Linear Regression and Regularization

Applied Machine Learning for Educational Data Science

true

10/12/2021

Contents

Linear Regression **1**

Model Description . . . 1

Model Estimation . . . 3

Model Evaluation . . . 3

Model Interpretation . . . 3

Linear Regression with Regularization **3**

Ridge Penalty . . . 3

Lasso Penalty . . . 3

Elastic Net . . . 3

Wrapping up: Predicting Readability Scores **3**

[Updated: Wed, Oct 13, 2021 - 11:29:33]

In the machine learning literature, the prediction algorithms are classified into two main categories: *supervised* and *unsupervised*. Supervised algorithms are being used when the dataset has an actual outcome of interest to predict (labels), and the goal is to build the "best" model predicting the outcome of interest that exists in the data. On the other side, unsupervised algorithms are being used when the dataset doesn't have an outcome of interest, and the goal is typically to identify similar groups of observations (rows of data) or similar groups of variables (columns of data) in data. In this course, we plan to cover a number of *supervised* algorithms. Linear regression is one of the simplest approach among supervised algorithms, and also one of the easiest to interpret.

Linear Regression

Model Description

In most general terms, the linear regression model with P predictors (X_1,X_2,X_3,...,X_p) to predict an outcome (Y) can be written as the following:

$$Y = \beta_0 + \sum_{p=1}^{P} \beta_p X_p + \epsilon.$$

In this model, Y represents the observed value for the outcome for an observation, X_p represents the observed value of the p^{th} variable for the same observation, and β_p is the associated model parameter for the p^{th} variable. ϵ is the model error (residual) for the observation.

This model includes only the main effects of each predictor and can be easily extended by including a quadratic or higher-order polynomial terms for all (or a specific subset of) predictors. For instance, the model below includes all first-order, second-order, and third-order polynomial terms for all predictors.

$$Y = \beta_0 + \sum_{p=1}^{P} \beta_p X_p + \sum_{k=1}^{P} \beta_{k+P} X_k^2 + \sum_{m=1}^{P} \beta_{m+2P} X_m^3 + \epsilon.$$

The simple first-order, second-order, and third-order polynomial terms can also be replaced by corresponding terms obtained from B-splines or natural splines.

Sometimes, the effect of predictor variables on the outcome variable are not additive, and the effect of one predictor on the response variable can depend on the levels of another predictor. These non-additive effects are also called interaction effects. The interaction effects can also be a first-order interaction (interaction between two variables, e.g., $X_1 * X_2$), second-order interaction ($X_1 * X_2 * X_3$), or higher orders. It is also possible to add the interaction effects to the model. For instance, the model below also adds the first-order interactions.

$$Y = \beta_0 + \sum_{p=1}^{P} \beta_p X_p + \sum_{k=1}^{P} \beta_{k+P} X_k^2 + \sum_{m=1}^{P} \beta_{m+2P} X_m^3 + \sum_{i=1}^{P} \sum_{j=i+1}^{P} \beta_{i,j} X_i X_j + \epsilon.$$

If you are not comfortable or confused with notational representation, below is a simple example for different models you can write with 5 predictors (X_1, X_2, X_3, X_4, X_5).

A model with only main-effects:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon.$$

A model with polynomial terms up to the 3rd degree added:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_1^2 + \beta_7 X_2^2 + \beta_8 X_3^2 + \beta 9 X_4^2 + \beta_{10} X_5^2 + \beta_{11} X_1^3 + \beta_{12} X_2^3 + \beta_{13} X_3^3 + \beta 14 X_4^3 + \beta_{15} X_5^3 + \epsilon$$

A model with both interaction terms and polynomial terms up to the 3rd degree added:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_1^2 + \beta_7 X_2^2 + \beta_8 X_3^2 + \beta 9 X_4^2 + \beta_{10} X_5^2 + \beta_{11} X_1^3 + \beta_{12} X_2^3 + \beta_{13} X_3^3 + \beta 14 X_4^3 + \beta_{15} X_5^3 + \beta_1$$

Model Estimation

Model Evaluation

Model Interpretation

Linear Regression with Regularization

Ridge Penalty

Lasso Penalty

Elastic Net

Wrapping up: Predicting Readability Scores