give us optimum results. But to get high accuracy, the model complexity has to be increased and, in this process, we would decrease the variance but that will slowly increase the bias. If bias is increased, accuracy will come down.

So, as we learned in the previous modules that we have to maintain the balance between simplicity, accuracy and complexity.