MLHW

Q1: (a) Symmetric matrix $M \in \mathbb{R}^n$ is a positive semi-definite if $\forall x \in \mathbb{R}^n$ $x^T M x \geq 0$, given matrix $A \in \mathbb{R}^{n \times n}$. Show that AA^T is a positive semedefinite

$\Rightarrow$ Assume a matrix $v \in \mathbb{R}^n$

$v^T A A^T v \geq 0$

$= (A^T v)^T (A^T v) = \|A^T v\|^2$

Because of the power of the matrix, $\|A^T v\|$ is alway $\geq 0 \ \forall v$

Then AA^T is positive semi-definite.

(b) If $f(x_1, x_2) = x_1 \sin(x_2) e^{-x_1 x_2}$, $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = ?$

$$\frac{\partial f}{\partial x_1} = \sin(x_2)\left[e^{-x_1 x_2} + x_1 \cdot e^{-x_1 x_2} \cdot (-x_2)\right]$$

$$= \boxed{e^{-x_1 x_2} \cdot \sin(x_2) \cdot (1 - x_1 x_2)} \ *$$

$$\frac{\partial f}{\partial x_2} = x_1\left[\cos(x_2) e^{-x_1 x_2} + \sin(x_2) e^{-x_1 x_2} \cdot (-x_1)\right]$$

$$= \boxed{x_1 e^{-x_1 x_2}\left[\cos(x_2) - x_1 \sin(x_2)\right]} \ *$$

(c) Given $f(x; p) = p^x (1-p)^{1-x}$ for $x \in \{0, 1\}$

The MLE of p is:

(1) construct $L(p) = f(x_1; p) f(x_2; p) \cdots f(x_n; p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}$

$= p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}$

(2) $\ln(L(p)) = \ln(p^{\sum_{i=1}^n x_i}) + \ln((1-p)^{n - \sum_{i=1}^n x_i}) = \sum_{i=1}^n x_i \ln(p) + (n - \sum_{i=1}^n x_i) \ln(1-p)$

given $\sum_{i=1}^n x_i = n\bar{x}$ where $\bar{x}$ is average of all sample

(3) $\frac{d}{dp} \ln(L(p)) = n\bar{x} \cdot \frac{1}{p} + (n - n\bar{x}) \cdot \frac{-1}{1-p}$, suppose $\frac{d}{dp} \ln(L(p)) = 0$

$\Rightarrow \frac{n\bar{x}}{p} = \frac{n - n\bar{x}}{1-p} \Rightarrow \boxed{p = \bar{x}} \ *$

make this correct

examine the 2nd derivative (4) $\frac{d^2}{dp^2} \ln(L(p)) = \frac{-n\bar{x}}{p^2} + \frac{-(n - n\bar{x})}{(1-p)^2}$, substitude $p = \bar{x} \Rightarrow \frac{-n}{\bar{x}} - \frac{n(\cancel{1-\bar{x}})}{(1-\bar{x})^{\cancel{2}}} = \frac{-n}{\bar{x}} - \frac{n}{1-\bar{x}}$

because of $x \in \{0, 1\}$, $0 \leq \bar{x} \leq 1$ and $1 - \bar{x} \geq 0 \Rightarrow \frac{d^2}{dp^2} \ln(L(p)) < 0$

ML HW

Q2, Linear Regression model $\vec{y} = \vec{X}\vec{\theta} + \epsilon$ where $\vec{y} \in \mathbb{R}^n$, $\vec{X} \in \mathbb{R}^{n\times d}$, $\vec{\theta} \in \mathbb{R}^d$, $\epsilon \in \mathbb{R}^n$

