Linear Regression

Fuhao Zou(邹复好)

Intelligent and Embedded Computing Lab,
Huazhong University of Science & Technology

fuhao_zou@hust.edu.cn

2019年04月14日

Table of contents

Table of Contents

Formal definition:

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

- Data Assumption: $y_i \in \mathbb{R}$
- Model Assumption: $y_i = \mathbf{w}^\top \mathbf{x}_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$

$$\Rightarrow y_i|\mathbf{x}_i \sim N(\mathbf{w}^\top \mathbf{x}_i, \sigma^2) \Rightarrow P(y_i|\mathbf{x}_i, \mathbf{w}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mathbf{x}_i^\top \mathbf{w} - y_i)^2}{2\sigma^2}}$$

Formal definition:

- In words, we assume that the data is drawn from a "line" $\mathbf{w}^\top \mathbf{x}$ through the origin (one can always add a bias / offset through an additional dimension, similar to the Perceptron). For each data point with features $\mathbf{x}_i$, the label y is drawn from a Gaussian with mean $\mathbf{w}^\top \mathbf{x}_i$ and variance σ^2. Our task is to estimate the slope $\mathbf{w}$ from the data.

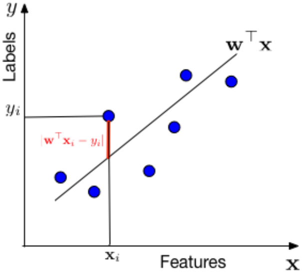
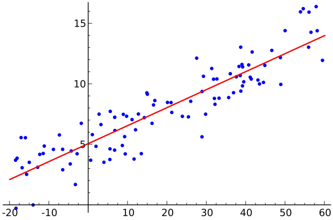


Table of Contents

Simple and multiple linear regression

The very simplest case of a single scalar predictor variable x and a single scalar response variable y is known as simple linear regression. The extension to multiple and/or vector-valued predictor variables (denoted with a capital X) is known as multiple linear regression, also known as multivariable linear regression. Nearly all real-world regression models involve multiple predictors, and basic descriptions of linear regression are often phrased in terms of the multiple regression model.



Generalized linear models

Generalized linear models (GLMs) are a framework for modeling response variables that are bounded or discrete. This is used, for example:

- when modeling positive quantities (e.g. prices or populations) that vary over a large scale— which are better described using a skewed distribution such as the log-normal distribution or Poisson distribution (although GLMs are not used for log-normal data, instead the response variable is simply transformed using the logarithm function);
- when modeling categorical data, such as the choice of a given candidate in an election (which is better described using a Bernoulli distribution/binomial distribution for binary choices, or a categorical distribution/multinomial distribution for multi-way choices), where there are a fixed number of choices that cannot be meaningfully ordered;
- when modeling ordinal data, e.g. ratings on a scale from 0 to 5, where the different outcomes can be ordered but where the quantity itself may not have any absolute meaning (e.g. a rating of 4 may not be "twice as good" in any objective sense as a rating of 2, but simply indicates that it is better than 2 or 3 but not as good as 5).

Hierarchical linear models

Hierarchical linear models (or multilevel regression) organizes the data into a hierarchy of regressions, for example where A is regressed on B, and B is regressed on C. It is often used where the variables of interest have a natural hierarchical structure such as in educational statistics, where students are nested in classrooms, classrooms are nested in schools, and schools are nested in some administrative grouping, such as a school district. The response variable might be a measure of student achievement such as a test score, and different covariates would be collected at the classroom, school, and school district levels.

Table of Contents

Estimating with MLE

$$
\begin{aligned}
\mathbf{w} &= \underset{\mathbf{w}}{\operatorname{argmax}}\, P(y_1, \mathbf{x}_1, \ldots, y_n, \mathbf{x}_n | \mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i, \mathbf{x}_i | \mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i | \mathbf{x}_i, \mathbf{w}) P(\mathbf{x}_i | \mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i | \mathbf{x}_i, \mathbf{w}) P(\mathbf{x}_i) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i | \mathbf{x}_i, \mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left[P(y_i | \mathbf{x}_i, \mathbf{w}) \right] \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \sum_{i=1}^{n} \left[\log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) + \log \left(e^{-\frac{(\mathbf{x}_i^\top \mathbf{w} - y_i)^2}{2\sigma^2}} \right) \right]
\end{aligned}
$$

Estimating with MLE

$$= \underset{\mathbf{w}}{\operatorname{argmax}} -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{x}_i^\top \mathbf{w} - y_i)^2$$

$$= \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^\top \mathbf{w} - y_i)^2$$

- We are minimizing a loss function, $l(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^\top \mathbf{w} - y_i)^2$. This particular loss function is also known as the squared loss or Ordinary Least Squares (OLS). OLS can be optimized with gradient descent, Newton's method, or in closed form. Closed Form: $\mathbf{w} = (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X}\mathbf{y}^\top$ where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $\mathbf{y} = [y_1, \ldots, y_n]$.

