ECE5984 Homework1

소프트웨어학과 2022710836 김민근

1. cost function of the linear regression

(a).

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^T x^{(i)} - y^{(i)})^2$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^T x^{(i)})^2 - 2\left(\theta^T x^{(i)} y^{(i)}\right) + (y^{(i)})^2$$

Let's partial derivative about θ.

$$\frac{\partial J}{\partial \theta} = \sum_{i=1}^{m}(x^{(i)})^2\theta^T - x^{(i)}y^{(i)}$$

$$\frac{\partial^2 J}{\partial \theta^2} = \sum_{i=1}^{m}(x^{(i)})^2$$

$$\sum_{i=1}^{m}\left(x^{(i)}\right)^2 > 0$$

∴ J(θ) can be written in a quadratic form, which is convex.

(b).

Let proof of $J(\lambda\theta_1 + (1-\lambda)\theta_2) - \lambda J(\theta_1) - (1-\lambda)J(\theta_2) \leq 0$ … Let **(A)**

About $X \in \mathbb{R}^{m\times d}$, $Y \in \mathbb{R}^m$, $X = \begin{bmatrix} \cdots & (x^{(1)})^T & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & (x^{(m)})^T & \cdots \end{bmatrix}$, $Y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}$,

$$J(\theta) = \frac{1}{2}(\theta^T X^T X\theta - 2Y^T X\theta + Y^T Y)$$

-> $\frac{1}{2}(-\ 2Y^T X\theta + Y^T Y)$ will be 0.

(Because $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$ in any linear function)

Left side of **(A)**:

$$\frac{1}{2}(\lambda\theta_1 - (1-\lambda)\theta_2)^T X^T X(\lambda\theta_1 - (1-\lambda)\theta_2) - \frac{1}{2}\lambda\theta_1^T X^T X\theta_1 - \frac{1}{2}(1-\lambda)\theta_2^T X^T X\theta_2$$

$$= \frac{1}{2}\lambda^2\theta_1^T X^T X\theta_1 + \frac{1}{2}(1-\lambda)^2\theta_2^T X^T X\theta_2 + \lambda(1-\lambda)\theta_1^T X^T X\theta_2 - \frac{1}{2}\lambda\theta_1^T X^T X\theta_1 - \frac{1}{2}(1-\lambda)\theta_2^T X^T X\theta_2$$

$$= \frac{1}{2}(\lambda^2-\lambda)\theta_1^T X^T X\theta_1 + \frac{1}{2}\big((1-\lambda)^2-(1-\lambda)\big)\theta_2^T X^T X\theta_2 + \lambda(1-\lambda)\theta_1^T X^T X\theta_2$$

(Because $\theta_2^T X^T X\theta_1 = (X\theta_2)^T X\theta_1 = ((X\theta_2)^T X\theta_1)^T = (X\theta_1)^T X\theta_2 = \theta_1^T X^T X\theta_2$)

$$= -\frac{1}{2}\lambda(1-\lambda)(\theta_1^T X^T X\theta_1 + \theta_2^T X^T X\theta_2 - 2\theta_1^T X^T X\theta_2)$$

$$= -\frac{1}{2}\lambda(1-\lambda)\big((\theta_1-\theta_2)^T X^T X(\theta_1-\theta_2)\big)$$

$$= -\frac{1}{2}\lambda(1-\lambda)\left(\big(X(\theta_1-\theta_2)\big)^T X(\theta_1-\theta_2)\right)$$

$$= -\frac{1}{2}\lambda(1-\lambda)\|X(\theta_1-\theta_2)\|^2$$

$$-\frac{1}{2}\lambda(1-\lambda)\|X(\theta_1-\theta_2)\|^2 \le 0$$

∴ J(θ) is a convex function.

2. Newton's method

$$\frac{\partial J}{\partial\theta_j} = \sum_{i=1}^{m}(\theta^T x^{(i)} - y^{(i)})\, x_j^{(i)}$$

$$\frac{\partial^2 J}{\partial\theta_j\cdot\partial\theta_k} = \sum_{i=1}^{m}\frac{\partial}{\partial\theta_k}(\theta^T x^{(i)} - y^{(i)})\, x_j^{(i)}$$

$$= \sum_{i=1}^{m} x_j^{(i)}\, x_k^{(i)} = (X^T X)_{jk}$$

$$\therefore \mathrm{H} = X^T X$$

$$\nabla_\theta J(\theta) = \nabla_\theta \frac{1}{2}(X\theta - Y)^T(X\theta - Y)$$

$$= \frac{1}{2}\nabla_\theta((X\theta)^T X\theta - (X\theta)^T Y - Y^T(X\theta) + Y^T Y)$$

$$= \frac{1}{2}\nabla_\theta\big(\theta^T(X^T X)\theta - Y^T(X\theta) - Y^T(X\theta)\big)$$

$$= \frac{1}{2}\nabla_\theta(\theta^T(X^T X)\theta - 2(X^T Y)^T\theta)$$

$$= \frac{1}{2}(2X^T X\theta - 2X^T Y)$$

$$= X^T X\theta - X^T Y$$

By Newton's method in this problem,

$$\theta^{(t+1)} \leftarrow \theta^t - (X^T X)^{-1}\nabla_\theta(J(\theta^{(t)}))$$

About any $\theta^{(0)}$,

$$\theta^{(1)} = \theta^0 - (X^TX)^{-1}\nabla_\theta(J(\theta^{(0)}))$$
$$= \theta^0 - (X^TX)^{-1}(X^TX\theta^{(0)} - X^TY)$$
$$= \theta^0 - \theta^0 + (X^TX)^{-1}X^TY$$
$$= (X^TX)^{-1}X^TY$$

∴ One iteration of Newton's method gives us the solution.

