Multivariate Ordinary Least Squares Derivation

Model

y = X beta + epsilon

Residual

e = y - X beta

Objective Function

J(beta) = (y - X beta)^T (y - X beta)

Expanded Form

J(beta) = y^T y - 2 beta^T X^T y + beta^T X^T X beta

Gradient with respect to beta

dJ/dbeta = -2 X^T y + 2 X^T X beta

Optimality Condition

X^T X beta = X^T y

OLS Solution

beta = (X^T X)^(-1) X^T y

Speed Comparison

OLS Normal Equation

Time complexity: O(n p squared + p cubed)

Gradient Descent

Update rule: beta(t+1) = beta(t) - alpha times gradient

Gradient computation cost per iteration: O(n p)

Total cost: O(k n p)

Conclusion

OLS is faster for small number of features.

Gradient descent is faster for large datasets and high dimensional data.