Table of Contents

Estimating with MAP

Additional Model Assumption: $P(\mathbf{w}) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{\mathbf{w}^\top \mathbf{w}}{2\tau^2}}$

$$
\begin{aligned}
\mathbf{w} &= \underset{\mathbf{w}}{\operatorname{argmax}}\, P(\mathbf{w} | y_1, \mathbf{x}_1, \ldots, y_n, \mathbf{x}_n) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}}\, \frac{P(y_1, \mathbf{x}_1, \ldots, y_n, \mathbf{x}_n | \mathbf{w}) P(\mathbf{w})}{P(y_1, \mathbf{x}_1, \ldots, y_n, \mathbf{x}_n)} \\
&= \underset{\mathbf{w}}{\operatorname{argmax}}\, P(y_1, \mathbf{x}_1, \ldots, y_n, \mathbf{x}_n | \mathbf{w}) P(\mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \left[\prod_{i=1}^{n} P(y_i, \mathbf{x}_i | \mathbf{w}) \right] P(\mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \left[\prod_{i=1}^{n} P(y_i | \mathbf{x}_i, \mathbf{w}) P(\mathbf{x}_i | \mathbf{w}) \right] P(\mathbf{w}) \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \left[\prod_{i=1}^{n} P(y_i | \mathbf{x}_i, \mathbf{w}) P(\mathbf{x}_i) \right] P(\mathbf{w})
\end{aligned}
$$

Estimating with MAP

$$= \operatorname*{argmax}_{\mathbf{w}} \left[\prod_{i=1}^{n} P(y_i|\mathbf{x}_i, \mathbf{w})\right] P(\mathbf{w})$$

$$= \operatorname*{argmax}_{\mathbf{w}} \sum_{i=1}^{n} \log P(y_i|\mathbf{x}_i, \mathbf{w}) + \log P(\mathbf{w})$$

$$= \operatorname*{argmin}_{\mathbf{w}} \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\mathbf{x}_i^\top \mathbf{w} - y_i)^2 + \frac{1}{2\tau^2} \mathbf{w}^\top \mathbf{w}$$

$$= \operatorname*{argmin}_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^\top \mathbf{w} - y_i)^2 + \lambda ||\mathbf{w}||_2^2$$

- This objective is known as Ridge Regression. It has a closed form solution of: $\mathbf{w} = (\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I})^{-1} \mathbf{X}\mathbf{y}^\top$, where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $\mathbf{y} = [y_1, \ldots, y_n]$.

Table of Contents

Finance

Finance

The capital asset pricing model uses linear regression as well as the concept of beta for analyzing and quantifying the systematic risk of an investment. This comes directly from the beta coefficient of the linear regression model that relates the return on the investment to the return on all risky assets.

Economics

Linear regression is the predominant empirical tool in economics. For example, it is used to predict consumption spending, fixed investment spending, inventory investment, purchases of a country's exports, spending on imports, the demand to hold liquid assets, labor demand, and labor supply.

Environmental science

Linear regression finds application in a wide range of environmental science applications. In Canada, the Environmental Effects Monitoring Program uses statistical analyses on fish and benthic surveys to measure the effects of pulp mill or metal mine effluent on the aquatic ecosystem.

Machine learning

Linear regression plays an important role in the field of artificial intelligence such as machine learning. The linear regression algorithm is one of the fundamental supervised machine-learning algorithms due to its relative simplicity and well-known properties.

Table of Contents

Summary

- **Ordinary Least Squares**:

  $\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^\top \mathbf{w} - y_i)^2$

  Squared loss

  No regularization

  Closed form: $\mathbf{w} = (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X}\mathbf{y}^\top$

- **Ridge Regression**:

  $\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^\top \mathbf{w} - y_i)^2 + \lambda ||\mathbf{w}||_2^2$

  Squared loss

  $l2$-regularization

  Closed form: $\mathbf{w} = (\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I})^{-1}\mathbf{X}\mathbf{y}^\top$

Simple linear regression implementation

Simple linear regression implementation code click here

housing price prediction

housing price prediction code click here

The End