The Simple Linear Regression: Formula Derivation

Hongjian YUAN

May 2020

1 Introduction

Simple Linear Regression(SLR) is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables: One variable, denoted x, is regarded as the predictor, explanatory, or independent variable; The other variable, denoted y, is regarded as the response, outcome, or dependent variable. [3]

2 Prerequisite of Derivation

2.1 Convex Function and Concave Function(Binary)

[2] Specific tests for convexity of Binary functions: Let:

$$a = f_{xx}; b = f_{xy}; c = f_{yy} \tag{1}$$

Then:
f is convex for $\boldsymbol{D}$ when $a < 0$ and $ac - b^2 >= 0$;
f is concave for $\boldsymbol{D}$ when $a > 0$ and $ac - b^2 >= 0$.

2.2 Partial Derivative and Minima & Maxima

Let f be a function with two variables with continuous second order partial derivatives f_{xx}, f_{yy} and f_{xy} at a critical point (a,b). Let:

$$D = f_{xx}(a,b)f_{yy}(a,b) - f_{xy}^2(a,b) \tag{2}$$

a) If $D > 0$ and $f_{xx}(a,b) > 0$, then f has a relative minimum at (a,b).
b) If $D > 0$ and $f_{xx}(a,b) < 0$, then f has a relative maximum at (a,b).
c) If $D < 0$, then f has a saddle point at (a,b).
d) If $D = 0$, then no conclusion can be drawn.

3 Process of Derivation

[1]

3.1 Convexity of Loss Function

Proof.

$$E(w,b) = \sum_{i=1}^{m}(y_i - f(x_i))^2 = \sum_{i=1}^{m}(y_i - wx_i - b)^2 \tag{3}$$

$$\frac{\partial E(w,b)}{\partial w} = \sum_{i=1}^{m}\frac{\partial}{\partial w}(y_i - wx_i - b)^2 = \sum_{i=1}^{m}\times 2(y_i - wx_i - b)(-x_i) = 2(w\sum_{i=1}^{m}x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i) \tag{4}$$

From (1):

$$a = \frac{\partial^2 E(w,b)}{\partial w^2} = 2\sum_{i=1}^{m}x_i^2 \tag{5}$$

$$b = \frac{\partial^2 E(w,b)}{\partial w \partial b} = 2\sum_{i=1}^{m}x_i \tag{6}$$

$$c = \frac{\partial^2 E(w,b)}{\partial b^2} = 2m \tag{7}$$

From (5)(6)(7):

$$ac - b^2 = 4m \times \sum_{i=1}^{m}x_i^2 - 4\sum_{i=1}^{m}x_i^2 = 4(m-1)\sum_{i-1}^{m}x_i^2 \geq 0 \tag{8}$$

From (5), we can easily know that $a \geq 0$, therefore we can conclude that **the loss function is convex**.

3.2 Formula Derivation of optimized *b*

$$\frac{\partial E(w,b)}{\partial b} = \sum_{i=1}^{m}\frac{\partial}{\partial b}(y_i - wx_i - b)^2 = \sum_{i=1}^{m}2(y_i - wx_i - b)(-1) = 2(w\sum_{i=1}^{m}x_i - \sum_{i=1}^{m}(yi - b)) \tag{9}$$

Let(9) = 0, then:

$$b = \frac{1}{m}\sum_{i=1}^{m}(y_i - wx_i) = \bar{y} - w\bar{x} \tag{10}$$

3.3 Formula Derivation of optimized $\boldsymbol{w}$

$$\frac{\partial E(w,b)}{\partial w} = 2(w\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i) \tag{11}$$

Let(11) = 0, from also (10), then:

$$w\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i y_i + \sum_{i=1}^{m} x_i(\bar{y} - w\bar{x}) = 0 \tag{12}$$

$$\begin{aligned} w(\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i\bar{x}) &= \sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i\bar{y} \\ w &= \frac{\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i\bar{y}}{\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i\bar{x}} \\ &= \frac{\sum_{i=1}^{m} y_i(x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2} \end{aligned} \tag{13}$$

3.4 Vectorization of $\boldsymbol{w}$

From (13) and $\frac{1}{m}(\sum_{i=1}^{m} x_i)^2 = \bar{x}\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} x_i\bar{x}$:

$$w = \frac{\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i\bar{y}}{\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i\bar{x}}$$

For:

$$\sum_{i=1}^{m} y_i\bar{x} = \bar{x}\sum_{i=1}^{m} y_i = \frac{1}{m}\sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i = \sum_{i=1}^{m} x_i * \frac{1}{m} * \sum_{i=1}^{m} y_i = \sum_{i=1}^{m} x_i\bar{y} \tag{14}$$

$$\sum_{i=1}^{m} y_i\bar{x} = barx\sum_{i=1}^{m} y_i = \bar{x} * m * \frac{1}{m} * \sum_{i=1}^{m} y_i = m\bar{x}\bar{y} = \sum_{i=1}^{m} \bar{x}\bar{y} \tag{15}$$

$$\sum_{i=1}^{m} x_i\bar{x} = \bar{x}\sum_{[} i = 1]^m x_i = \bar{x} * m * \frac{1}{m} * \sum_{i=1}^{m} x_i = m\bar{x}^2 = \sum_{i=1}^{m} \bar{x}^2 \tag{16}$$

From (14)(15)(16):

$$\begin{aligned} w = \frac{\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i\bar{y}}{\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i\bar{x}} &= \frac{\sum_{i=1}^{m}(y_i x_i - y_i\bar{x} - y_i\bar{x} + y_i\bar{x})}{\sum_{i=1}^{m}(x_i^2 - x_i\bar{x} - x_i\bar{x} + x_i\bar{x})} \\ = \frac{\sum_{i=1}^{m}(y_i x_i - y_i\bar{x} - x_i\bar{y} + \bar{x}\bar{y})}{\sum_{i=1}^{m}(x_i^2 - x_i\bar{x} - x_i\bar{x} + \bar{x}^2)} &= \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2} \end{aligned} \tag{17}$$

Suppose:

$$x = (x_1, x_2, x_3, ..., x_m)^T, x_d = (x_1 - \bar{x}, x_2 - \bar{x}, ..., x_m - \bar{x})^T \tag{18}$$

Similarly:

$$y = (y_1, y_2, y_3, ..., y_m)^T y_d = (y_1 - \bar{y}, y_2 - \bar{y}, ..., y_m - \bar{y})^T \tag{19}$$

From (17)(18)(19):

$$w = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2} = \frac{{X_d}^T y_d}{{X_d}^T X_d} \tag{20}$$

References

[1] *Machine Learning*. Tsinghua University, 2016.

[2] W. Kaplan. *Maxima and Minima with Applications: Practical Optimization and Duality*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.

[3] The Pennsylvania State University. 2.1 - what is simple linear regression? https://online.stat.psu.edu/stat462/node/91/ Accessed May 4, 2020.