(a) Find general form of θ^* that minimize the weighted MSE

Given: $L(\theta) = (y - X\theta)^T \Omega (y - X\theta)$

$$= (y^T - \theta^T X^T)\Omega(y - X\theta) = (y^T\Omega - \theta^T X^T \Omega)(y - X\theta)$$

$$= y^T\Omega y - \underbrace{\theta^T X^T \Omega y}_{③} - \underbrace{y^T \Omega X \theta}_{②} + \underbrace{\theta^T X^T \Omega X \theta}_{①} \rightarrow \text{2nd order term}$$

$$= (\theta - \phi)^T X^T \Omega X (\theta - \phi) + \text{complementary term}$$

↖ shorted as CT

$$= (\theta^T - \phi^T)(X^T\Omega X)(\theta - \phi) + CT$$

$$= (\theta^T X^T \Omega X - \phi^T X^T \Omega X)(\theta - \phi) + CT$$

$$= \underbrace{\theta^T X^T \Omega X \theta}_{\triangle 1} - \underbrace{\phi^T X^T \Omega X \theta}_{\triangle 2} - \underbrace{\theta^T X^T \Omega X \phi}_{\triangle 3} + \phi^T X^T \Omega X \phi + CT$$

the same as the last term

$$\left.\begin{array}{l} \triangle 1 = ① \\ \triangle 2 = ② \\ \triangle 3 = ③ \end{array}\right\} \quad \theta^T X^T \Omega y = \theta^T X^T \Omega X \phi \Rightarrow \phi = (X^T\Omega X)^{-1} X^T \Omega y$$

put into $\triangle 2 = ②$ $\Rightarrow$ $\theta^T X^T \Omega X \theta = \theta^T X^T \Omega X (X^T\Omega X)^{-1} X^T \Omega y = \theta^T X^T \Omega y$

the same with ②

$$CT = y^T\Omega y - \phi^T X^T \Omega X \phi = y^T \Omega y - y^T \Omega^T X [(X^T\Omega X)^{-1}]^T \cancel{X^T \Omega X (X^T \Omega X)^{-1}} X^T \Omega y$$

$$= y^T\Omega y - y^T \Omega^T X [(X^T\Omega X)^{-1}]^T X^T \Omega y$$

$$\boxed{\begin{array}{l} L(\theta) = (\theta - \phi)^T X^T \Omega X (\theta - \phi) + y^T\Omega y - y^T\Omega^T X[(X^T\Omega X)^{-1}]^T X^T \Omega y \\ \quad \text{where } \phi = (X^T \Omega X)^{-1} X^T \Omega y \end{array}}$$ *

(b) $L(\theta) = (y_i - X_i\theta)^T (y_i - X_i\theta) + \lambda w w^T \quad \because w^T = \theta$

$$\therefore = (y_i - X_i\theta)^T (y_i - X_i\theta) + \lambda \theta^T \theta$$

$$\frac{\partial L(\theta)}{\partial \theta} \overset{①}{=} \frac{\partial (y_i - X_i\theta)}{\partial \theta}(y_i - X_i\theta) + \frac{\partial (y_i - X_i\theta)}{\partial \theta}(y_i - X_i\theta) + \frac{\partial \lambda \theta^T \theta}{\partial \theta}$$

$$= 2\left[\overset{②}{} -X_i^T (y_i - X_i\theta)\right] + \overset{③}{2\lambda\theta} \Rightarrow \text{when } \frac{\partial L(\theta)}{\partial \theta} = 0, \text{ it has extreme value}$$

$$\Rightarrow -2X_i^T y_i + 2X_i^T X_i \theta + 2\lambda\theta = 0 \Rightarrow X_i^T X_i \theta + \lambda\theta = X_i^T y_i = (X_i^T X_i + \lambda I)\theta$$

$$\boxed{\theta = (X_i^T X_i + \lambda I)^{-1} X_i^T y_i}$$ *

Cont. for Q2.b

① $\frac{\partial \vec{u}^T\vec{v}}{\partial \vec{x}} = \frac{\partial \vec{u}}{\partial \vec{x}}\vec{v} + \frac{\partial \vec{v}}{\partial \vec{x}}\vec{u} = \frac{\partial \sum_k u_k v_k}{\partial \vec{x}} = \sum_k \frac{\partial u_k v_k}{\partial \vec{x}} = \sum_k \left(u_k \frac{\partial v_k}{\partial \vec{x}} + v_k \frac{\partial u_k}{\partial \vec{x}}\right)$

