Statistical Consulting

Concrete compressive strength prediction

Shao-Ning, Chen

Department of Statistics, National Cheng Kung University

May 31, 2022

1 Recap

2 Solved Challenges

3 Methods

4 Experimental Result

5 Conclusion

1 Recap

2 Solved Challenges

3 Methods

4 Experimental Result

5 Conclusion

Recap

- The dataset comes from UCI Concrete Compressive Strength Data Set [1]
- Number of instances (observations): 1030
- Attribute breakdown: 8 quantitative input variables, and 1 quantitative output variable
- Missing Attribute Values: None

	Cement	Slag	Fly_Ash	Water	Superplastic	Coarse_Aggr	Fine_Aggr	Age	CCStr
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.30

Variable Information

Name	Unit	Description	Dtype
Cement	kg/m3	Cement	float
Blast Furnace Slag	kg/m3	Metal oxides and SiO2 mix	float
Fly Ash	kg/m3	Coal combustion product	float
Water	kg/m3	Water	float
Superplasticizer	kg/m3	Making high-strength concrete	float
Coarse Aggregate	kg/m3	Larger than 4.75mm aggregate	float
Fine Aggregate	kg/m3	Small than 4.75mm aggregate	float
Age	Day	Age	int
CCStrength	MPa	Output Variable	float

Recap

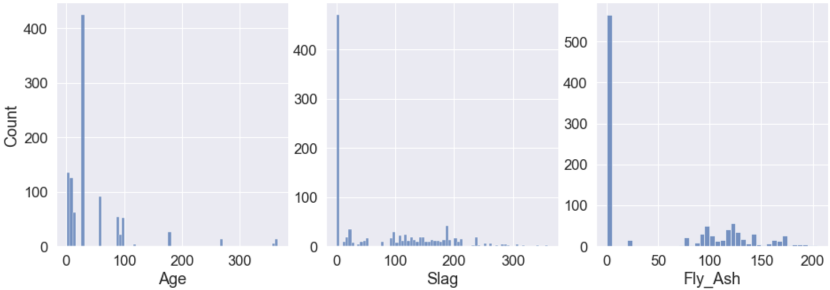
- The following table are the descriptive statistics of the dataset
- There is no missing value and outlier on this dataset

	Cement	Slag	Fly_Ash	Water	Superplastic	Coarse_Aggr	Fine_Aggr	Age	CCStr
count	1030.00	1030.00	1030.00	1030.00	1030.00	1030.00	1030.00	1030.00	1030.00
mean	281.17	73.90	54.19	181.57	6.20	972.92	773.58	45.66	35.82
std	104.51	86.28	64.00	21.35	5.97	77.75	80.18	63.17	16.71
min	102.00	0.00	0.00	121.80	0.00	801.00	594.00	1.00	2.33
25%	192.38	0.00	0.00	164.90	0.00	932.00	730.95	7.00	23.71
50%	272.90	22.00	0.00	185.00	6.40	968.00	779.50	28.00	34.44
75%	350.00	142.95	118.30	192.00	10.20	1029.40	824.00	56.00	46.14
max	540.00	359.40	200.10	247.00	32.20	1145.00	992.60	365.00	82.60

1 Recap
2 Solved Challenges
3 Methods
4 Experimental Result
5 Conclusion

Problem

- Q : Some of variables are imbalanced



Problem Solved

- Ans : Data binning
- Some variables has a few extreme values. To mitigate the bias in this dataset, I using the quantiles method to transform the data

	Slag	Fly_Ash	Age
0	0.0	0.0	28
1	0.0	0.0	28
2	142.5	0.0	270
3	142.5	0.0	365
4	132.4	0.0	360
5	114.0	0.0	90
6	95.0	0.0	365

	Slag	Fly_Ash	Age
0	1	1	2
1	1	1	2
2	3	1	5
3	3	1	5
4	3	1	5
5	2	1	4
6	2	1	5

1 Recap

2 Solved Challenges

3 Methods

4 Experimental Result

5 Conclusion

Methods

- Linear regression and its variant [2]
- Machine learning without tree based methods [2]
- Machine learning with tree based methods
- Training and testing on different testing ratio
- Training and testing on different regression methods
- Compare the result with/without data binning

