

Question-1:

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?

Answer:

Optimal Value of Alpha:

The optimal value of alpha for both Ridge and Lasso regression was determined through techniques like cross-validation. It aims to balance the trade-off between bias and variance.

Ridge Best alpha: 3.8535285937105273

Lasso Best alpha: 100.0

Doubling Alpha:

Doubling the alpha value in both Ridge and Lasso increases the strength of regularization. For Ridge, it further shrinks the coefficients towards zero, potentially increasing bias and reducing variance. For Lasso, it might lead to more coefficients being set to zero, enhancing the feature selection effect.

Predictors After Change: In Ridge, the most important predictors remain the same but their influence (coefficients) is reduced. In Lasso, some predictors (e.g., feature 107 ExterQual) are completely eliminated (coefficients set to zero), altering the set of important variables.

Considering only the positive correlated features as sample in each method, below shows the impact of doubling the co-efficients.

Top 10 Positive Features - Lasso:

	Feature	Coefficient	Original Feature	Category	Description
76	RoofMatl_WdShngl	95943.537870	RoofMatl	WdShngl	Roof material
36	Neighborhood_NoRidge	89510.353708	Neighborhood	NoRidge	Physical locations within Ames city limits
43	Neighborhood_StoneBr	59091.896173	Neighborhood	StoneBr	Physical locations within Ames city limits
59	HouseStyle_2.5Fin	58518.311284	HouseStyle	2.5Fin	Style of dwelling
120	BsmtQual_Ex	56842.899095	BsmtQual	Ex	Evaluates the height of the basement
152	KitchenQual_Ex	51677.942293	KitchenQual	Ex	Kitchen quality
107	ExterQual_Ex	41342.794979	ExterQual	Ex	Evaluates the quality of the material on the e...
37	Neighborhood_NridgHt	40210.161908	Neighborhood	NridgHt	Physical locations within Ames city limits
27	Neighborhood_Crawfor	29415.565242	Neighborhood	Crawfor	Physical locations within Ames city limits
128	BsmtExposure_Gd	29075.005643	BsmtExposure	Gd	Refers to walkout or garden level walls

Top 10 Positive Features - Lasso (Updated):

	Feature	Coefficient	Original Feature	Category	Description
36	Neighborhood_NoRidge	87184.389148	Neighborhood	NoRidge	Physical locations within Ames city limits
76	RoofMatl_WdShngl	80590.757383	RoofMatl	WdShngl	Roof material
120	BsmtQual_Ex	57140.586968	BsmtQual	Ex	Evaluates the height of the basement
43	Neighborhood_StoneBr	53914.671856	Neighborhood	StoneBr	Physical locations within Ames city limits
152	KitchenQual_Ex	50951.276340	KitchenQual	Ex	Kitchen quality
37	Neighborhood_NridgHt	36696.268587	Neighborhood	NridgHt	Physical locations within Ames city limits
59	HouseStyle_2.5Fin	34609.033696	HouseStyle	2.5Fin	Style of dwelling
27	Neighborhood_Crawfor	28072.786053	Neighborhood	Crawfor	Physical locations within Ames city limits
128	BsmtExposure_Gd	28036.825766	BsmtExposure	Gd	Refers to walkout or garden level walls
79	Exterior1st_BrkFace	25755.781896	Exterior1st	BrkFace	Exterior covering on house

Top 10 Positive Features - Ridge:

	Feature	Coefficient	Original Feature	Category	Description
36	Neighborhood_NoRidge	75542.275975	Neighborhood	NoRidge	Physical locations within Ames city limits
76	RoofMatl_WdShngl	64261.516088	RoofMatl	WdShngl	Roof material
43	Neighborhood_StoneBr	47469.462522	Neighborhood	StoneBr	Physical locations within Ames city limits
59	HouseStyle_2.5Fin	42617.929033	HouseStyle	2.5Fin	Style of dwelling
120	BsmtQual_Ex	39378.473199	BsmtQual	Ex	Evaluates the height of the basement
152	KitchenQual_Ex	37037.518581	KitchenQual	Ex	Kitchen quality
37	Neighborhood_NridgHt	35056.077793	Neighborhood	NridgHt	Physical locations within Ames city limits
107	ExterQual_Ex	25407.496812	ExterQual	Ex	Evaluates the quality of the material on the e...
79	Exterior1st_BrkFace	24515.432533	Exterior1st	BrkFace	Exterior covering on house
128	BsmtExposure_Gd	24459.702003	BsmtExposure	Gd	Refers to walkout or garden level walls