$= \sum_k \frac{\partial v_k}{\partial \vec{x}} u_k + \sum_k \frac{\partial u_k}{\partial \vec{x}} v_k = \frac{\partial \vec{v}}{\partial \vec{x}}\vec{u} + \frac{\partial \vec{u}}{\partial \vec{x}}\vec{v}$

② $\frac{\partial A\vec{u}}{\partial \vec{x}} = \frac{\partial \vec{u}}{\partial \vec{x}} A^T$

$\left(\frac{\partial A\vec{u}}{\partial \vec{x}}\right)_{ij} = \frac{\partial \sum_k a_{jk} u_k}{\partial x_i} = \sum_k a_{jk} \frac{\partial u_k}{\partial x_i} = \sum_k \left(\frac{\partial \vec{u}}{\partial \vec{x}}\right)_{ik} (A^T)_{kj} = \left(\frac{\partial \vec{u}}{\partial \vec{x}} A^T\right)_{ij}$

③ $\frac{\partial \vec{x}^T\vec{x}}{\partial \vec{x}} = 2X$, Use $\frac{\partial X^TAX}{\partial X} = (A+A^T)X$ and replace A to I.

$\frac{\partial \vec{x}^T I \vec{x}}{\partial X} = \frac{\partial X^T I X}{\partial X} = \frac{\partial X}{\partial X} IX + \frac{\partial X}{\partial X} I^T X = IIX + II^T X$

$= X + X = 2X$

What is the extreme value that $\frac{\partial L(\theta)}{\partial \theta}$ is

$\frac{\partial^2}{\partial \theta^2}\left(-2X_i^T y_i + 2X_i^T X_i \theta + 2\lambda\theta\right) = 2(X_i^T X)^T + 2\lambda = 2(\|X\|^2)^T + 2\lambda > 0$

curved open up so $\frac{\partial L(\theta)}{\partial \theta} = 0$ has minimum value ✱

ML HW

Q.3: Logistic Sigmoid Function and Hyperbolic Tangent Function

Given: $\sigma(a) = \frac{1}{1+e^{-a}}$, $\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

(1) Show $\tanh(a) = 2\sigma(2a) - 1$

$$\Rightarrow \tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} = \frac{1 - e^{-2a}}{1 + e^{-2a}} = \frac{1}{1+e^{-2a}} - \frac{e^{-2a}}{1+e^{-2a}}$$

$$= \frac{1}{1+e^{-2a}} - \frac{1}{e^{2a}+1} = \sigma(2a) - \sigma(-2a)$$

$$\because \sigma(z) = 1 - \sigma(-z) \quad \therefore \tanh(a) = \sigma(2a) - 1 + \sigma(2a)$$

$$= 2\sigma(2a) - 1$$

※

(2) Given $\begin{cases} y(x, \vec{w}) = w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x-\mu_j}{s}\right) \\ y(x, \vec{u}) = u_0 + \sum_{j=1}^{M} u_j \tanh\left(\frac{x-\mu_j}{2s}\right) \end{cases}$

Assume $a = \frac{x-\mu_j}{2s}$

$$\Rightarrow w_j \sigma(a) = u_j \tanh(a) = u_j\left[2\sigma(2a) - 1\right] = u_j\left[2\frac{1}{1+e^{-2a}} - 1\right]$$

$$= w_j \frac{1}{1+e^{-a}} \Rightarrow \frac{w_j}{u_j} = 2\frac{(1+e^{-a})}{1+e^{-2a}} - (1+e^{-a}) = \frac{2(1+e^{-a}) - (1+e^{-a})(1+e^{-2a})}{1+e^{-2a}}$$

