

Linear Regression: Pseudo Inverse - Gradient Descent Alternative

If you are familiar with the concept of *Pseudo Inverse* in Linear Algebra, the parameters θ can be obtained by this formula:

$$\theta = (X^T X)^{-1} X y$$

In Multivariate Linear Regression, the formula is the same as above. But, what if the Normal Equation is non-invertible? Then consider deleting redundant features or using the regularization. The advantages in using Normal Equation are:

- No need to choose learning rate α
- No need to iterate
- Feature scaling is not necessary

If the number of features is very large, the Normal Equation works extremely slow, while the Gradient Descent Algorithm still works well.