

上海大学 2019 — 2020 学年 冬季学期

课程名称: 机器学习与力学

课 程 号: 01826249

授课教师: 胡国辉

论文名称: 混凝土抗压强度线性模型

学 号: 17121632

姓 名: 翟晗锋

专 业: 理论与应用力学

成 绩: _____

日 期: 2020/2/10

评 语:

混凝土抗压强度的线性模型

翟晗锋

(上海大学力学与工程科学学院, 上海 200444)

摘要

混凝土为一类基础且重要的基础型抗压性材料, 其被广泛应用于桥梁建筑等土木工程领域。在人类制备及应用混凝土的长久历史中, 混凝土的抗压性能与大量已知或潜在变量均存在直接或间接关系。因此, 通过实验的方式及数学推导的传统方式很难给出混凝土抗压能力与各种自变量的关系及模型。所以, 利用机器学习中线性回归算法, 我们可以通过大量的数据及叠加计算拟合出混凝土抗压极限与大量已有记录数据的关系, 并得到相应线性模型。本文中, 作者通过收集混凝土粘合度、高炉渣含量、吹入灰尘量、水含量等八个系数与混凝土抗压极限的 1030 组数据, 拟合计算线性模型。通过对获得的线性模型进行讨论, 我们可以基本分析出在已测量数据中不同自变量与混凝土抗压强度的关系, 与此分析出在混凝土制备过程中控制条件及对于制备混凝土的力学性能进行可能定型预估。

关键词: 机器学习, 混凝土, 抗压强度, 线性模型

1 引言

在人类长久的建筑与工程学的历史中, 混凝土扮演了极其重要的地位。在 5000 多年前, 考古学家就发现古人有类似烧至混凝土的技艺来建造房屋; 且直至今日, 混凝土仍然作为一种主要建筑材料服务和造福人类的生产生活。其低廉的成本和优异的抗压性能受到了人们的青睐。作为一种抗压材料, 对我们力学学科学生来讲起抗压性能与造成该性能的原因是我们好奇且想要分析的。

通过课堂上对于线性模型的学习, 我们学习到了如何利用数据拟合得到线性模型, 尤其是通过多元线性回归的方法我们可以拟合出多种自变量与一种因变量之间, 即是混凝土力学系数及抗压强度的关系。利用这种方法, 我们将其应用于混凝土制备的数据中, 便可以得到混凝土抗压强度与其各个参数之间的线性模型

2 背景概况

2.1 混凝土

混凝土的广义概念被定义为以水泥为主要的胶凝材料，总体可理解为水、砂、石子，以及合成时掺入化学外加剂和其他矿物原料，用适当比例配备，再通过均匀的搅拌和密实成型及养护硬化而形成的人造抗压材料。混凝土制造主要分为两个阶段与状态：凝结硬化前的塑性状态，即新拌混凝土或混凝土拌合物；硬化之后的坚硬状态，即硬化混凝土或本文中提到的混凝土。混凝土强度等级是以立方体抗压强度标准值来定义或划分，在中国普通混凝土强度等级划分为 14 级：C15、C20、C25、C30、C35、C40、C45、C50、C55、C60、C65、C70、C75 及 C80。

2.2 抗压强度

抗压强度是材料或结构承受趋于减小尺寸的载荷的能力。换句话说，抗压强度抵抗压缩（抵抗物体挤压）。在材料强度研究中，可以独立地分析抗张强度，抗压强度和剪切强度。

一些材料在其抗压强度极限时断裂或破坏；其他变形不可逆地变形，因此可以将给定的变形量视为压缩载荷的极限。抗压强度是结构设计的重要指标。

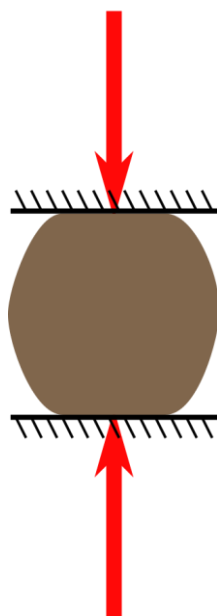


图5 物体抗压强度测量方法。

Fig.5. Method to test materials' compressive strength.

