Linear Regression Mathematics

Charlie Clark

January 6, 2022

1 Simple Linear Regression

Model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where $1 \leq i \leq N$ and $N \geq 1$ is the sample size.

Ordinary Least Squares Cost Function:

$$J\left(\hat{\beta}_0, \hat{\beta}_1\right) = \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

Solve for Coefficients:

$$\begin{aligned}
\min_{\left(\hat{\beta}_0, \hat{\beta}_1\right)} J\left(\hat{\beta}_0, \hat{\beta}_1\right) &= \min_{\left(\hat{\beta}_0, \hat{\beta}_1\right)} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 && \because \text{Minimize cost function.} \\
\Rightarrow \frac{\partial J}{\partial \hat{\beta}_0} &= \frac{\partial}{\partial \hat{\beta}_0} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 && \because \text{Partial with respect to } \hat{\beta}_0. \\
&= -2 \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) && \because \text{Chain rule.} \\
&= -2 \left(\sum_{i=1}^{N} y_i - N\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} x_i\right) && \because \text{Distribute summation.} \\
&= 0 && \because \frac{\partial J}{\partial \hat{\beta}_0} = 0 \text{ at minimum.} \\
\Leftrightarrow 0 &= -2 \left(\sum_{i=1}^{N} y_i - N\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} x_i\right) && \because \text{Rewrite.} \\
&= \sum_{i=1}^{N} y_i - N\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} x_i && \because \text{Divide by } -2. \\
\Leftrightarrow N\hat{\beta}_0 &= \sum_{i=1}^{N} y_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i && \because \text{Add } N\hat{\beta}_0. \\
\Leftrightarrow \hat{\beta}_0 &= \overline{y} - \hat{\beta}_1 \overline{x} && \because \text{Divide by} N, \text{ substitute averages.}
\end{aligned}$$

$$
\begin{aligned}
\Rightarrow \frac{\partial J}{\partial \hat{\beta}_1} &= \frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 && \because \text{Partial with respect to } \hat{\beta}_1. \\
&= -2 \sum_{i=1}^{N} x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) && \because \text{Chain rule.} \\
&= -2 \sum_{i=1}^{N} \left(y_i x_i - \hat{\beta}_0 x_i - \hat{\beta}_1 x_i^2\right) && \because \text{Distribute } x_i. \\
&= -2 \left(\sum_{i=1}^{N} y_i x_i - \hat{\beta}_0 \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2\right) && \because \text{Distribute summation.} \\
&= 0 && \because \frac{\partial J}{\partial \hat{\beta}_1} = 0 \text{ at minimum.} \\
\Leftrightarrow 0 &= -2 \left(\sum_{i=1}^{N} y_i x_i - \hat{\beta}_0 \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2\right) && \because \text{Rewrite.} \\
&= \sum_{i=1}^{N} y_i x_i - \hat{\beta}_0 \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 && \because \text{Divide by } -2. \\
&= \overline{yx} - \hat{\beta}_0 \overline{x} - \hat{\beta}_1 \overline{x^2} && \because \text{Divide by } N. \\
&= \overline{yx} - \left(\overline{y} - \hat{\beta}_1 \overline{x}\right) \overline{x} - \hat{\beta}_1 \overline{x^2} && \because \text{Substitute } \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}. \\
&= \overline{yx} - \overline{y} \cdot \overline{x} - \hat{\beta}_1 \overline{x^2} + \hat{\beta}_1 \overline{x}^2 && \\
&= \overline{yx} - \overline{y} \cdot \overline{x} - \hat{\beta}_1 \left(\overline{x^2} - \overline{x}^2\right) && \because \text{Factor } -\hat{\beta}_0. \\
\Leftrightarrow \hat{\beta}_1 \left(\overline{x^2} - \overline{x}^2\right) &= \overline{yx} - \overline{y} \cdot \overline{x} && \because \text{Add } \hat{\beta}_1 \left(\overline{x^2} - \overline{x}^2\right). \\
\Leftrightarrow \hat{\beta}_1 &= \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} && \because \text{Divide by } \overline{x^2} - \overline{x}^2.
\end{aligned}
$$

Therefore,

$$
\hat{\beta}_1 = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} \text{ and } \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x},
$$

where

$$
\begin{aligned}
\overline{y} &= \frac{1}{N} \sum_{i=1}^{N} y_i, \\
\overline{x} &= \frac{1}{N} \sum_{i=1}^{N} x_i, \\
\overline{yx} &= \frac{1}{N} \sum_{i=1}^{N} y_i x_i, \\
\text{and } \overline{x^2} &= \frac{1}{N} \sum_{i=1}^{N} x_i^2.
\end{aligned}
$$

2 Multiple Linear Regression

Model:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_K x_{i,K} + \varepsilon_i,$$

where $1 \leq i \leq N, N \geq 1$ is the sample size, and $K \geq 2$ is the number of coefficients -1.

Model in Matrix Form:

$$\vec{y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon},$$

where

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,K} \end{bmatrix}, \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \text{ and } \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Ordinary Least Squares Cost Function:

$$J\left(\vec{\hat{\beta}}\right) = \left(\vec{y} - \mathbf{X}\vec{\hat{\beta}}\right)^{\mathrm{T}} \left(\vec{y} - \mathbf{X}\vec{\hat{\beta}}\right) = \vec{y}^{\mathrm{T}}\vec{y} - \vec{y}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}} - \vec{\hat{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\vec{y} + \vec{\hat{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}}.$$

Solve for Coefficients:

$$\begin{aligned}
\min_{\left(\vec{\hat{\beta}}\right)} J\left(\vec{\hat{\beta}}\right) \Rightarrow \frac{\partial J}{\partial \vec{\hat{\beta}}} = 0 & \qquad \because \text{Minimum when slop is 0.} \\
\Leftrightarrow \frac{\partial}{\partial \vec{\hat{\beta}}} \left(\vec{y}^{\mathrm{T}}\vec{y} - \vec{y}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}} - \vec{\hat{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\vec{y} + \vec{\hat{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}}\right) = 0 & \qquad \because \text{Definition of } J\left(\vec{\hat{\beta}}\right) \\
\Leftrightarrow 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}} - 2\mathbf{X}^{\mathrm{T}}\vec{y} = 0 & \qquad \because \text{Simplify.} \\
\Leftrightarrow 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}} = 2\mathbf{X}^{\mathrm{T}}\vec{y} & \qquad \because \text{Add } 2\mathbf{X}^{\mathrm{T}}\vec{y}. \\
\Leftrightarrow \mathbf{X}^{\mathrm{T}}\mathbf{X}\vec{\hat{\beta}} = \mathbf{X}^{\mathrm{T}}\vec{y} & \qquad \because \text{Divide by 2.} \\
\Leftrightarrow \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\vec{\hat{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\vec{y} & \qquad \because \text{Multiply by } \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}. \\
\Leftrightarrow \vec{\hat{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\vec{y} & \qquad \because \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) = \mathbf{I},\ \mathbf{I}\vec{\hat{\beta}} = \vec{\hat{\beta}}.
\end{aligned}$$

Therefore,

$$\vec{\hat{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\vec{y},$$

where

$$\vec{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}.$$