Ordinary Least Squares derivation

TODO: derive optimization condition for OLS regression

Linear regression model:

$$y = \omega_1 x_1 + \omega_2 x_2 + \dots$$

Goal is to minimize mean squared error:

$$\text{MSE} = \frac{1}{N} \sum (y_{true} - y)^2$$

We can write the linear model in the form of a matrix equation:

$$\begin{aligned} Y &= X\omega \\ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} &= \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{k1} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{bmatrix} \end{aligned}$$

representing a dataset with k variables and n rows.

The loss term (squared error) then becomes:

$$L = ||Y - X\omega||^2$$

(remember: Y are the true values, X is the dataset of features, ω are the weights to be learned)

This can be re-written as:

$$L = (Y - X\omega)^T (Y - X\omega)$$

Now optimize by taking the derivative and setting it equal to zero:

$$\begin{aligned} \frac{\partial L}{\partial \omega} &= 2(Y - X\omega)^T \frac{\partial (Y - X\omega)}{\partial \omega} \\ &= 2(Y - X\omega)^T (-X) \\ &= 0 \end{aligned}$$

Divide both sides by -2

$$(Y - X\omega)^T X = 0$$

Take the transpose of both sides and solve for ω

$$
\begin{aligned}
[(Y - X\omega)^T X]^T &= 0 \\
X^T(Y - X\omega) &= 0 \\
X^T Y - X^T X\omega &= 0 \\
X^T Y &= X^T X\omega \\
(X^T X)^{-1} X^T Y &= \omega
\end{aligned}
$$

Some useful matrix identities

$$
\begin{aligned}
\frac{\partial(u^T A\nu)}{\partial x} &= u^T A \frac{\partial \nu}{\partial x} + \nu^T A \frac{\partial u}{\partial x} \\
\frac{\partial u^T u}{\partial x} &= 2u^T \frac{\partial u}{\partial x}
\end{aligned}
$$

$$
(AB)^T = B^T A^T
$$

References

- https://towardsdatascience.com/ordinary-least-squares-ols-deriving-the-normal-equation-8da168d740c