Methods

- Linear Regression*
- Lasso Regression*
- Ridge Regression*
- ElasticNet Regression
- Decision Tree Regression*
- KNN Regression
- MLP Regression
- SVM Regression
- LightGBM Regression
- XGBoost Regression
- CatBoost Regression
- Random Forest Regression*

1 Recap

2 Solved Challenges

3 Methods

4 Experimental Result

5 Conclusion

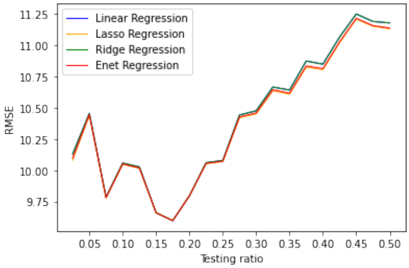
Linear Regression and Variant

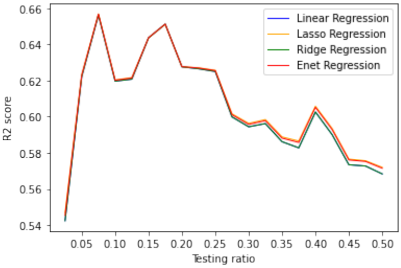
Methods

- Linear Regression*
- Lasso Regression*
- Ridge Regression*
- ElasticNet Regression

Linear Regression and Variant

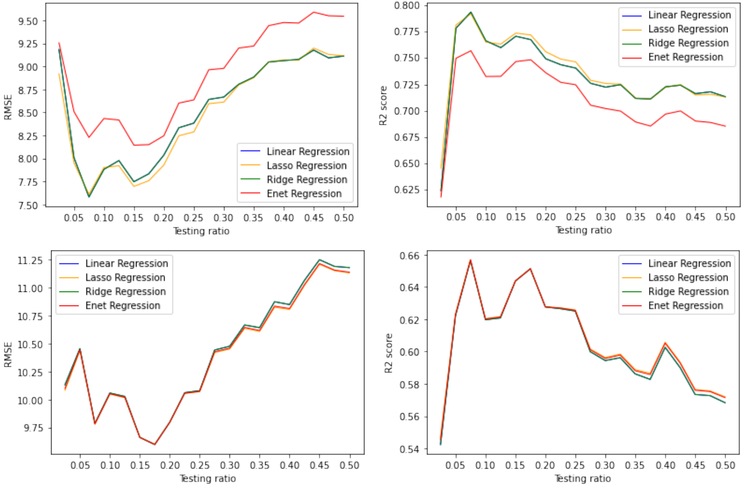
- The test size set between 0.025 to 0.5, and compare the performance between different testing ratio
- The following plots are the performance without data binning





