Proposal
Prediction on the Concrete Compressive Strength Using Supervised Learning

Yinyin Lu
June 2015

Background and Motivation

The high performance concrete is a highly complex material that consists of cement, water, blast furnace slag, super-plasticizer, coarse and fine aggregate. All of the materials play a certain role in the compressive strength of concrete. Combined with the information about age, the compressive strength of concrete is determined by eight attributes:

1. Cement (kg/m^3)
2. Fly ash (kg/m^3)
3. Blast furnace slag (kg/m^3)
4. Water (kg/m^3)
5. Super-plasticizer (kg/m^3)
6. Coarse aggregate (kg/m^3)
7. Fine aggregate (kg/m^3)
8. Age of testing (days)

The objective of this study is to predict the compressive strength of high performance concrete under supervised learning. Several methods will be discussed and compared, including Linear Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Shrinkage method, Random Forest, K Nearest Neighbor, Support Vector Machine, and Neural Network, etc. Cross validation, forward selection, and backward selection will be used to assess different models. Both of the interpretability and bias will be considered. The model with lowest prediction error rate will be chosen to predict the strength. Also, certain data transformation might be used during the stage of data pre-processing.

In this study, three questions will be answered:
1). What is the best model to the prediction? (Choosing criteria)
2). What is the most predictive parameter in model?
3). How does model with high interpretability (eg. Linear regression) compare to the more complex models?

Data Description

The response variable in this data set given by UCI machine learning repository is the compressive strength of high performance concrete. There are eight explanatory variables that were shown above and 1030 observations. Table 1 includes the information for each feature:

Table 1. Features summary

	Cement	Slag	Ash	Water	Superplasticizer
Min	102	0	0	121.8	0
1st Qu	192.4	0	0	164.9	0
Median	272.9	22	0	185	6.4
Mean	281.2	73.9	54.19	181.6	6.205
3rd Qu	350	142.9	118.3	192	10.2
Max	540	359.4	200.1	247	32.2
	Coarse	**Fine**	**Age**	**Strength**	
Min	801	594	1	2.33	
1st Qu	932	731	7	23.71	
Median	968	779.5	28	34.45	
Mean	972.9	773.6	45.66	35.82	
3rd Qu	1029.4	824	56	46.13	
Max	1145	992.6	365	82.6	

Reference

1. I-Cheng Yeh, "Modeling of strength of high performance concrete using artificial neural networks," Cement and Concrete Research, Vol. 28, No. 12, pp. 1797-1808 (1998).

2. I-Cheng Yeh, "Modeling Concrete Strength with Augment-Neuron Networks," J. of Materials in Civil Engineering, ASCE, Vol. 10, No. 4, pp. 263-268 (1998).