3 方法

3.1 数学模型

首先，对于一组具有 d 个属性（自变量）的数据集，其中，我们将这些属性总体表达为 $\mathbf{x} = (x_1; x_2; \dots; x_d)$ 。其中 x_i 是 \mathbf{x} 在第 i 个属性取值。我们可以通过建立线性模型学得一个线性模型进行函数预测

$$f(\mathbf{x}) = \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_d x_d + b$$

对具有 d 个属性的数据集 D 的多元线性回归问题，我们用一般的向量形式表达

$$f(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x}_i + b, \text{ 使 } f(\mathbf{x}_i) \cong y_i$$

同时，我们可以用最小二乘法来对 $\boldsymbol{\omega}$ 和 b 进行估计。为方便表示，引入向量形式 $\hat{\boldsymbol{\omega}} = (\boldsymbol{\omega}; b)$ 。因此，即可把数据集 D 表示为一个 $m \times (d + 1)$ 大小的矩阵 \mathbf{X} ，其中每行对应一个示例（数据集中对应的行），该行前 d 个元素对应实例中 d 个属性值，最后一行元素恒为 1。

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} & 1 \\ x_{21} & x_{22} & \dots & x_{2d} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{md} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T & 1 \\ \mathbf{x}_2^T & 1 \\ \dots & \dots \\ \mathbf{x}_m^T & 1 \end{pmatrix}$$

再把标记也写成向量形式 $\mathbf{y} = (y_1; y_2; \dots; y_m)$

$$\hat{\boldsymbol{\omega}}^* = \underset{\hat{\boldsymbol{\omega}}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\omega}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\omega}})$$

令 $E_{\hat{\boldsymbol{\omega}}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\omega}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\omega}})$ ，通过对 $\hat{\boldsymbol{\omega}}$ 求导得到

$$\frac{\partial E_{\hat{\boldsymbol{\omega}}}}{\partial \hat{\boldsymbol{\omega}}} = 2\mathbf{X}^T (\mathbf{X}\hat{\boldsymbol{\omega}} - \mathbf{y})$$

令上式为零即得到 $\hat{\boldsymbol{\omega}}$ 最优解的闭式解。当 $\mathbf{X}^T \mathbf{X}$ 为满秩矩阵和正定矩阵时，令 $\frac{\partial E_{\hat{\boldsymbol{\omega}}}}{\partial \hat{\boldsymbol{\omega}}} = 0$ 可得：

$$\hat{\boldsymbol{\omega}}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

其中 $(\mathbf{X}^T \mathbf{X})^{-1}$ 为 $\mathbf{X}^T \mathbf{X}$ 逆矩阵。令 $\hat{\mathbf{x}}_i = (\mathbf{x}_i; 1)$ ，则最终学得多元线性回归模型

$$f(\hat{\mathbf{x}}_i) = \hat{\mathbf{x}}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

3.2 变量分析

在该问题中，总共存在 8 个属性值，分别为为水泥、高炉渣、粉煤灰、水、超增塑剂、粗聚合剂、精细聚合物的物质含量及年龄八个自变量；具体英文名称及其单位如下表所示。

Cement (kg in a m ³ mixture)	Blast Furnace Slag (kg in a m ³ mixture)	Fly Ash (kg in a m ³ mixture)
Water (kg in a m ³ mixture)	Superplasticizer (kg in a m ³ mixture)	Coarse Aggregate (kg in a m ³ mixture)
Fine Aggregate (kg in a m ³ mixture)	Age (day)	Concrete compressive strength (MPa, megapascals)

表 1 数据集中的人为控制的 7 个自变量及混凝土年龄和抗压强度

Table 1. The decision tree results obtained from the second classification method.