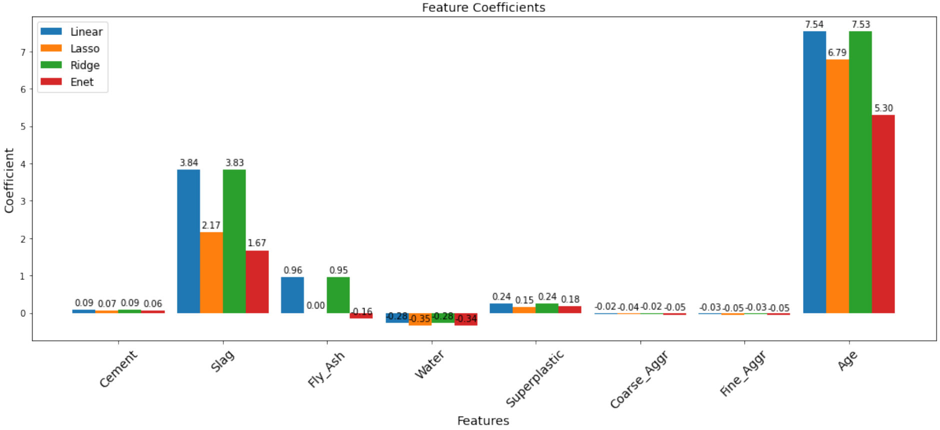
Linear Regression and Variant

- Performance with(upper)/without(lower) data binning



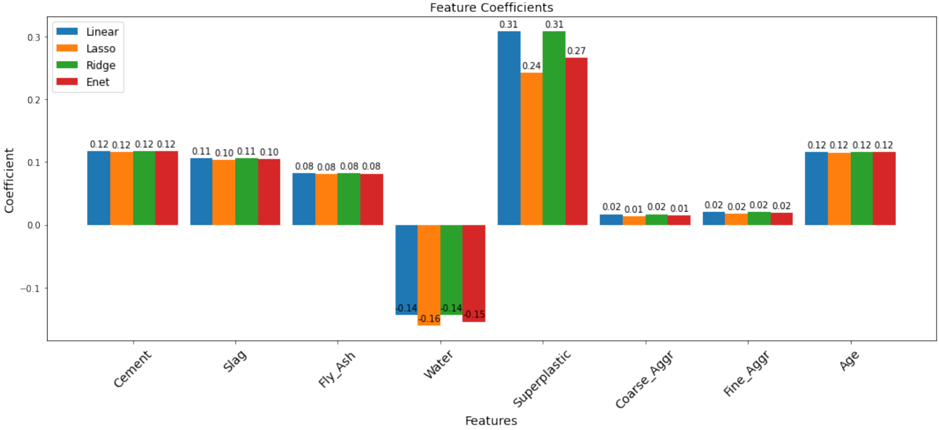
Linear Regression and Variant

- Coefficients with binning dataset (testing ratio=0.15)



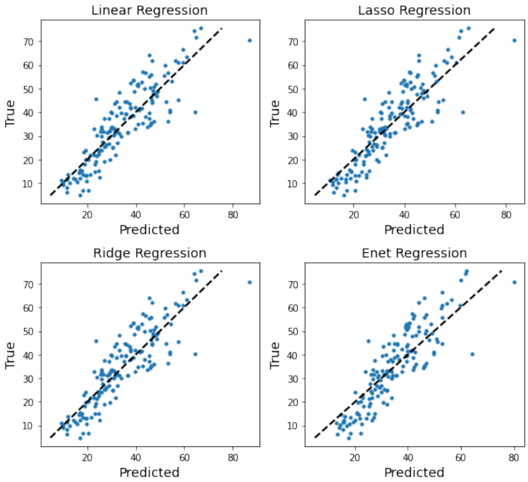
Linear Regression and Variant

- Coefficients without binning dataset (testing ratio=0.15)



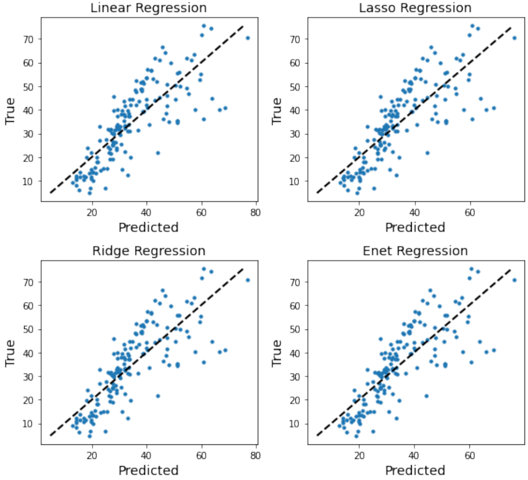
Linear Regression and Variant

- Predicted value vs True value (with data binning, test=0.15)



Linear Regression and Variant

- Predicted value vs True value (without binning, test=0.15)



