Matrix Solution to Minimization Problem

The MLE solution for linear regression, the weight vector $\mathbf{w}$, can be solved analytically using the label vector $\mathbf{y}$ and data matrix $\mathbf{X}$. Recall that the objective with MLE was to find:

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

Before you can derive a closed form solution, you must first change this optimization to matrix notation. Stack your n data points horizontally to form an $n \times d$ data matrix and your labels into a single $n \times 1$ vector:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

☆ **Key Points**

If you use MLE to fit your model, you find the weight vector that minimizes the mean squared error of the predicted y values.

To find the minimum, take the gradient of this expression with respect to weight vector $\mathbf{w}$ and set it to zero.

Solving for $\mathbf{w}$, you get a final expression in terms of $\mathbf{X}$ *and* $\mathbf{y}$ that can be easily evaluated with matrix multiplication in NumPy.

This notation allows us to rewrite the squared-loss objective function as:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \frac{1}{n} (\mathbf{X}\mathbf{w} - \mathbf{y})^2$$

Closed Form Solution

To start, let's expand the objective function:

$$\begin{aligned}(\mathbf{X}\mathbf{w} - \mathbf{y})^2 &= (\mathbf{X}\mathbf{w} - \mathbf{y})^\top (\mathbf{X}\mathbf{w} - \mathbf{y}) \\ &= (\mathbf{X}\mathbf{w})^\top (\mathbf{X}\mathbf{w}) - (\mathbf{X}\mathbf{w})^\top \mathbf{y} - \mathbf{y}^\top (\mathbf{X}\mathbf{w}) + \mathbf{y}^\top \mathbf{y} \\ &= \mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\mathbf{w} - 2\mathbf{w}^\top (\mathbf{X}^\top \mathbf{y}) + \mathbf{y}^\top \mathbf{y}\end{aligned}$$

You can find the minimizer of this final term by taking the gradient with respect to $\mathbf{w}$ and equating the resulting expression with zero, namely:

$$2(\mathbf{X}^\top \mathbf{X})\mathbf{w} - 2\mathbf{X}^\top \mathbf{y} = 0$$

Using some elementary algebra, you can obtain a closed form matrix solution for $\mathbf{w}$. If your multivariate calculus is a little bit rusty, you might find the **matrix cookbook (https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf)** to be helpful.

$$\begin{aligned} 2(\mathbf{X}^\top \mathbf{X})\mathbf{w} &= 2\mathbf{X}^\top \mathbf{y} \\ \mathbf{w} &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y} \end{aligned}$$