# Subjective Questions on Surprise Housing Assignment

**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?

Answer:

The Optimal value of lambda for Ridge Regression = 9

The Optimal value of lambda for Lasso = 0.001

Below are the metric changes with the change in alpha values

Changes in Ridge Regression metrics:

- R2 score of train set decreased from 0.92 to 0.91

- R2 score of test set remained same at 0.89

Changes in Lasso metrics:

- R2 score of train set decreased from 0.90 to 0.88

- R2 score of test set decreased from 0.88 to 0.87

The most important predictor variables after the changes are:

As per Ridge model top 10 predictors are:

- OverallQual_9

- OverallQual_8

- Neighborhood_Crawfor

- Neighborhood_StoneBr

- Exterior1st_BrkFace

- Neighborhood_NridgHt

- CentralAir_Y

- Functional_Typ

- GrLivArea

- BsmtCond_TA

As per Lasso model top 10 predictors are:

- OverallQual_9

- GrLivArea

- OverallQual_8

- Neighborhood_Crawfor

- CentralAir_Y

- Functional_Typ

- GarageCars

- Exterior1st_BrkFace

- Neighborhood_NridgHt

- Condition1_Norm

**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?

Answer:

In this case we can go with the lasso model as the difference in r2 score between train and test is less compared to the Ridge model

### Comparision Between Ridge and Lasso Models

- `Ridge` -> r2 score is `0.92(Train), 0.89(Test)` and Root mea
- `Lasso` -> r2 score is `0.90(Train), 0.88(Test)` and Root mea

Also as there are more features it helps in easier feature elimination.

```
# View the features removed by Lasso
lasso_cols_removed = list(betas[betas['Lasso']==0].index)
print(lasso_cols_removed)
```

```
['MasVnrArea', 'BsmtFinSF1', '1stFlrSF', 'OpenPorchSF', 'MSSubClass_40', 'MSSubClass_45', 'MSSubClass_50', 'MSSubClass_75', 'MS
SubClass_80', 'MSSubClass_85', 'MSSubClass_90', 'MSSubClass_120', 'MSSubClass_180', 'MSSubClass_190', 'MSZoning_RH', 'Street_Pa
ve', 'Alley_None', 'LotShape_IR3', 'LotShape_Reg', 'LandContour_Low', 'Utilities_NoSeWa', 'LotConfig_FR2', 'LotConfig_FR3', 'La
ndSlope_Mod', 'LandSlope_Sev', 'Neighborhood_Blueste', 'Neighborhood_BrDale', 'Neighborhood_CollgCr', 'Neighborhood_Gilbert',
'Neighborhood_IDOTRR', 'Neighborhood_Mitchel', 'Neighborhood_NPkVill', 'Neighborhood_NWAmes', 'Neighborhood_OldTown', 'Neighbor
hood_SWISU', 'Neighborhood_Sawyer', 'Neighborhood_SawyerW', 'Neighborhood_Timber', 'Neighborhood_Veenker', 'Condition1_Feedr',
'Condition1_PosA', 'Condition1_PosN', 'Condition1_RRAe', 'Condition1_RRAn', 'Condition1_RRNe', 'Condition1_RRNn', 'Condition2_F
eedr', 'Condition2_Norm', 'Condition2_PosA', 'Condition2_PosN', 'Condition2_RRAe', 'Condition2_RRAn', 'Condition2_RRNn', 'BldgT
ype_2fmCon', 'BldgType_Duplex', 'HouseStyle_1.5Unf', 'HouseStyle_2.5Fin', 'HouseStyle_2.5Unf', 'HouseStyle_SFoyer', 'HouseStyle
_SLvl', 'OverallQual_2', 'OverallQual_6', 'OverallQual_10', 'OverallCond_2', 'OverallCond_6', 'RoofStyle_Gambrel', 'RoofStyle_H
ip', 'RoofStyle_Mansard', 'RoofStyle_Shed', 'RoofMatl_CompShg', 'RoofMatl_Membran', 'RoofMatl_Metal', 'RoofMatl_Roll', 'RoofMat
l_Tar&Grv', 'RoofMatl_WdShake', 'RoofMatl_WdShngl', 'Exterior1st_AsphShn', 'Exterior1st_BrkComm', 'Exterior1st_CBlock', 'Exteri
or1st_CemntBd', 'Exterior1st_HdBoard', 'Exterior1st_ImStucc', 'Exterior1st_Plywood', 'Exterior1st_Stone', 'Exterior1st_Stucco',
'Exterior1st_VinylSd', 'Exterior1st_Wd Sdng', 'Exterior1st_WdShing', 'Exterior2nd_AsphShn', 'Exterior2nd_Brk Cmn', 'Exterior2nd
_BrkFace', 'Exterior2nd_CBlock', 'Exterior2nd_CmentBd', 'Exterior2nd_HdBoard', 'Exterior2nd_ImStucc', 'Exterior2nd_MetalSd', 'E
xterior2nd_Other', 'Exterior2nd_Plywood', 'Exterior2nd_Stone', 'Exterior2nd_Stucco', 'Exterior2nd_VinylSd', 'Exterior2nd_Wd Shn
g', 'MasVnrType_BrkFace', 'MasVnrType_None', 'MasVnrType_Stone', 'ExterQual_Fa', 'ExterQual_Gd', 'ExterCond_Fa', 'ExterCond_G
d', 'ExterCond_Po', 'ExterCond_TA', 'Foundation_CBlock', 'Foundation_Slab', 'Foundation_Stone', 'Foundation_Wood', 'BsmtQual_F
a', 'BsmtCond_None', 'BsmtCond_Po', 'BsmtExposure_Mn', 'BsmtFinType1_BLQ', 'BsmtFinType1_LwQ', 'BsmtFinType1_Rec', 'BsmtFinType
```

It has removed 179 insignificant features.

**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?

Answer:

After removing the top 5 important predictors from lasso model the new 5 important features are:

### After dropping our top 5 lasso predictors, we get the below new top 5 predictors and coefficients

```
- Neighborhood_StoneBr    ,0.120437
- Neighborhood_NridgHt    ,0.108747
- Exterior1st_BrkFace     ,0.106814
- 2ndFlrSF                ,0.093194
- BsmtCond_TA             ,0.085863
```

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

Answer:

We can tell that the model is generalisable when the model performance is not affected when implemented on unseen dataset and it is robust when any change in the input may not affect the performance of the model

We can ensure the below measures:

1. Model should not be over fitted, that is completely learn the train data and give poor performance on unseen data, we can use regularization techniques like ridge and lasso to avoid this
2. Model should we simple, robust and generalisable, we have to ensure the proper tradeoff between bias and variance.