SUBJECTIVE QUESTIONS & ANSWERS:

Answer-1:

The optimal value of alpha (hyperparameter) for the Ridge regression = 10

The optimal value of alpha (hyperparameter) for the Lasso regression = 100

The optimal value has been choosen in such a way that the model does not overfit and at the same time the r squared value does not drag below 80 % range which can potentially lead to underfit of the ML model

If we choose to double the value of alpha for both Ridge and Lasso regression, then the regularization would be more i.e more feature coefficients will tend towards 0 in case of Ridge regression and more feature coefficients will become 0 in case of the Lasso regression. On top of this the r squared value will reduce too which will help us get rid of the overfit problem however it is very important to choose optimal value of the alpha so that the model does not underfit with too high alpha value either

The most important predictor variable post changes implemented – "**LotArea**"

Answer-2:

We have figured out the optimal value of lambda for both the ridge and lasso regression during the assignment however we would prefer to choose the Lasso regression since in case of the lasso regression we would not only be able to derive the optimal lambda value but at the same time we would be able to figure out the relevant feature variables of the model as well & thereby can come up with a business decision to chunk off the unwanted irrelevant variables from the dataset while coming up with predictions.

Answer-3:

The 5 most important predictor variables now are:

LotFrontage

LotArea

RoofStyle

PoolArea

Fence

Answer-4:

We can ensure that the model is robust and generalizable by choosing optimal value of lambda. That is the value of lambda should not be too low so that the model overfits and at the same time the value should not be too high so that the r squared value decreases significantly and the model underfits hence the model accuracy thereby very much depends on the choosen lambda value. The model accuracy would get significantly impacted if we are not choosing correct lambda value and we would need to tune the model with multiple iterations to decide and come up with a conclusion on the correct value of lambda. On top of these not choosing correct value of lambda would lead to incorrect coefficient assignments for feature variables and especially in case of the lasso regression if we choose too high lambda for instance then the coefficients of lot of important feature variables might become 0 which can pose problems in the overall ML model & thereby the model won't be robust leading to false predictions.