Top 10 Positive Features - Ridge (Updated):

	Feature	Coefficient	Original Feature	Category	Description
36	Neighborhood_NoRidge	67194.029453	Neighborhood	NoRidge	Physical locations within Ames city limits
76	RoofMatl_WdShngl	45941.122570	RoofMatl	WdShngl	Roof material
43	Neighborhood_StoneBr	38759.744251	Neighborhood	StoneBr	Physical locations within Ames city limits
120	BsmtQual_Ex	38241.675391	BsmtQual	Ex	Evaluates the height of the basement
152	KitchenQual_Ex	36753.557379	KitchenQual	Ex	Kitchen quality
37	Neighborhood_NridgHt	30476.699883	Neighborhood	NridgHt	Physical locations within Ames city limits
59	HouseStyle_2.5Fin	29996.464317	HouseStyle	2.5Fin	Style of dwelling
107	ExterQual_Ex	24957.986267	ExterQual	Ex	Evaluates the quality of the material on the e...
128	BsmtExposure_Gd	23229.385961	BsmtExposure	Gd	Refers to walkout or garden level walls
79	Exterior1st_BrkFace	21821.045629	Exterior1st	BrkFace	Exterior covering on house

Question-2:

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?

Answer:

The choice between Ridge and Lasso regression is made based on their Test metrics on Unseen data. Ridge might be a better choice as it is slightly better than Lasso on multiple metrics like MSE, RMSE and R-Squared value.

Ridge R-squared (Train): 0.8204142635357901

Ridge Mean Squared Error (Train): 1169110877.1003654

Ridge Root Mean Squared Error (Train): 34192.263410022526

Ridge R-squared (Test): 0.7661618987504312

Ridge Mean Squared Error (Test): 1185046663.7764618

Ridge Root Mean Squared Error (Test): 34424.50673250761

Lasso R-squared (Train): 0.818516707319751

Lasso Mean Squared Error (Train): 1181464049.772978

Lasso Root Mean Squared Error (Train): 34372.43153710511

Lasso R-squared (Test): 0.762599290231023

Lasso Mean Squared Error (Test): 1203101280.7003326

Lasso Root Mean Squared Error (Test): 34685.750398403274

Question-3:

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?

Answer:

Identifying the next five most important predictor variables required re-running the Lasso regression without the initially identified top five predictors. The new important variables are those with the highest coefficients in this revised model. See below for revised ranking.

Top 10 Positive Features - Lasso:

	Feature	Coefficient	Original Feature	Category	Description
76	RoofMatl_WdShngl	95943.537870	RoofMatl	WdShngl	Roof material
36	Neighborhood_NoRidge	89510.353708	Neighborhood	NoRidge	Physical locations within Ames city limits
43	Neighborhood_StoneBr	59091.896173	Neighborhood	StoneBr	Physical locations within Ames city limits
59	HouseStyle_2.5Fin	58518.311284	HouseStyle	2.5Fin	Style of dwelling
120	BsmtQual_Ex	56842.899095	BsmtQual	Ex	Evaluates the height of the basement
152	KitchenQual_Ex	51677.942293	KitchenQual	Ex	Kitchen quality
107	ExterQual_Ex	41342.794979	ExterQual	Ex	Evaluates the quality of the material on the e...
37	Neighborhood_NridgHt	40210.161908	Neighborhood	NridgHt	Physical locations within Ames city limits
27	Neighborhood_Crawfor	29415.565242	Neighborhood	Crawfor	Physical locations within Ames city limits
128	BsmtExposure_Gd	29075.005643	BsmtExposure	Gd	Refers to walkout or garden level walls

