Derivation of MSE closed-form solution

Chieh Wu

January 22, 2025

1. Define the MSE Objective

The Mean Squared Error (MSE) objective for linear regression is defined as:

$$\mathcal{L}(w) = \frac{1}{n}\sum_{i=1}^{n}(\phi(x_i)^\top w - y_i)^2 \tag{1}$$

Here:

- $\phi(x_i)^\top$ is a row vector of size $1 \times d$, representing the feature map for the i-th data point.
- w is a column vector of size $d \times 1$, representing the model parameters.
- y_i is a scalar, representing the target value for the i-th data point.

2. Rewrite the MSE in Matrix Form

The MSE objective can be rewritten into a matrix format. Let's start by replacing the inner term with $z_i = (\phi(x_i)^\top w - y_i)$, we have

$$\frac{1}{n}\sum_{i=1}^{n}(\phi(x_i)^\top w - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\underbrace{\begin{bmatrix} z_1 & z_2 & \ldots \end{bmatrix}}_{z^\top}\underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \ldots \end{bmatrix}}_{z} = \frac{1}{n}z^\top z. \tag{2}$$

Let's now take a closer look at the z vector. We refill each z_i term back into the matrix. This allows us to combine all $\phi(x_i)$ into a single matrix, giving us

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \ldots \end{bmatrix} = \begin{bmatrix} (\phi(x_1)^\top w - y_1) \\ (\phi(x_2)^\top w - y_2) \\ (\phi(x_3)^\top w - y_3) \\ \ldots \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \phi(x_1)^\top \\ \phi(x_2)^\top \\ \phi(x_3)^\top \\ \ldots \end{bmatrix} w - \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \end{bmatrix} \end{bmatrix}. \tag{3}$$

At this point, note that we now have the feature map Matrix Φ and the label vector y.

$$\Phi = \begin{bmatrix} \phi(x_1)^\top \\ \phi(x_2)^\top \\ \phi(x_3)^\top \\ \ldots \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \end{bmatrix} \tag{4}$$

This allows us to represent the equation more compactly as

$$z = (\Phi w - y) \quad \implies \quad \frac{1}{n}z^\top z = \frac{1}{n}(\Phi w - y)^\top(\Phi w - y)$$

3. Expand the MSE Objective

Expand the matrix form of the MSE:

$$\mathcal{L}(w) = \frac{1}{n}\left((\Phi w)^\top \Phi w - (\Phi w)^\top y - y^\top(\Phi w) + y^\top y\right) \tag{5}$$

Simplify using the fact that $(\Phi w)^\top y = y^\top \Phi w$ (since both are scalars):

$$\mathcal{L}(w) = \frac{1}{n}\left(w^\top \Phi^\top \Phi w - 2w^\top \Phi^\top y + y^\top y\right) \tag{6}$$

4. Take the Derivative of MSE with Respect to w

To find the optimal w, we take the derivative of the MSE with respect to w and set it to 0:

$$\frac{\partial \mathcal{L}(w)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{n}\left(w^\top \Phi^\top \Phi w - 2w^\top \Phi^\top y + y^\top y\right)\right) \tag{7}$$

Compute the derivative term by term:

$$\frac{\partial}{\partial w}(w^\top \Phi^\top \Phi w) = 2\Phi^\top \Phi w \tag{8}$$

$$\frac{\partial}{\partial w}(-2w^\top \Phi^\top y) = -2\Phi^\top y \tag{9}$$

$$\frac{\partial}{\partial w}(y^\top y) = 0 \tag{10}$$

Thus, the derivative of the MSE is:

$$\frac{\partial \mathcal{L}(w)}{\partial w} = \frac{1}{n}\left(2\Phi^\top \Phi w - 2\Phi^\top y\right) \tag{11}$$

5. Set the Derivative to 0 and Solve for w

Set the derivative to 0 to find the optimal w:

$$\frac{1}{n}\left(2\Phi^\top \Phi w - 2\Phi^\top y\right) = 0 \tag{12}$$

Multiply through by n and divide by 2:

$$\Phi^\top \Phi w - \Phi^\top y = 0 \tag{13}$$

Rearrange to solve for w:

$$\Phi^\top \Phi w = \Phi^\top y \tag{14}$$

Finally, solve for w:

$$w = (\Phi^\top \Phi)^{-1}\Phi^\top y \tag{15}$$

6. Closed-Form Solution

The closed-form solution for the linear regression parameters is:

$$w = (\Phi^\top \Phi)^{-1}\Phi^\top y \tag{16}$$

This is the optimal value of w that minimizes the MSE objective.

Notes

- $\Phi^\top \Phi$ must be invertible for this solution to exist. If $\Phi^\top \Phi$ is not invertible, regularization (e.g., ridge regression) or pseudoinverses can be used.
- This derivation assumes that the data is centered (mean of y is 0) and that there is no regularization term in the objective.