接下来，我们可以通过将所有数据描绘在坐标图上，显示出每个自变量与混凝土抗压系数之间的总体关系区间，如下图所示。其中，每一种颜色代表一个自变量的属性。

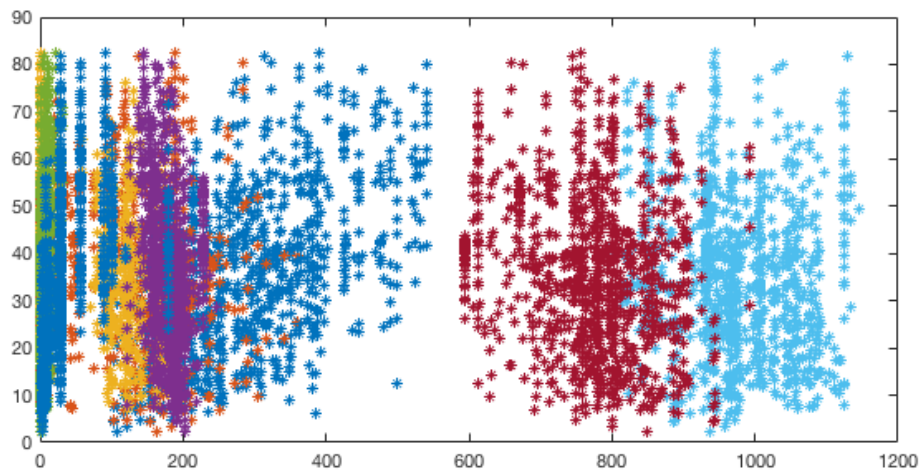


图5 混凝土抗压强度与八种自变量的关系。

Fig.5. Relationships between concrete compressive strength and eight variables.

4 结果

4.1 数据可视化

首先，我们可以先对数据集中所有的属性或自变量进行可视化以便大体分析这些数据属性之间的关联。混凝土抗压自变量间的数据关联如下图所示。可以观察得，数据之间并没有呈现出明显的比例关系；其中，颜色的分别是由混凝土年龄的不同而改变。由图像可得，第 1、2 组数据（水泥和高炉渣含量）间大体呈反比关系。第 7、9 组数据间（粗聚合剂含量与抗压强度）也粗略呈反比关系。在所有数据中，我们还可以通过单独显示两组数据判断其关联性。

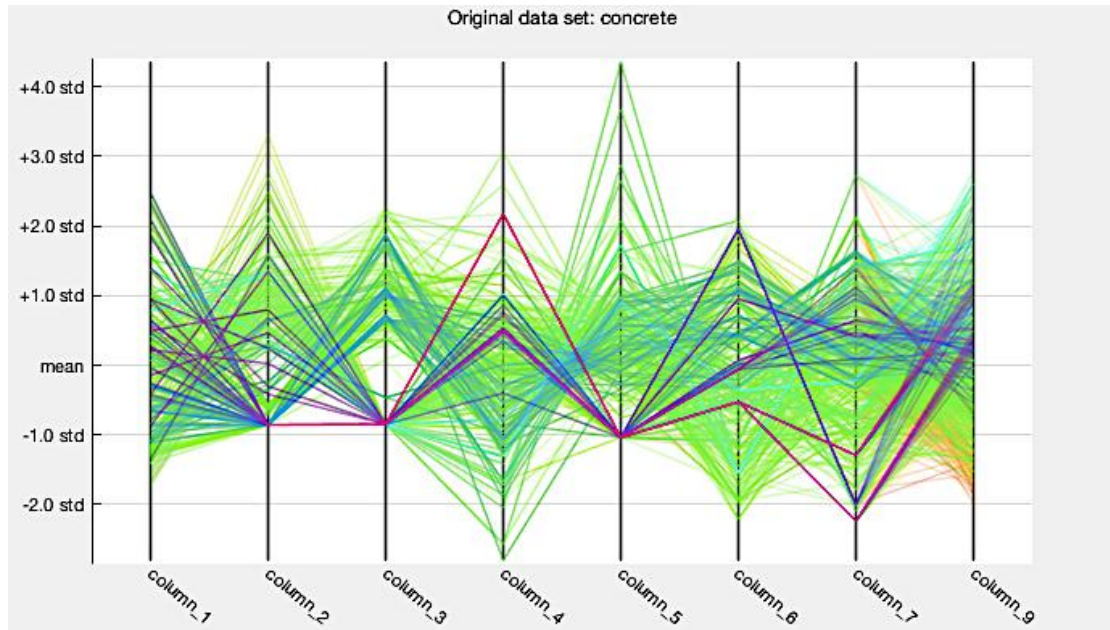


图 5 数据库中包含所有变量的数据可视化。其中，每种颜色代表一种混凝土的属性。

Fig.5. Visible data in the database. Among the data, each color indicates a concrete attribute.

