Composition of concrete and its influence on compressive strength

Filipe P. de Farias
Department of Teleinformatics Engineering
Federal University of Ceará
Fortaleza, Brazil
filipepfarias@fisica.ufc.br

Yvo J. M. Sales
Department of Teleinformatics Engineering
Federal University of Ceará
Fortaleza, Brazil
yvo@gtel.ufc.br

***Abstract*—The compressive strength of concrete impacts directly on its application. The difference between the concrete for columns or beams and the concrete for pavements is mainly due the compressive strength it is able to resist. In this work, we perform an unconditional and a class-conditional mono-variate analysis as well as unconditional bi-variate and multi-variate analysis of the UCI concrete composition database.**

***Index Terms*—concrete, compresive, strength, machine, learning, pre-processing**

I. Introduction

A material formed by aggregates bonded together by a fluid material that hardens over time has been used by humans for construction since many years ago [3]. Nowadays this material is known as concrete and it's widely used in the construction field. The aggregates used in the concrete affect directly its compressive strength which highly impacts its applications. For instance, in general, the concrete for columns or beams needs to have a greater compressive strength than the one for pavement. On this paper, we carry out an statistical analysis on a dataset extracted from the UCI Machine Learning Repository (University of California, Irvine) [4] that collects information about the concentration of some aggregates used to form different types of concretes and their resulting compressive strenght.

This work is divided as follows. A description of the data is given in Section II. The Section **??** brings an unconditional and a class-conditional mono-variate analysis as well as unconditional bi-variate and multi-variate analysis of the data. Finally, the conclusions and considerations are exposed in Section V.

II. Data description

The composition of each of the N concrete samples is given by the concentrations (kg/m^3) of D components: Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate and Fine Aggregate. Each sample has its Age (day) and the measured Concrete compressive strength (MPa), as described in Table I.

The concrete was stratified into $L = 3$ classes [1]. The concrete which is weak and not recommended for structures, the *Non-standard*, was labeled with L_1 and has 295 samples in it. The concrete whose strength is in a range that can be applied to structures is classified as L_2, or *Standard*, and has 525 samples. The high performance concrete L_3, *High-strength* and has 210 samples. The observations are the measured compressive strengths of each sample and, as the predictors $D_1 - D_7$, are continuous. The Age (D_8) of the concrete is extremely discrete. The output is the concrete strength

TABLE I
Data description

Component	Description	Unit
D_1	Cement	kg/m^3
D_2	Blast Furnace Slag	kg/m^3
D_3	Fly Ash	kg/m^3
D_4	Water	kg/m^3
D_5	Superplasticizer	kg/m^3
D_6	Coarse Aggregate	kg/m^3
D_7	Fine Aggregate	kg/m^3
D_8	Age	days
Y	Compressive strength	MPa
N	1030 samples	

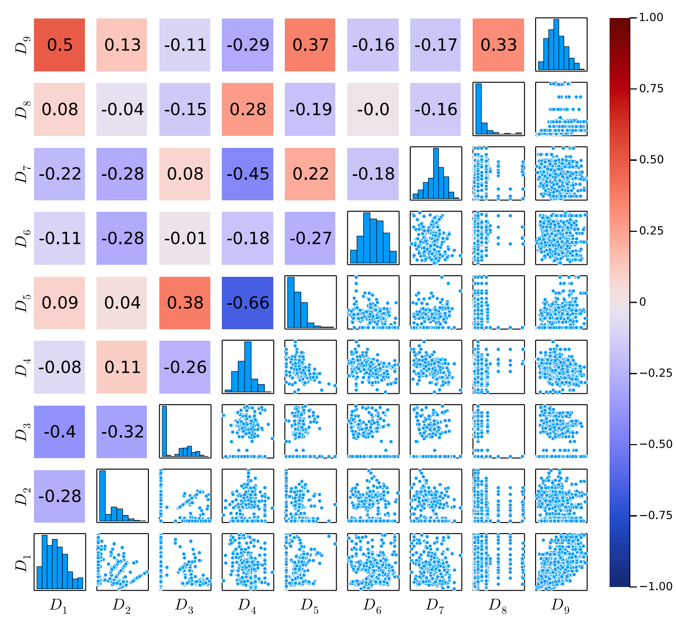

Fig. 1. Pairwise scatter plot of each concrete component.

III. Methods

Regression models try to find relations between the *independent variables* and the *dependent variables*, which are named, respectively, predictors and outcomes in this work. These relations can occur in different forms. The simplest one is the linear relationship, which is when the curve predictors *vs* outcomes, in the case that both are one-dimemsional, forms a simple line and, in the general case, a hyperplane. In what folows, we formluate the *Linear regression*, which is a subclass of regressions dedicated to find a linear model to explain the relation between the predictors and the outcomes.

