**Predicting Concrete Compressive Strength using Linear Regression**

Oswaldo David García Rodríguez

Tecnológico de Monterrey, Campus Querétaro

April 9, 2020

*Abstract: This document presents the training process of a linear regression algorithm to estimate concrete compressive strength using a dataset of 1030 examples*

**INTRODUCCION**

During construction, it is important to choose the most appropriate type of concrete that ensures the strength of the structures. The most used attribute is the concrete compressive strength, which depends on the type of mixture.

According to the article “Testing Compressive Strength of Concrete”, this is measured using a compression-test machine that tries to break a cylindrical concrete sample. Then it is calculated the failure load divided by the cross-sectional area resisting the load. The result is represented in pound-force per square inch (psi) or megapascals (MPa).

Figure 1. Example of a compression test machine

The use of a linear regression algorithm will try to make an estimation of the compressive strength of concrete given the concentrations for each material that makes up the mixture.

**DATASET**

It contains 1030 samples of concrete represented by 8 quantitative values and the compressive strength given in MPa.

The following attributes represents the components of the mixture, represented in kg per cubic meter:

1. Cement
2. Blast furnace slag
3. Fly ash
4. Water
5. Superplasticizer
6. Coarse aggregate
7. Fine aggregate

The last attribute is the age of the concrete, expressed in days.

**LINEAR REGRESSION**

This is a model which attempts to make a relationship between a set of inputs (xi) and an output (y) using a linear equation.

The formula of a linear equation with multiple variables is represented as:

where *n* represents the number of attributes, *mi* is the scope for each *xi,xi* is the attribute and *b* (bias) is a constant.

In linear regression, the hypothesis function is the way to make an approximation to the observed data (y) through calculating the scopes for all variables.

where *θi* are the coefficients and represents the scopes of the *x*i’s. It’s important to mention that *θ0*is the bias and *x0* it’s always 1.

The first time, the values of coefficients can be random as they will be updated with each iteration to be closer to observed data. The difference between values obtained with hypothesis function and real values are calculated using the mean squared error (MSE) function:

where *m* is the number of samples, *h0(xi)* is the hypothesis function applied to a sample and *yi* is the true output of that sample.

SME won’t affect the calculations of new coefficients, but it’s useful to observe how the error changes with each iteration. If it is increasing, then it would mean the linear regression model should be modified.

To update the coefficients, each one changes using the cost function with respect to the corresponding coefficient multiplied with a learning rate *α*:

**REFERENCES**

Jamal, H. (2017). *Procedure for Concrete Compression Test.* Retrieved from: <https://www.aboutcivil.org/method-process-compression-test.html>

NRMCA. (2014). *CIP 35 – Testing Compressive Strength of Concrete.* Retrieved from: <https://www.nrmca.org/aboutconcrete/cips/35pr.pdf>

UCI. (2007). *Concrete Compressive Strength Data Set*. Retrieved from: <https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>