Jake Schinto

Homework 3

Linear Regression

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

$J(\theta) = \frac{1}{4}\left((0-(\theta_0+\theta_1+2\theta_2))^2 + (-1-(\theta_0+2\theta_1+2\theta_2))^2\right)$

$= \frac{1}{4}(\theta_0^2 + 2\theta_0\theta_1 + \theta_1^2 + 4\theta_0\theta_2 + 4\theta_1\theta_2 + 4\theta_2^2 +$
$1 + 2\theta_0 + \theta_0^2 + 4\theta_1 + 4\theta_0\theta_1 + 4\theta_1^2 + 4\theta_2 + 4\theta_0\theta_2 + 8\theta_1\theta_2 + 4\theta_2^2)$

$= \frac{1}{4}(1 + 2\theta_0^2 + 6\theta_0\theta_1 + 5\theta_1^2 + 8\theta_0\theta_2 + 12\theta_1\theta_2 + 8\theta_2^2 + 2\theta_0 + 4\theta_1 + 4\theta_2)$

2. $J(\theta) = 0$ if $y^{(i)} = h(x^{(i)})$

$0 = \theta_0 + \theta_1 + 2\theta_2$ $\quad -1 = \theta_0 + 2\theta_1 + 2\theta_2$

$\theta_0 = -\theta_1 - 2\theta_2$ $\quad -1 = -\theta_1 - 2\theta_2 + 2\theta_2 + 2\theta_1$

$\theta_1 = -1 \rightarrow 1 = \theta_0 + 2\theta_2$

if $\theta_0 = 0 \rightarrow \theta_2 = 1/2$ $\quad h_1(x) = -x_1 + x_2/2$

if $\theta_2 = 0 \rightarrow \theta_0 = 1$ $\quad h_2(x) = 1 - x_1$

3. $A^+ = (AA^T)^{-1}A$ $\quad$ let $AA^T = B$ where B is $n \times n$

$A^+ = B^{-1}A$

$A^+ \cdot A^T = B^{-1}A \cdot A^T = B^{-1} \cdot B$

$= I_{n \times n} = A^+ \cdot A^T$

4. For $A \cdot A^T$ to be invertible, it must be a square matrix with Linearly independant columns.

Linear Independant of cols $= N(A^T) = \{\vec{0}\}$

for $\vec{v} \in N(AA^T) \rightarrow v^T A A^T v = v^T \vec{0} = 0$

$v^T \cdot A = (A^T v)^T \rightarrow (A^T v)^T A^T v = (A^T v)(A^T v) = \|A^T \cdot v\|^2 = 0$

Therefore, $N(AA^T) = N(A^T)$ and AA^T will be linearly independant if A^T is as well.

a. if X^T is not full column rank then A^T cannot be linnearly independant. Because of this, AA^T cannot be LI as well and not invertible

b. if $m < n$; there are more columns than rows. Again it is impossible for A^T to be LI if there are more columns, so AA^T cannot be inverted in this case.

5. $X = m \times 2 =$

$$\begin{array}{cc} X_1 & X_2 \\ \hline X_1^1 & \alpha X_1^1 \\ X_1^2 & \alpha X_1^2 \\ X_1^3 & \alpha X_1^3 \\ \vdots X_1^m & \alpha X_1^m \end{array} \rightarrow$$

divide each row by $X_1^{(n)}$ where n is the row

$$\begin{bmatrix} 1 & \alpha \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \leftarrow \begin{bmatrix} 1 & \alpha \\ 1 & \alpha \\ 1 & \alpha \end{bmatrix}$$

$\text{Rank}(X) = 1$, which is < 2

6. a. $P(y^{(i)} | x^{(i)}; \theta, \sigma) = N(\theta^T x^{(i)}, \sigma)$

likelyhood function =

$$\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma}} e^{\left\{\frac{-(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma}\right\}}$$

b. $$= \sum_{i=1}^{m} \ln(N(y^{(i)} | \theta^T x^{(i)}, \sigma))$$

$$= \underbrace{\frac{m}{2}\ln(\sigma) - \frac{m}{2}\ln(2\pi)}_{\text{rest}} - \underbrace{\sigma\left(\frac{1}{2}\sum_{i=1}^{m}(y^{(i)} - \theta^T x^{(i)})^2\right)}_{\text{Error function}}$$

Since the error function is subtracted within the log-likelyhood, you would need to minimize it in order to maximize likelyhood. Also, θ only apears within the error function part, so when taking the derivative wrt θ the rest is treated as a constant, therefore the θ that maxes likelyhood is the same that minimizes error cost.

Logistic Regression

1. $f(x) = \frac{1}{1+e^{-x}}$ $\quad 1 - \frac{1}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}$

$= \frac{e^{-x}}{1+e^{-x}} \cdot \frac{e^x}{e^x} = \frac{1}{e^x+1} = \frac{1}{1+e^x} = 1 - f(x)$

$\frac{d}{dx} f(x) = \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) = \frac{(1+e^{-x})(0) - (1)(-e^{-x})}{(1+e^{-x})^2}$

$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{(1+e^{x})(1+e^{-x})} = \frac{1}{(1+e^x)} \cdot \frac{1}{(1+e^{-x})} = f(x) \cdot (1 - f(x))$

2. $\tanh(x) = 2f(2x) - 1 = \frac{2}{1+e^{-2x}} - 1 = \frac{2-1-e^{-2x}}{1+e^{-2x}}$

$= \frac{1-e^{-2x}}{1+e^{-2x}} \cdot \frac{e^{2x}}{e^{2x}} = \frac{e^{2x}-1}{e^{2x}+1} = \tanh(x)$

