DSCI 35600 – HW 05 (24 points) Due: Friday, 2/15 Name Key

Problem 1 (8 pts). Assume that you are provided with three 2D arrays, named X_train, X_val, and X_test. The contents of these arrays are provided below.

X_train

6	15
8	17
2	23
5	25
10	21
4	16

X_val

5	20
4	18
12	15
7	16

X_test

3	16
1	25
4	12
9	24

note that for the first column of X_train MIN=2 and MAX=10. So the transformed values for X_train will be computed from (X-2)/(10-2). For exampple, when X=6, we get (6-2)/(10-2)=0.5, etc. We use THE SAME MIN and MAX for X_val and X_test.

Complete the tables below to show the contents of the arrays Xs_train, Xs_val, and Xs_test after executing the following code:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
scaler.fit(X_train)
Xs_train = scaler.transform(X_train)
Xs_val = scaler.transform(X_val)
Xs_test = scaler.transform(X_test)
```

X_train

0.5	0
0.75	0.2
0	0.8
0.375	1
1	0.6
0.25	0.1

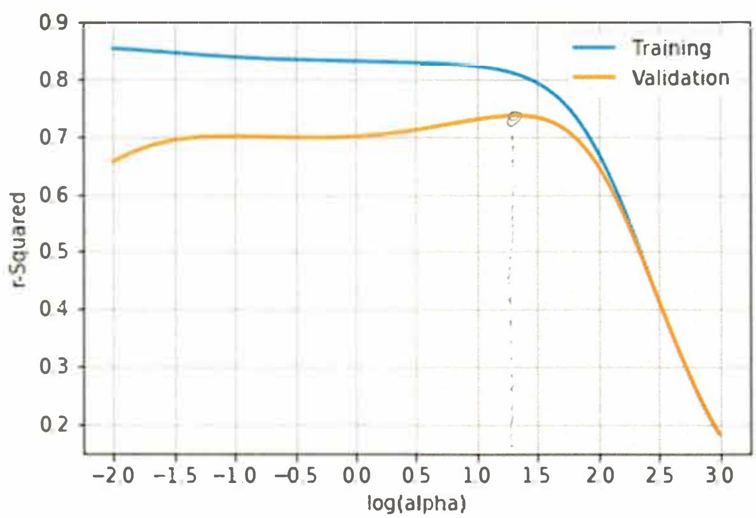
X_val

0.375	0.5
0.25	0.3
1.25	0
0.625	0.1

X_test

0.125	0.1
-0.125	1
0.25	-0.3
0.875	0.9

Problem 2 (4 pts). The plot below shows the training and validation r-squared scores of a ridge regression model for varying values of the rgularization hyperparameter. State the (approximate) value of α that would produce the model with the best abilitity to generalize. Briefly explain why you selected that value.



$\alpha \approx 10^{1.3} = \boxed{19.95}$

This is where the validation r^2 is maximized.

Problem 3 (10 pts). Assume you are provided with a dataset containing 5 observations and three features to use in creating a regression model. The table below contains information for three different proposed models, with each row relating to a different model. The first four columns provide the proposed coefficient estimates for each model. Columns 5 – 9 contain the **predicted** y values resulting from each of the models. The **true** y values are provided in the bottom row of the table.

For each of the three proposed models, calculate its loss using the linear regression loss function, the ridge regression loss function with $\alpha = 0.5$, and the LASSO regression loss function with $\alpha = 0.5$. In each case, use SSE rather than MSE as the "baseline" loss.

After calculating each of the loss values, circle the best loss for each of the three different loss functions. In other words, for each of the last three columns, circle the best loss value that appears in that specific column.

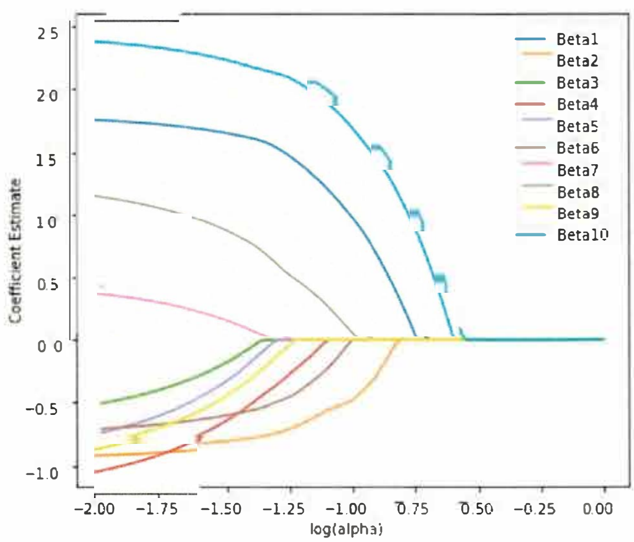
$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$	$\hat{y}_5$	Linear Regression	Ridge Regresssion	LASSO Regression
6	5.6	-3.2	2.4	8.1	12.9	13.8	17.6	18.3	9.22	32.9	14.82
7	3.2	-2.2	1.8	9.5	9.8	16.4	16.4	18.1	11.07	20.23	14.67
5	4.4	-0.2	0.8	11.8	11.7	16.5	15.2	17.6	11.63	21.65	14.33
True y Values →				9.8	12.1	14.6	15.7	19.5			

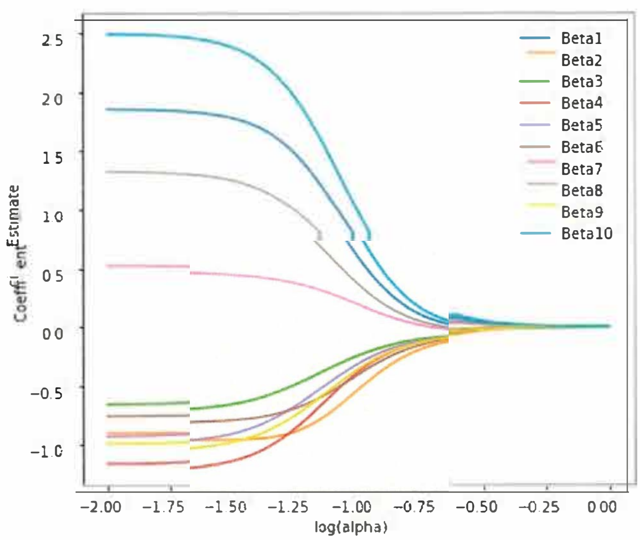
For example: Linear Regression Loss = SSE and for the first row is
(8.1-9.8)^2 + (12.9-12.1)^2+(13.8-14.6)^2+(17.6-15.7)^2+(18.3-19.5)^2 = 9.22

Ridge Regression Loss = SSE + alpha (beta1^2 +beta2^2 +beta3^2) =
9.22 + 0.5*(5.6^2 + 3.2^2 + 2.4^2) = 32.9

Lasso Regression Loss = SSE + alpha (|beta1| +|beta2| +|beta3|) =
9.22 + 0.5*(5.6 + 3.2 + 2.4) = 14.82

Problem 4 (2 pts). A dataset with 10 features is to be used for creating a regression model. Several ridge and LASSO models are fit to the data, each with varying values of the regularization parameter. The two plots below show how the model coefficients change with respect to the regularization parameter for each of the two model types. Identify which plot was generated using ridge models, and which was generated using LASSO models. Provide your answer by writing "Ridge" or "LASSO" below the plots.





LASSO

Ridge