A. Linear regression

This method tries to find the linear regression between the predictors and outcomes by fitting a line with linear coefficient β_0 and angular coefficient β_1, defined in Eq. (1), through the data. The fitting is done by minimising a cost function that can take different forms. Each cost function yields to different optimal parameters and two of them are described in the following, the *Ordinary least squares* and *L_2-penalized least squares*. Regard the fact that β_1 can be a vector of the size such as the number of predictors in X. An interesting fact to observe is when the predictors correlates with the outcome, we can observe the error will be small. This is because the correlation evaluates linear variations between the variables as such as the linear regression.

$$Y = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix} \begin{bmatrix} 1 \\ X \end{bmatrix} + \varepsilon \qquad (1)$$

The *Ordinary least squares* defines the cost function to find the optimal parameters for Eq. (1) as

$$L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2, \qquad (2)$$

where $\mathbf{y} = [y_0, ..., y_p)]^T$ and $\boldsymbol{\beta} = [\beta_0, ..., \beta_p)]^T$.

The *L_2-penalized least squares* modifies Eq. (2) adding a term that penalize large values of the parameters, yielding to

$$L(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2, \qquad (3)$$

where $\mathbf{y} = [y_0, ..., y_p)]^T$, $\boldsymbol{\beta} = [\beta_0, ..., \beta_p)]^T$ and λ is the penalization coefficient, which is a tuning parameter.

Partial least squares

IV. Results

V. Conclusion

The database of concrete is not easy to analyse if there is no previous knowledge about the problem of the components mixture. In none of the analysis the data have been shown as separable on the initially determined classes. The next step is to try to perform regression to model the compressive strength itself before trying to classify the samples.

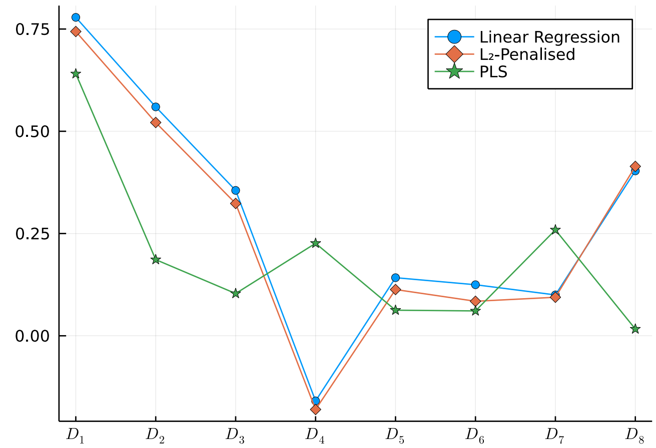

Fig. 2. Values of the weights of each predictor obtained after 30%/70% cross-validation strategy.

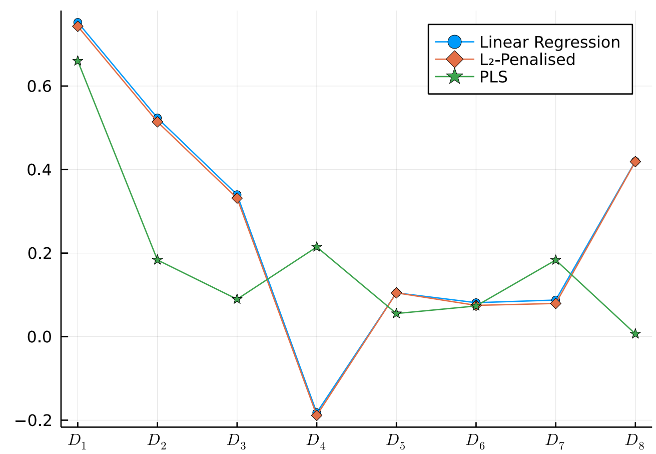

Fig. 3. Values of the weights of each predictor obtained after 5-fold cross-validation strategy.

TABLE II
OLS Linear Regression

CV	RMSE	R^2
70% Train / 30% Test	0.614014	0.612271
1-th fold	0.614187	0.634496
2-th fold	0.647336	0.597003
3-th fold	0.5952	0.599484
4-th fold	0.686377	0.530463
5-th fold	**0.582266**	**0.661407**

TABLE III
Ridge Regression

CV	RMSE	R^2
70% Train / 30% Test	0.599004	0.640309
1-th fold	0.613885	0.634857
2-th fold	0.647344	0.596994
3-th fold	0.594722	0.600127
4-th fold	0.68599	0.530993
5-th fold	**0.582357**	**0.661302**

TABLE IV
PLS REGRESSION

CV	RMSE	R^2
70% Train / 30% Test	0.599264	0.639996
1-th fold	0.614106	0.634593
2-th fold	0.646593	0.597928
3-th fold	0.594709	0.600143
4-th fold	0.6859	0.531116
5-th fold	**0.58221**	**0.661472**

REFERENCES

[1] ACI Manual of Concrete Practice 2000, Part 1: Materials and General Properties of Concrete. American Concrete Institute. Farmington Hills, MI.

[2] Tibshirani, Robert, et al. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Alemanha, Springer New York, 2009.

[3] Mindess, S., and Young, J.F. Concrete. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.

[4] 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)