$(1+e^{-a})$

$$= \frac{\cancel{2} + 2\cancel{e^{-a}} - \cancel{1} - \cancel{e^{-a}} - e^{-2a} - e^{-3a}}{1+e^{-2a}} = \frac{1 + e^{-a} - e^{-2a} - e^{-3a}}{1+e^{-2a}}$$

$$= \frac{1-e^{-2a}}{1+e^{-2a}} + \frac{e^{-a} - e^{-3a}}{1+e^{-2a}} = \frac{e^a - e^{-a}}{e^a + e^{-a}} + e^{-a}\left(\frac{1-e^{-2a}}{1+e^{-2a}}\right)$$

$$= \tanh(a) + e^{-a}\left(\frac{e^a - e^{-a}}{e^a + e^{-a}}\right) = \tanh(a) + e^{-a}\tanh(a) = \boxed{(1+e^{-a})\tanh(a)}$$

※

$$w_j \sigma(2a) = u_j \tanh(a) = u_j\left[2\sigma(2a) - 1\right], \Rightarrow \frac{w_j}{u_j} = \frac{2\sigma(2a) - 1}{\sigma(2a)} = 2 - \frac{1}{\sigma(2a)}$$

$$= \boxed{2 - \frac{1}{\sigma\left(\frac{x-\mu_j}{s}\right)}}$$

※

Next Page for Q3.2

Q3.2

Given
$$\begin{cases} y(x, \vec{w}) = w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x-\mu_j}{s}\right) \\ y(x, \vec{u}) = u_0 + \sum_{j=1}^{M} u_j \tanh\left(\frac{x-\mu_j}{2s}\right) \end{cases}$$

Show these 2 expression are the same.

$$w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x-\mu_j}{s}\right) = u_0 + \sum_{j=1}^{M} u_j \tanh\left(\frac{x-\mu_j}{2s}\right)$$ Assume $a = \frac{x-\mu_j}{2s}$

$$\Rightarrow w_0 + \sum_{j=1}^{M} w_j \sigma(2a) = u_0 + \sum_{j=1}^{M} u_j \tanh(a) = u_0 + \sum_{j=1}^{M} u_j (2\sigma(2a) - 1)$$

Refer to 黃浩然

$$\Rightarrow w_0 + \frac{w_1}{1+\exp\left(\frac{-x+\mu_1}{s}\right)} + \frac{w_2}{1+\exp\left(\frac{-x+\mu_2}{s}\right)} + \cdots + \frac{w_M}{1+\exp\left(\frac{-x+\mu_M}{s}\right)}$$

$$= u_0 + u_1 \frac{2-\left(1+\exp\left(\frac{-x+\mu_1}{s}\right)\right)}{1+\exp\left(\frac{-x+\mu_1}{s}\right)} + \frac{u_2\left(2-\left(1+\exp\left(\frac{-x+\mu_2}{s}\right)\right)\right)}{1+\exp\left(\frac{-x+\mu_2}{s}\right)} + \cdots + \frac{w_M}{1+\exp\left(\frac{-x+\mu_M}{s}\right)}$$

$$\frac{[w_1, w_2, w_3 \ldots w_M]}{[u_1, u_2, u_3 \ldots u_M]} = 1 - \exp\left(\frac{-x+\mu_j}{s}\right), \quad j = \{1, 2, \ldots M\}$$

※

$$w_0 + \sum_{j=1}^{M} w_j \sigma(2a) = u_0 + \sum_{j=1}^{M} u_j (2\sigma(2a) - 1) = u_0 + \sum_{j=1}^{M} 2u_j \sigma(2a) - \sum_{j=1}^{M} u_j$$

$$= u_0 - \sum_{j=1}^{M} u_j + \sum_{j=1}^{M} 2u_j \sigma(2a)$$

Because $w_0 \neq u_0$, cannot use the method above (red cross area)

$$w_0 = u_0 - \sum_{j=1}^{M} u_j$$

$$\sum_{j=1}^{M} w_j \sigma(2a) = \sum_{j=1}^{M} 2u_j \sigma(2a) \Rightarrow \boxed{\frac{u_j}{w_j} = \frac{1}{2}} \text{ for } j \in \{1, 2 \ldots M\}$$