通过对大量自变量进行显示判断，最终我们发现第 1 组数据（水泥含量）和混凝土抗压强度之间呈现相对其他自变量较好的线性关联，尽管这种关联并不明显。具体呈现效果如下图所示。

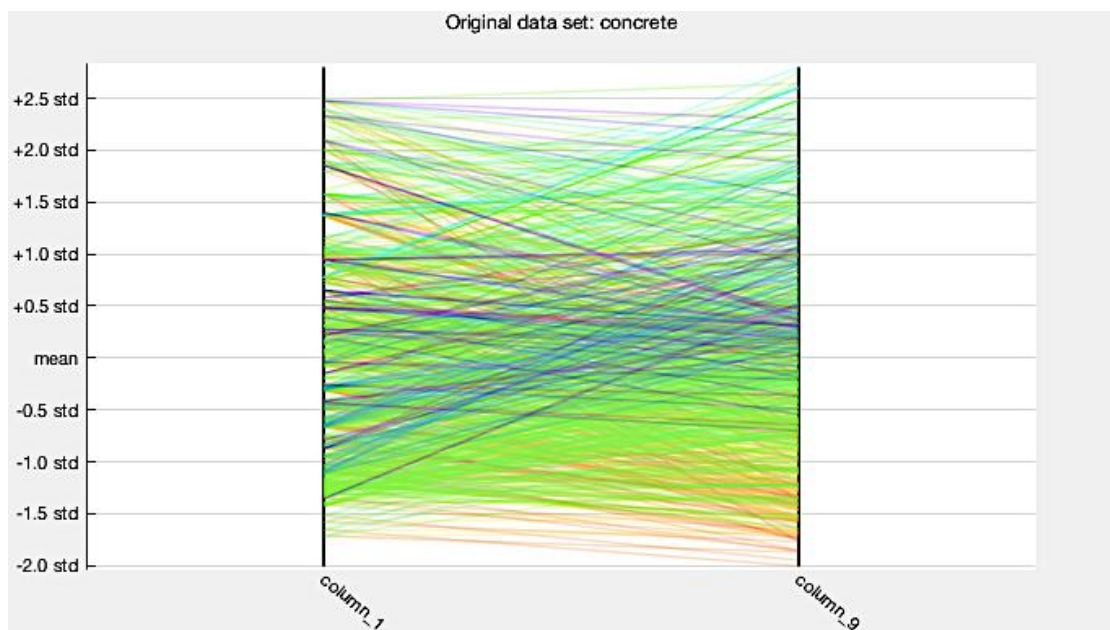


图 5 水泥含量和混凝土抗压强度之间关系。

Fig.5. Relationship between cement and compressive strength.

4.2 线性模型

如上文所述，我们需要对已有的八个自变量和混凝土的抗压强度进行多元线性拟合。根据上述的数学模型，数据库中的自变量可用矩阵表示为

$$\mathbf{X}_{variables} = \begin{pmatrix} \text{Cement} & \text{BlastFurnace} & \text{Water} & \dots (8 \text{ variables}) \dots & \text{Age} & 1 \\ \text{value}_{c1} & \text{value}_{b1} & \text{value}_{w1} & \dots (\text{values}) \dots & \text{value}_{a1} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \text{value}_{c1030} & \text{value}_{b1030} & \text{value}_{w1030} & \dots & \text{value}_{a1030} & 1 \end{pmatrix}$$

其中，*Cement*, *BlastFurnace* 等代表混凝土属性名称，其下的列代表该属性的数据： value_{c1} 代表水泥的第一个数据， value_{c1030} 代表其第 1030 个数据，即数据库中最后一个数据。以此类推，剩下的数据都是其他属性的数据。将这八组数据归入 \mathbf{X} 矩阵，并求解逆矩阵及转置，便后续计算模型。

再对因变量进行分析，在该问题中，我们要研究的因变量是混凝土的抗压强度，即其被归为 \mathbf{y} 矩阵

$$\mathbf{y}_{CompressiveStrengthValue} = (\text{value}_1; \text{value}_2; \dots; \text{value}_{1030})$$