3. Positive definite matrices

(a).

Let $x \in \mathbb{R}^n$ be any non-zero vectors

$$x^TIx = x^Tx = \sum_{i=1}^{n}(x_i)^2 > 0$$

And the identity matrix I is symmetric ($I^T = I$)

∴ I is positive definite matrix.

(b).

Let $x \in \mathbb{R}^n$

$$x^TAx = x^Tzz^Tx = (z^Tx)^Tz^Tx = \sum_{i=1}^{n}(z_i^Tx_i)^2 \geq 0$$

And z^Tx is 1×1 matrix, so it is symmetric

∴ A is positive semidefinite.

(c).

BAB^T is PSD.

Let $x \in \mathbb{R}^n$

$$x^T(BAB^T)x = (B^Tx)^TA(B^Tx) \geq 0$$

(Because A is PSD, always $z^TAz \geq 0$ about $x \in \mathbb{R}^n$)

∴ $x^T(BAB^T)x \geq 0$

And

$$(BAB^T)^T = BA^TB^T = BAB^T$$

(Because A is PSD, A is symmetric)

$\therefore BAB^T$ is symmetric

$\therefore BAB^T$ is positive semidefinite.

(d).

$$J(\theta) = -\sum_{i=1}^{N} y^{(i)} log \left(\sigma(\theta^T x^{(i)})\right) + \sum_{i=1}^{N}(1-y^{(i)}) log \left(1-\sigma(\theta^T x^{(i)})\right)$$

In case $y^{(i)} = 1$)

$$\nabla_{\theta_j}(-\sum_{i=1}^{N} y^{(i)} log \left(\sigma(\theta^T x^{(i)})\right))$$

$$= -\sum_{i=1}^{N} \nabla_{\theta_j}(y^{(i)} log \left(\sigma(\theta^T x^{(i)})\right))$$

$$= -y^{(i)}.\sum_{i=1}^{N} \frac{1}{\sigma(\theta^T x^{(i)})} \nabla_{\theta_j}\sigma(\theta^T x^{(i)})$$

$$= \sum_{i=1}^{N} y^{(i)} \left(1-\sigma(\theta^T x^{(i)})\right) x_j^{(i)} \qquad \ldots Let \ (A)$$

(Because $\nabla\sigma(z) = \sigma(z)(1-\sigma(z))$)

In case $y^{(i)} = 0$)

$$\nabla_{\theta_j}\sum_{i=1}^{N}(1-y^{(i)}) log \left(1-\sigma(\theta^T x^{(i)})\right)$$

$$= \left(1-y^{(i)}\right).\sum_{i=1}^{N} \nabla_{\theta_j}\left(log \left(1-\sigma(\theta^T x^{(i)})\right)\right)$$

$$= \left(1-y^{(i)}\right).\sum_{i=1}^{N} \frac{1}{1-\sigma(\theta^T x^{(i)})} \nabla_{\theta_j}\left(1-\sigma(\theta^T x^{(i)})\right)$$

$$= \sum_{i=1}^{N}(1-y^{(i)})(-\sigma(\theta^T x^{(i)}) x_j^{(i)}) \qquad \ldots Let \ (B)$$

$$\therefore \frac{\partial J}{\partial \theta_j} = \sum_{i=1}^{N} x_j^{(i)}\left(\sigma\left(\theta^T x^{(i)}\right) - y^{(i)}\right) \quad \ldots \left(\text{by (A)} + \text{(B)}\right)$$

$$\frac{\partial^2 J}{\partial \theta_j \cdot \partial \theta_k} = \sum_{i=1}^{N} \nabla_{\theta_k}(x_j^{(i)}\left(\sigma\left(\theta^T x^{(i)}\right) - y^{(i)}\right))$$

$$= \sum_{i=1}^{N} x_j^{(i)} x_k^{(i)} . \sigma\left(\theta^T x^{(i)}\right)\left(1 - \sigma\left(\theta^T x^{(i)}\right)\right)$$

$$\therefore \frac{\partial^2 J}{\partial \theta \cdot \partial \theta^T} = H(\theta) = \sum_{i=1}^{N} x^{(i)}\left(x^{(i)}\right)^T . \sigma\left(\theta^T x^{(i)}\right)\left(1 - \sigma\left(\theta^T x^{(i)}\right)\right)$$

Let $XX^T = \sum_{i=1}^{N} x^{(i)}\left(x^{(i)}\right)^T$,

Let $A_{(i)} = \sigma\left(\theta^T x^{(i)}\right)\left(1 - \sigma\left(\theta^T x^{(i)}\right)\right)$

(The scalar terms are combined in a diagonal matrix A)

$$\left(H(\theta)\right)^T = (XAX^T)^T = XA^T X^T = XAX^T \ \ldots Let \ (C)$$

(Because A is a diagonal matrix)

$$x^T H(\theta) x = x^T XAX^T x = x^T XA^{\left(\frac{1}{2}\right)}\left(x^T XA^{\left(\frac{1}{2}\right)}\right)^T \geq 0 \ \ldots Let \ (D)$$

H is symmetric by (C), and H ≥ 0 by (D)

$\therefore$ The Hessian H of the cost function is positive semidefinite.