MLHW

Q4. Given $f_{w,b}(x) = w^T x + b$ where $w \in \mathbb{R}^k, b \in \mathbb{R}, x_i \in \mathbb{R}^k, \eta_i \in \mathbb{R}^k$

$$\tilde{L}_{ss}(w,b) = \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{w,b}(x_i+\eta_i) - y_i\right)^2\right]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left[\mathbb{E}\left[\left(w^T(x_i+\eta_i)+b-y_i\right)^2\right]\right]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left[\mathbb{E}\left[\left(\underline{w^T x_i} + w^T\eta_i \underline{+ b - y_i}\right)^2\right]\right]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left[\mathbb{E}\left[\left(f_{w,b}(x_i) - y_i + w^T\eta_i\right)^2\right]\right]$$

Independent

$$= \frac{1}{2N}\sum_{i=1}^{N}\left[\mathbb{E}\left[\underbrace{(f_{w,b}(x_i)-y_i)^2}_{a^2} + \underbrace{2}_{2}\underbrace{(f_{w,b}(x_i)-y_i)}_{a}\underbrace{(w^T\eta_i)}_{b} + \underbrace{(w^T\eta_i)^2}_{b^2}\right]\right]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left[\mathbb{E}\left[\underbrace{(f_{w,b}(x_i)-y_i)^2}_{\hookrightarrow \text{ constant to } \mathbb{E}}\right] + 2\mathbb{E}\left[f_{w,b}(x_i)-y_i\right]\mathbb{E}\left[w^T\eta_i\right] + \mathbb{E}\left[(w^T\eta_i)^2\right]\right]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\left[(f_{w,b}(x_i)-y_i)^2 + 2(f_{w,b}(x_i)-y_i)\,w^T\underbrace{\mathbb{E}[\eta_i]}_{0} + (w^T)^2\,\mathbb{E}[\eta_i^2]\right]$$

where $\mathbb{E}[\eta_{i,j}] = 0$

$$= \frac{1}{2N}\sum_{i=1}^{N}(f_{w,b}(x_i)-y_i)^2 + \frac{(w^T)^2}{2N}\sum_{i=1}^{N}\mathbb{E}[(\eta_i)^2]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}(f_{w,b}(x_i)-y_i)^2 + \frac{\|w\|^2}{2N}\sum_{i=1}^{N}\delta_{ii}\delta_{ii'}\sigma^2 \quad \text{when } \delta_{i,i'} = \begin{cases} 1, & \text{if } i = i' \\ 0, & \text{otherwise} \end{cases}$$

$$= \frac{1}{2N}\sum_{i=1}^{N}(f_{w,b}(x_i)-y_i)^2 + \frac{\|w\|^2 \cdot N\sigma^2}{2N} =$$

$$= \boxed{\frac{1}{2N}\sum_{i=1}^{N}(f_{w,b}(x_i)-y_i)^2 + \frac{\|w\|^2\sigma^2}{2}}$$

※

Q5: Given linear function of the feature vector $\vec{x} = [x_1, x_2, x_3 \dots x_n]^T \in \mathbb{R}^n$ and $\vec{w} = [w_1, w_2, w_3 \dots w_n]^T \in \mathbb{R}^n$, so that

$$f_{\vec{w},b}(\vec{x}) = p(C_1 | \vec{x}) = \sigma\left(\sum_i w_i x_i + b\right) = \sigma(\vec{w}^T \vec{x} + b)$$

with $p(C_2 | \vec{x}) = 1 - p(C_1 | \vec{x}) = 1 - f_{\vec{w},b}(\vec{x})$

(a) Suppose $\vec{w} = [-1, 2, -1, 5]^T$, $\vec{x} = [7, 0, 3, 10]^T$ and $b = 3$

$$\Rightarrow f_{\vec{w},b}(\vec{x}) = \sigma(\vec{w}^T \vec{x} + b) = \sigma([-1, 2, -1, 5][7, 0, 3, 10]^T + 3) = \sigma(-7 + 0 - 3 + 50 + 3)$$

$$= \sigma(43) = \frac{1}{1 + e^{-43}} = \boxed{1} \; *$$

(b) Given training dataset $\{x_i, y_i\}_{i=1}^{N}$ ($N \to$ quantity of dataset), where $y_i \in \{0, 1\}$