此时，即算得 $\hat{\omega}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ ，即获得了线性项。因此，由最终式 $f(\hat{\mathbf{x}}_i) = \hat{\mathbf{x}}_i^T \hat{\omega}^*$ 可知学习得混凝土抗压线性模型。计算结果如下。

Linear regression model:				
y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8				
Estimated Coefficients:				
	Estimate	SE	tStat	pValue
(Intercept)	-23.164	26.588	-0.8712	0.38385
x1(Cement)	0.11979	0.0084894	14.11	1.9628e-41
x2(Blast Furnace Slag)	0.10385	0.010136	10.245	1.6331e-23
x3(Fly Ash)	0.087943	0.012585	6.9879	5.03e-12
x4(Water)	-0.1503	0.040179	-3.7407	0.00019373
x5(Superplasticizer)	0.29069	0.09346	3.1103	0.0019209
x6(Coarse Aggregate)	0.01803	0.0093942	1.9193	0.055227
x7(Fine Aggregate)	0.020154	0.010703	1.8831	0.059968
x8(Age)	0.11423	0.0054275	21.046	5.841e-82
Number of observations: 1030, Error degrees of freedom: 1021				

Root Mean Squared Error: 10.4
R-squared: 0.615, Adjusted R-Squared 0.612
F-statistic vs. constant model: 204, p-value = 6.76e-206

表 1 线性模型拟合结果

Table 1. The result of the linear regression.

获得结果后，我们学得线性模型。根据数据，我们可以获得方差为 0.615，协方差为一 8×8 矩阵；同时，获得 p 如上述表中最右列所示(pValue)，通过运算，我们还可以算得该模型的 t 值为为一矩阵 $[26.7001, 18.4993, 10.0893, -11.3592, 2.9932, 3.3653, 3.1974, 21.0409]^T$ 。绘制拟合得到曲线，如下图所示。

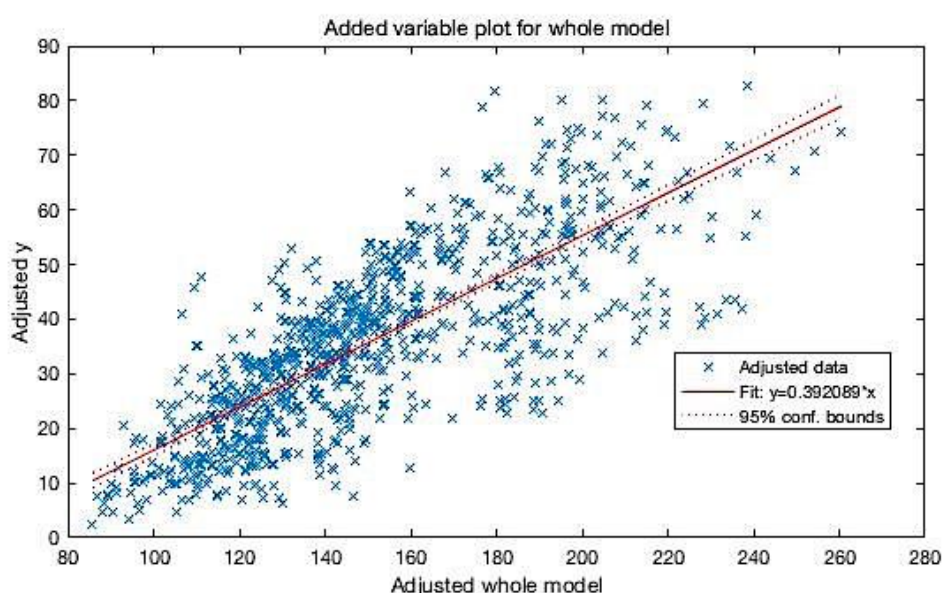


图5 拟合得到线性模型曲线图

Fig.5. Linear regression model fitted based on concrete data

5 结论

通过将混凝土制备过程中的数据库导入 MATLAB 软件计算，我们可以分析出其制备过程的 8 个变量和混凝土抗压强度的关系，拟合成多元线性回归模型。最终结果中，我们看到水泥含量与混凝土抗压强度线性常数为 0.11979，高炉渣与混凝土抗压强度线性常数为 0.10385，粉煤灰线性常数 0.087943，水含量线性常数 -0.1503，超增塑剂线性常数 0.29069，粗聚合剂线性常数 0.01803，精细聚合物线性常数 0.020154，混凝土年龄线性常数 0.11423。并且在所有数据中，和混凝土线性关系相对较强的自变量为水泥含量。