Top 10 Positive Features - Lasso (Updated):

	Feature	Coefficient	Original Feature	Category	Description
115	BsmtQual_Ex	57130.229725	BsmtQual	Ex	Evaluates the height of the basement
147	KitchenQual_Ex	52351.436844	KitchenQual	Ex	Kitchen quality
102	ExterQual_Ex	43358.477109	ExterQual	Ex	Evaluates the quality of the material on the e...
167	GarageQual_Ex	31257.564712	GarageQual	Ex	Garage quality
123	BsmtExposure_Gd	28341.024638	BsmtExposure	Gd	Refers to walkout or garden level walls
74	Exterior1st_BrkFace	27771.056314	Exterior1st	BrkFace	Exterior covering on house
133	Heating_GasW	23184.623058	Heating	GasW	Type of heating
104	ExterQual_Gd	21696.298339	ExterQual	Gd	Evaluates the quality of the material on the e...
169	GarageQual_Gd	20869.985584	GarageQual	Gd	Garage quality
76	Exterior1st_CemntBd	17017.422057	Exterior1st	CemntBd	Exterior covering on house

Top 10 Positive Features - Ridge:

	Feature	Coefficient	Original Feature	Category	Description
36	Neighborhood_NoRidge	75542.275975	Neighborhood	NoRidge	Physical locations within Ames city limits
76	RoofMatl_WdShngl	64261.516088	RoofMatl	WdShngl	Roof material
43	Neighborhood_StoneBr	47469.462522	Neighborhood	StoneBr	Physical locations within Ames city limits
59	HouseStyle_2.5Fin	42617.929033	HouseStyle	2.5Fin	Style of dwelling
120	BsmtQual_Ex	39378.473199	BsmtQual	Ex	Evaluates the height of the basement
152	KitchenQual_Ex	37037.518581	KitchenQual	Ex	Kitchen quality
37	Neighborhood_NridgHt	35056.077793	Neighborhood	NridgHt	Physical locations within Ames city limits
107	ExterQual_Ex	25407.496812	ExterQual	Ex	Evaluates the quality of the material on the e...
79	Exterior1st_BrkFace	24515.432533	Exterior1st	BrkFace	Exterior covering on house
128	BsmtExposure_Gd	24459.702003	BsmtExposure	Gd	Refers to walkout or garden level walls

Top 10 Positive Features - Ridge (Updated):

	Feature	Coefficient	Original Feature	Category	Description
115	BsmtQual_Ex	39821.561031	BsmtQual	Ex	Evaluates the height of the basement
147	KitchenQual_Ex	38009.068750	KitchenQual	Ex	Kitchen quality
167	GarageQual_Ex	27535.305401	GarageQual	Ex	Garage quality
102	ExterQual_Ex	27018.415468	ExterQual	Ex	Evaluates the quality of the material on the e...
123	BsmtExposure_Gd	24654.781985	BsmtExposure	Gd	Refers to walkout or garden level walls
49	BldgType_1Fam	24316.073776	BldgType	1Fam	Type of dwelling
74	Exterior1st_BrkFace	23672.675984	Exterior1st	BrkFace	Exterior covering on house
93	Exterior2nd_ImStucc	21195.252284	Exterior2nd	ImStucc	Exterior covering on house (if more than one m...
51	BldgType_Duplex	20659.023331	BldgType	Duplex	Type of dwelling
76	Exterior1st_CemntBd	18776.767853	Exterior1st	CemntBd	Exterior covering on house

Question-4:

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?

Answer:

Ensuring Robustness and Generalizability:

Techniques such as cross-validation, regularization and careful feature engineering were employed in building robust and generalizable models. Simplifying the model to avoid overfitting is also crucial.

In our context, it was deliberate to reduce the selected features to 200 in RFE post one-hot-encoding. Otherwise using full set of features overfits the model.

Implications for Accuracy:

There's a trade-off between a model's complexity and its ability to generalize. Overly complex models may fit the training data well (high accuracy) but perform poorly on new data (low generalizability). Conversely, simpler models may generalize better but at the cost of potentially higher bias.