$\Rightarrow$ likelihood function: $L(\theta) = f(x_1; \theta) f(x_2; \theta) f(x_3; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$

$\Rightarrow L(w,b) = f_{w,b}(x^1) f_{w,b}(x^2) \dots f_{w,b}(x^N)$

$\Rightarrow$ Let $L(w,b)$ likelihood function maximize means minimize $-\ln L(w,b)$

$\Rightarrow -\ln L(w,b) = -\ln f_{w,b}(x^1) \cdot -\ln f_{w,b}(x^2) \cdot -\ln f_{w,b}(x^3) \cdots -\ln f_{w,b}(x^N)$

$= -[y_1 \ln f_{w,b}(x_1) + (1-y_1)\ln(1 - f_{w,b}(x_1))] \cdot -[y_2 \ln f_{w,b}(x_2) + (1-y_2)\ln(1 - f_{w,b}(x_2))] \cdots$

$= \prod_{i=1}^{N} -[y_i \ln f_{w,b}(x_i) + (1-y_i)\ln(1 - f_{w,b}(x_i))] \; *$

(c) Observe $(x^i, \hat{y}^i)_{i=1}^{N}$ where $\hat{y}^i \in \{0, 1\}$, $x^i \in \mathbb{R}^d$

Given: $P_\theta(C_1 | x) = \sigma(w^T x + b)$ is equivalent to minimize $\sum_{i=1}^{N} -[y_i \ln(\sigma(w^T x + b)) + (1-y_i)\ln(1 - \sigma(w^T x + b))]$

$\Rightarrow$ Apply Gradient Descent $\frac{\partial L}{\partial w_i}(w,b)$ 、 $\frac{\partial L}{\partial b_i}(w,b)$

$\Rightarrow \frac{\partial L}{\partial w_i}(w,b) = -\sum_{i=1}^{N}\left[y_i \frac{\partial}{\partial w_i} \ln(\sigma(w^T x + b)) + (1-y_i)\frac{\partial}{\partial w_i}\ln(1 - \sigma(w^T x + b))\right]$

$\frac{\partial}{\partial z}\sigma(z) = \sigma(z)\sigma(-z) = \sigma(z)(1 - \sigma(z))$

$w_1 x_1 + w_2 x_2 \cdots \underline{\underline{w_i x_i}} + \cdots w_N x_N$

$$= -\sum_{n=1}^{N}\left[y_i \cdot \frac{\sigma(w^T x^n + b)(1 - \sigma(w^T x^n + b))}{\sigma(w^T x^n + b)} \cdot x_i^n + (1-y_i)\frac{(-1)\sigma(w^T x^n + b)(1 - \sigma(w^T x^n + b))}{1 - \sigma(w^T x^n + b)} \cdot x_i^n\right]$$

$$= -\sum_{n=1}^{N} x_i^n \left[y_i(1 - \sigma(w^T x^n + b)) + (-1 + y_i)\sigma(w^T x^n + b)\right]$$

$$= -\sum_{n=1}^{N} x_i^n \left[y_i - y_i\sigma(w^T x^n + b) + -\sigma(w^T x + b) + y_i\sigma(w^T x^n + b)\right]$$

$$= -\sum_{n=1}^{N} x_i^n \left[y_i - \sigma(w^T x^n + b)\right] = -\sum_{n=1}^{N} x_i^n \left[y_i - f_{w,b}(x_i^n)\right]$$

$$\Rightarrow w_i^{(t+1)} = w_i^{(t)} - \eta \sum_n -(y^i - f_{w,b}(x_i^n))\, x_i^n \; *$$