3. $P(Y=1|X) = P(Y=1)P(X|Y=1)/P(X)$

$P(Y=0|X) = P(Y=0)P(X|Y=0)/P(X)$

$\frac{P(Y=1|X)}{P(Y=0|X)} = \frac{P(Y=1)P(X|Y=1)}{P(Y=0)P(X|Y=0)}$

$\ln\left(\frac{P(Y=1)P(X|Y=1)}{P(Y=0)P(X|Y=0)}\right) = \ln\left(\frac{P(Y=1)}{P(Y=0)}\right) + \ln\left(\frac{P(X|Y=1)}{P(X|Y=0)}\right)$

$= \underbrace{\ln\left(\frac{P(Y=1)}{P(Y=0)}\right)}_{w_0} + \underbrace{\sum_{i=1}^{n} \ln\left(\frac{P(X_i|Y=1)}{P(X_i|Y=0)}\right)}_{\sum_{i=1}^{n} \ln\left(\frac{\theta^T X}{(1-\theta^T)X}\right)}$

$= w_0 + \sum_{i=1}^{n} \theta^T X$

4. $P(w|\mu, \sigma^2 I) \prod_{i=1}^{m} P(y^{(i)}|x^{(i)}, w)$

$\ln(P(w|\mu, \sigma^2 I)) + \sum_{i=1}^{m} \ln(p(y^{(i)}|x^{(i)}, w))$

$\ln\left(\frac{1}{\sqrt{2\pi\sigma^2 I}} e \left\{-\frac{(w-\mu)^2}{2\sigma^2}\right\}\right) + \frac{1}{2}\sum_{i=1}^{m} \ln(y^{(i)} - wx^{(i)})^2$

$$-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(w-\mu)^2}{2\sigma^2} + \frac{1}{2m}\sum_{i=1}^{m}(y^{(i)} - wx)^2$$

$$\frac{d}{dw} = \frac{-(w-\mu)}{\sigma^2} - \frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - wx^{(i)})^2$$

Convex Optimization

1. $\frac{df}{dw_1} = w_1 + 2w_2$ $\qquad$ $\frac{df}{dw_2} = w_2 + 2w_1$

$$H = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \det\left[\lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\right) = 0$$

$$\det\left(\begin{bmatrix} \lambda-1 & -2 \\ -2 & \lambda-1 \end{bmatrix}\right) \quad (\lambda-1)^2 - 4 = 0$$

$$\lambda^2 - 2\lambda + 1 - 4 = 0 = \lambda^2 - 2\lambda - 3$$

$$\lambda = \frac{2 \pm \sqrt{4+12}}{2} = 1 \pm 2 = 3, -1$$

it has a negative eiganvalue, so it is not convex

2. $SE = \sum (y^{(i)} - \theta^T x^{(i)})^2$

$$(y^{(i)} - \theta^T x^{(i)})^2 = y^{2(i)} - 2y^{(i)}\theta^T x^{(i)} + \theta^{T^2} x^{(i)^2}$$

$$\frac{d}{d\theta} = 2y^{(i)}x^{(i)} + 2\theta^T x^{(i)^2} \Rightarrow \frac{d}{d\theta} = 2x^{(i)^2}$$

the second derivative is always positive, and its differentiable everywhere. Therefore it must be convex

3. $J(w) = -y\left(w_0 + \sum_j^n w_j x_j\right) + \ln\left(1 + e^{\{w_0 + \sum_j^n w_j x_j\}}\right)$

$$\frac{dJ}{dw} = -y\left(\sum_j^n x_j\right) + \left(\sum_j^n x_j\right) e^{\{w_0 + \sum_j^n w_j x_j\}} \cdot \frac{1}{1+e^{\{w_0 + \sum_j^n w_j x_j\}}}$$

$$= \left(\sum_j^n x_j\right)\left(-y + \frac{e^{\{\}}}{1+e^{\{\}}}\right)$$

$$\frac{dJ}{dx} = \left(\sum_j^n w_j\right)\left(\quad \right)$$

$$H = \begin{bmatrix} \dfrac{\left(\sum_j^n x_j\right)^2 e^{\{\}}}{(1+e^{\{\}})^2} & \text{same} \\ -yn + \dfrac{\left(\sum_j^n w_j\right)\left(\sum_j^n x_j\right) e^{\{\}} + n e^{\{\}}}{(1+e^{\{\}})^2} & \dfrac{\left(\sum_j^n w_j\right)^2 e^{\{\}}}{(1+e^{\{\}})^2} \end{bmatrix}$$

$$\lambda^2 - \lambda H_{11} - \lambda H_{22} + H_{11}H_{22} - 2H_{12} = 0$$

$$= \lambda^2 - \lambda\left(\frac{e^{\{\}}\left(\sum_j^n x_j^2 + \sum_j^n w_j^2\right)}{(1+e^{\{\}})^2}\right) + \frac{\left(\sum_j^n w_j\right)^2\left(\sum_j^n x_j\right)^2 e^{\{\}\cdot 2}}{(1+e^{\{\}})^4} + 2yn - \frac{2\left(\sum_j^n w_j\right)\left(\sum_j^n x_j\right) e^{\{\}} + 2n e^{\{\}}}{(1+e^{\{\}})^2}$$

$$\lambda = \frac{\dfrac{e^{\{\}}\left(\sum_j^n x_j^2 + \sum w_j^2\right)}{(1+e^{\{\}})^2} \pm \sqrt{\dfrac{e^{2\{\}}\left(\sum x_j^2 + \sum w_j^2\right)^2}{(1+e^{\{\}})^4} - 4\left(H_{11}H_{22} - 2H_{12}\right)}}{2}$$

if $\lambda \geq 0$ then H is positive semi-definite and is convex