Problem Set 1

1 Conditions for Normal Equation

Prove the following theorem: The matrix A^TA is invertible if and only if the columns of A are linearly independent.

Solution: We first prove A^TA is invertible if the columns of A are linear independent. We denote by $\mathcal{N}(A)$ the null space of A. Due to the linear independence of A, $\mathcal{N}(A) = \{\vec{0}\}$ where $\vec{0}$ is a vector where all elements are zeros. Let $\mathcal{N}(A^TA)$ be the null space of the square matrix A^TA. For each $v \in \mathcal{N}(A^TA)$, we have $A^TAv = \vec{0}$, and thus $v^TA^TAv = 0$. Therefore, $(Av)^TAv = \|Av\|_2^2 = 0$, which implies $Av = \vec{0}$. It is demonstrated that, if $v \in \mathcal{N}(A^TA)$, then $Av = \vec{0}$ and $v \in \mathcal{N}(A)$. Recalling $\mathcal{N}(A) = \{\vec{0}\}$, we have $v = \vec{0}$. Finally, we have $\mathcal{N}(A^TA) = \{\vec{0}\}$. Hence, the columns of A^TA is linearly independent with each other, and we conclude that A^TA is invertible.

We then prove that if A^TA is invertible, the columns of A are linearly independent. Since A^TA is invertible, we have $\mathcal{N}(A^TA) = \{\vec{0}\}$. For $\forall v \in \mathcal{N}(A)$, we have $Av = \vec{0}$ and hence $A^TAv = \vec{0}$. Therefore, v also belongs to the null space of A^TA, i.e., $v \in \mathcal{N}(A^TA)$. Recalling $\mathcal{N}(A^TA) = \{\vec{0}\}$, then $v = \vec{0}$. We are done.

2 Regularized Normal Equation for Linear Regression

Given a data set $\{x^{(i)}, y^{(i)}\}_{i=1,\cdots,m}$ with $x^{(i)} \in \mathbb{R}^n$ and $y^{(i)} \in \mathbb{R}$, the general form of regularized linear regression is as follows

$$\min_{\theta} \frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right] \tag{1}$$

Derive the normal equation.

Solution: The objective function can be reformulated as

$$J(\theta) = \frac{1}{2m}\left[(X\theta - \vec{y})^T(X\theta - \vec{y}) + \lambda\theta^TL\theta\right]$$

where $X = \begin{bmatrix} -x^{(1)}- \\ \vdots \\ -x^{(m)}- \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}$ and $L = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$. Therefore, we have

$$\begin{aligned}
\nabla_\theta J(\theta) &= \frac{1}{2m}\left[\nabla_\theta (X\theta - \vec{y})^T (X\theta - \vec{y}) + \nabla_\theta \lambda \theta^T L\theta)\right] \\
&= \frac{1}{2m}[\nabla_\theta (\theta^T X^T X\theta - \theta^T X^T \vec{y} - \vec{y}^T X\theta + \vec{y}^T \vec{y}) + \nabla_\theta \lambda \theta^T L\theta)] \\
&= \frac{1}{2m}(2X^T X\theta - 2X^T \vec{y} + 2\lambda L\theta) \\
&= \frac{1}{m}(X^T X\theta - X^T \vec{y} + \lambda L\theta) \\
&= 0
\end{aligned}$$

Hence,

$$\theta = (X^T X + \lambda L)^{-1} X^T \vec{y}$$