6 总结

线性回归是统计学中回归分析的一种，是确定两种或多种变量间(混凝土实验力学系数与其抗压强度)相互依赖或指标性的定量关系的一种统计分析方法，运用广泛。在回归分析中，若回归分析中包括两个或两个以上的自变量(6个力学参数)，并且因变量和自变量之间是线性关系或满足广义线性关系，即为多元线性回归分析。在该多元回归问题中，混凝土的抗压形式是作为自变量被用来评估其制备过程中的多种材料含量；在实际制备过程中，很难给出确切的公式和数学推导来分析混凝土力学性能和自变量的推导关系，因此，本次研究可以用来更好地评估分析混凝土制备工艺和方法。最后，作为一名力学学科的学生，分析建筑材料的抗压性能是我们必备技能之一，而通过机器学习的方法来分析研究这样的过程能够让我们更加好的理解算法的奥秘与强大。

感谢上海大学力学与工程科学学院胡国辉教授的讲授与讨论。

参考文献

- [1]史大全,方光秀,李普,杨旭.磨细稻壳和再生骨料对混凝土抗压强度影响研究[J].山西建筑,2020,46(03):96-98.
- [2]王莘晴.浅析提高预填骨料混凝土抗压强度的工程措施[J].建材与装饰,2020(04):54-55.
- [3] P. Harrington(compile), Li.Y, Li.P, Qu.Y, Wang.B(translate)Machines Learning in Action 2013 (The People's Posts and Telecommunications Press) (in Chinese) [P. Harrington 编, 李锐等 译 机器学习实战 2013 (人民邮电出版社)]
- [4] Zhou Zhihua. Machines Learning 2018 (Tsinghua University Press) (in Chinese)
[周志华 机器学习 2018 (清华大学出版社)]
- [5] J. Mata, Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models, Engineering Structures, Volume 33, Issue 3, 2011, Pages 903-910, ISSN 0141-0296,
<https://doi.org/10.1016/j.engstruct.2010.12.011>.
- [6] Faezehossadat Khademi, Sayed Mohammadmehdi Jamal, Neela Deshpande, Shreenivas Londhe, Predicting strength of recycled aggregate concrete using Artificial

Neural Network, Adaptive Neuro-Fuzzy Inference System and Multiple Linear Regression, International Journal of Sustainable Built Environment, Volume 5, Issue 2, 2016, Pages 355-369, ISSN 2212-6090, <https://doi.org/10.1016/j.ijsbe.2016.09.003>.

Linear Regression on Concrete Compressive Strength

Zhai Hanfeng

(School of Mechanics and Engineering Science, Shanghai University, Shanghai 200444, China)

Abstract

As a sort of fundamental and significantly important compression material, concrete is widely applied in civil engineering fields such as bridge construction and building construction. In the long history of humans preparing and applying concrete, the compressive properties of concrete have a direct or indirect relationship with a large number of known or potential variables. Therefore, it is difficult to discern the relationship and model of concrete compressive capacity to various independent variables through experimental methods and traditional methods of mathematical derivation. Therefore, using the linear regression algorithm in machine learning, we can fit the relationship between the concrete compressive limit and a large amount of existing recorded data through a large amount of data and superposition calculation, and obtain a corresponding linear model. In this paper, the author collects 1030 sets of data including eight coefficients of concrete cement, blast furnace slag content, blown dust amount, and water content, and the compression limit of the concrete to fit and calculate a linear model. By discussing the obtained linear model, we can basically analyze the relationship between the different independent variables and the compressive strength of the concrete in the measured data. With this estimation, we can analyze the control conditions and the possible mechanical properties of the concrete during the concrete preparation process. Furthermore, these analysis provide decent insights for manufacturing and civil engineering.

Keywords: machine learning; concrete; compressive strength; linear regression model