Regularization

Statistical Prediction Modeling *DATA 2204*

Jumana Obeid

100832624

Assignment #4
24th of March 2022

DURHAM COLLEGE
SUCCESS MATTERS

b. Rational Statement (summary of the problem or problems to be addressed by the PPT) – 2%

In linear regression, overfitting can be a problem when both the prediction performance and generalization power and not good.We need to add some noise in order to better predict our results when new data is tested.

In other words, the regularized machine learning model, is a model that its loss function contains another element that should be eliminated as well.In our model:

Independent Variables:
X1 - Relative Compactness
X2 - Surface Area X3 - Wall Area
X4 - Roof Area X5 - Overall Height
X6 - Orientation X7 - Glazing Area
X8 - Glazing Area Distribution
Dependent Variable:
Y- Cooling Load

c. Present and explain three (3) key insights from the key metrics (Adj. R2 , MAE, RMSE) from each of the Optimized Regularization models (i.e. LASSO, Ridge, and Elastic Nets), but first use Tukey to remove any outliers. Note: nine (9) insights in total are required – 10% **Let us first see Linear regression:**

Linear Regression

```
Intercept: 24.57

Model coefficients:
X1 -7.36
X2 -3.66
X3 0.03
X4 -3.58
X5 7.99
X6 0.21
X7 2.01
X8 0.06

R2: 0.86
Adj_R2: 0.86

Mean Absolute Error: 2.46
Mean Squared Error: 11.51
Root Mean Squared Error: 3.39
```

• Adj_R2: .86 states that 86% of the variance of the dependent variable being studied is explained by the variance of the independent variable.

• MAE: 2.46 which is the magnitude of difference between the prediction of an observation and the true value of that observation.Lower is better.

• RMSE 3.39is not very close to 10% of the output mean. We can say we could have a better model.

Lasso

```
Intercept: 24.57

Model coefficients:
X1 5.09
X2 -0.77
X3 4.71
X4 -0.98
X5 1.29
X6 0.11
X7 1.98
X8 0.0

R2: 0.79
Adj_R2: 0.79

Mean Absolute Error: 2.95
Mean Squared Error: 17.36
Root Mean Squared Error: 4.17
```

- Adj_R2: .79 states that 79% of the variance of the dependent variable being studied is explained by the variance of the independent variable.

- MAE: 2.95 which is the magnitude of difference between the prediction of an observation and the true value of that observation.Lower is better.

- RMSE 4.17is not very close to 10% of the output mean. We can say we could have a better model.

Ridge

```
Intercept: 24.57

Model coefficients:
X1 -7.2
X2 -3.56
X3 0.05
X4 -3.49
X5 8.03
X6 0.21
X7 2.01
X8 0.06

R2: 0.86
Adj_R2: 0.86

Mean Absolute Error: 2.46
Mean Squared Error: 11.51
Root Mean Squared Error: 3.39
```

- Adj_R2: .86 states that 86% of the variance of the dependent variable being studied is explained by the variance of the independent variable.

- MAE: 2.46 which is the magnitude of difference between the prediction of an observation and the true value of that observation.Lower is better.

- RMSE 3.39is not very close to 10% of the output mean. We can say we could have a better model.So far , this is the best results.

Elastic Nets

```
Intercept: 24.57

Model coefficients:
X1 4.57
X2 -1.19
X3 4.5
X4 -1.11
X5 1.29
X6 0.15
X7 1.92
X8 0.06

R2: 0.80
Adj_R2: 0.80

Mean Absolute Error: 2.87
Mean Squared Error: 16.97
Root Mean Squared Error: 4.12
```

- Adj_R2: .80 states that 80% of the variance of the dependent variable being studied is explained by the variance of the independent variable.

- MAE: 2.87 which is the magnitude of difference between the prediction of an observation and the true value of that observation.Lower is better

- RMSE 4.12 is not very close to 10% of the output mean. We can say we could have a better model.

d. State and explain three (3) recommendations for Mr. John Hughes for next steps. – 3% Attention: Please ensure that all key facts are in your slides and not in the notes section.

We recommend the following:

- Apply a feature selection technique to select important features only.
- Apply a rule (AIC or BIC-penalized measure to fit.for deciding which variables to include -unless a variable improves AIC or BIC don’t include it.
- Use Boruta, which is a method to find all features carrying information usable for prediction,rather than relying on error factor.