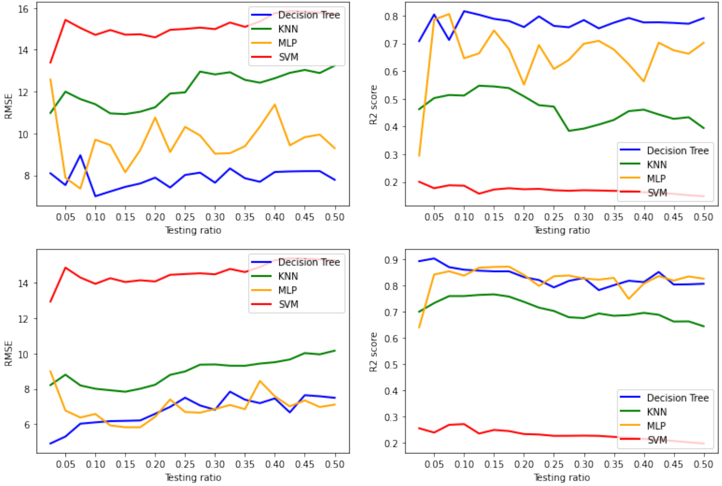
Machine Learning Methods

Methods

- Decision Trees Regression*
- KNN Regression
- MLP Regression
- SVM Regression

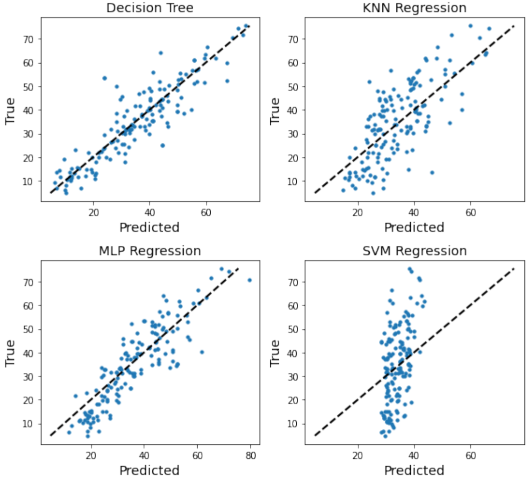
Machine Learning Methods

- Performance with(upper)/without(lower) data binning



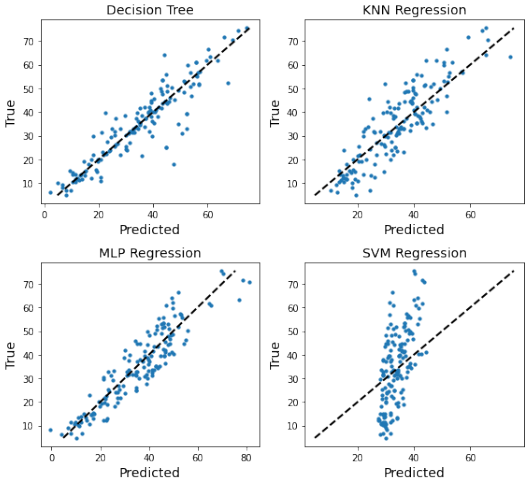
Machine Learning Methods

- Predicted value vs True value (with data binning, test=0.15)



Machine Learning Methods

- Predicted value vs True value (without binning, test=0.15)



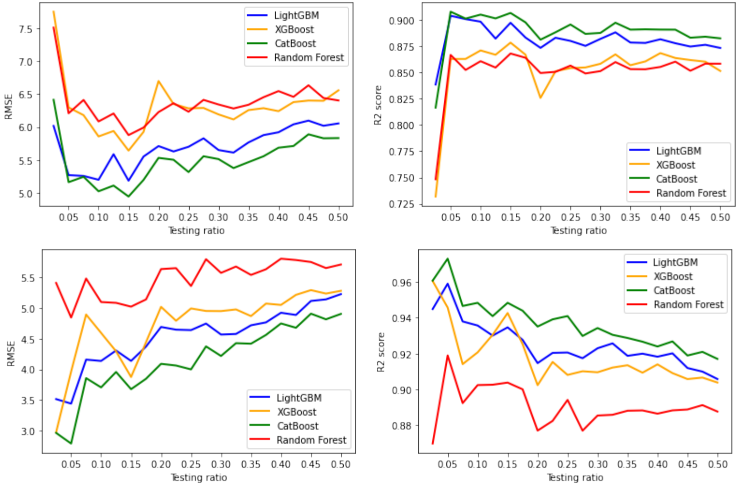
Tree based Methods

Methods

- LightGBM Regression
- XGBoost Regression
- CatBoost Regression
- Random Forests Regression*

