Nguyen Vu Anh Ngoc
Machine Learning 1
September 21, 2023

Homework Week 4: Linear Regression

Question 1. Normal Equation Proof

Likelihood of the dataset is given by:

$$
\begin{aligned}
L = P(Y|X, w, \sum^{2}) &= \prod_{i=1}^{N} P(y_i|x_i, w, \sigma^2) \\
&= \prod_{i=1}^{N} N(y_i|y(x_i, w), \sigma^2) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - y(x_i,w))^2}{2\sigma^2}} \\
\log L &= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - y(x_i,w))^2}{2\sigma^2}}\right) \\
&= N \log \frac{1}{\sqrt{2\pi\sigma^2}} - \Sigma_{i=1}^{N} \frac{(y_i - y(x_i, w))^2}{2\sigma^2}
\end{aligned}
$$

To maximize the likelihood is to minimize the loss function:

$$
max_w \log L = -max_w \Sigma_{i=1}^{N} \frac{(y_i - y(x_i, w))^2}{2} = min_w \Sigma_{i=1}^{N} \frac{(y_i - y(x_i, w))^2}{2}
$$

Loss function:

$$
\begin{aligned}
J(W) &= \sum_{i=1}^{N} \frac{(y_i - y(x_i, w))^2}{2} \\
&= \frac{1}{2} \Sigma_{i=1}^{N} (\hat{y}_i - y_i)^2 \\
&= \frac{1}{2} (XW - Y)^T (XW - Y) \\
\nabla w J(W) &= \frac{1}{2} \nabla w (XW - Y)^T (XW - Y) \\
&= \frac{1}{2} \nabla w (W^T X^T X W - W^T X^T Y - T^T W X - Y^T Y) \\
&= \frac{1}{2} [X^T X W + X^T X W - X^T Y - X^T Y] \\
&= X^T X W - X^T Y
\end{aligned}
$$

Loss function reaches minimum when $\nabla w J(W) = 0$:

$$
\begin{aligned}
\nabla w J(W) = 0 &\leftrightarrow \; X^T X W - X^T Y = 0 \\
&\leftrightarrow W = (X^T X)^{-1} X^T Y
\end{aligned}
$$

Question 2. Prove if the matrix X is of full rank, then the square matrix X^TX is invertible, denoted as $(X^TX)^{-1}$.

Let X be an $m \times n$ matrix with full rank, where $m \geq n$.

X^TX is **symmetric**: $(X^TX)^T = X^T(X^T)^T = X^TX$

X^TX is **positive semi-definite**: For any vector $v \in \mathbb{R}^n$, the quadratic form $v^T(X^TX)v = (Xv)^T(Xv) = ||Xv||^2$ is always non-negative. This implies that all the eigenvalues of X^TX are non-negative.

$Xv = 0$ (and hence $||Xv||^2 = 0$) if and only if the columns of X are linearly dependent (X is full rank), so if X has full column rank then X^TX is positive definite.

When X is of full rank, the matrix X^TX is invertible, and $(X^TX)^{-1}$ exists.