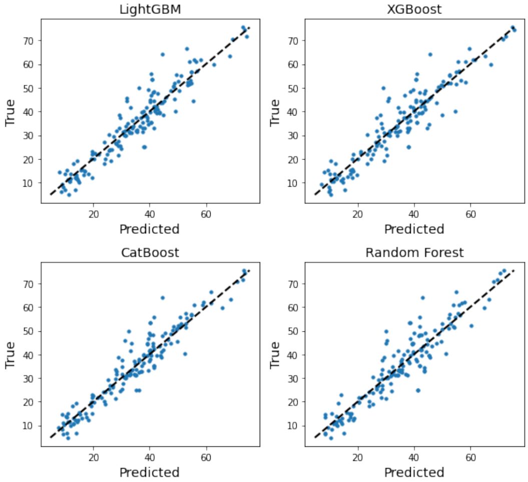
Tree based Methods

- Performance with(upper)/without(lower) data binning



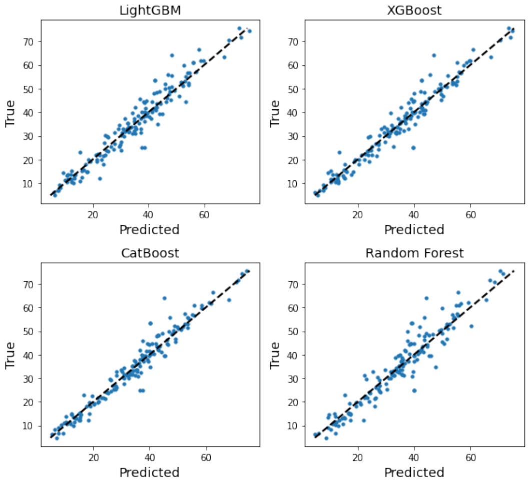
Tree based Methods

- Predicted value vs True value (with data binning, test=0.15)



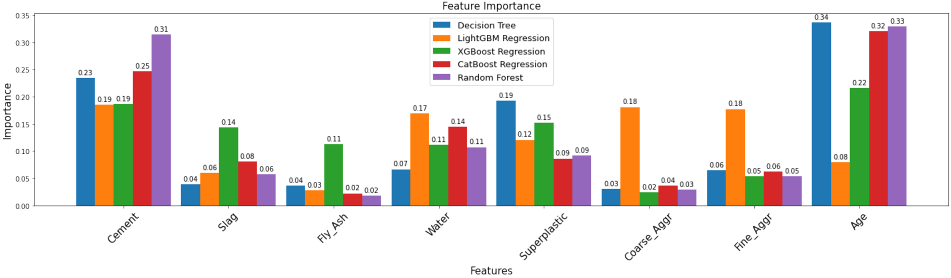
Tree based Methods

- Predicted value vs True value (without binning, test=0.15)



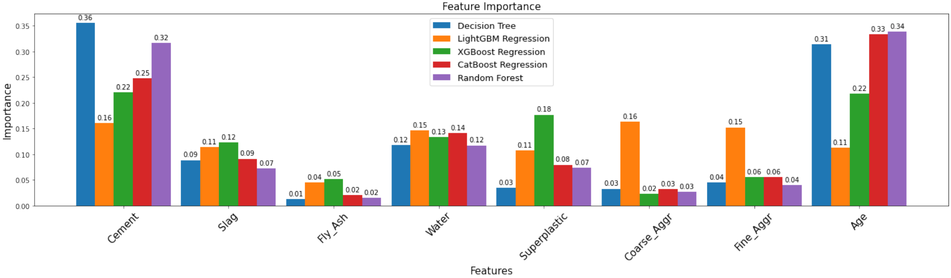
Tree based Methods

- Feature importance of tree based methods (with data binning, test=0.15)



Tree based Methods

- Feature importance of tree based methods (without data binning, test=0.15)



1 Recap

2 Solved Challenges

3 Methods

4 Experimental Result

5 Conclusion

Conclusion

- The lower RMSE occur between testing ratio of 0.05 and 0.2
- In linear regression and its variant, data binning can improve performance
- In machine learning or Tree based methods, data binning can not improve performance
- In Tree baesd methods, as the testing ratio growth, the RMSE without data binning increase rapidly than the dataset with data binning
- Performance : Tree based methods $>$ Linear Regression and variant $\geq$ Machine Learning without tree based methods
- Best model : CatBoost

Reference

[1] I.-C. Yeh. "Modeling of strength of high-performance concrete using artificial neural networks". In: *Cement and Concrete Research* 28.12 (1998), pp. 1797–1808. ISSN: 0008-8846. DOI: https://doi.org/10.1016/S0008-8846(98)00165-3. URL: https://www.sciencedirect.com/science/article/pii/S0008884698001653.

[2] Ahsanul Kabir, Monjurul Hasan, and Md Khasro Miah. "Strength prediction model for concrete". In: *International Journal on Civil and Environmental Engineering* 2.1 (2013), p. 14.

Thank you!