Xin Qian xing HW1 - 11741

1. (a) Lemma: w is orthogonal to the line $w^T x + b = 0$

Proof: Consider x_1, x_2 on the line

$$\begin{cases} w^T x_1 - b = 0 & ① \\ w^T x_2 - b = 0 & ② \end{cases} \quad ① - ② \quad w^T(x_1 - x_2) = 0$$

$\therefore$ perpendicular

Given w orthogonal to the line, suppose there's point a^* on the line who is closest to the origin

$a^* \perp$ hyperplane h

$a^* = \alpha w$ α is some scalar

$$f(a^*) = w^T \alpha w - b = \alpha w^T w - b = 0$$

$$\alpha = \frac{b}{w^T w}$$

$$a^* = \frac{bw}{w^T w}$$

$$\text{dist}(a^* \text{ to origin}) = \sqrt{a^{*T} a} = \sqrt{\frac{b^T b (w^T w)}{(w^T w)^2}} = \frac{b}{\|w\|}$$

(b) $y \in \{0, 1\}$, $x = z + \frac{yw}{\|w\|} \cdot d$ (d is the perpendicular distance)

$$f(x) = w^T\left(z + \frac{yw}{\|w\|} \cdot d\right) - b$$

$$= (w^T z - b) + \frac{y w^T w}{\|w\|} \cdot d \quad (z \text{ is on the hyperplane})$$

$$= 0 + \|w\| d y$$

$$d = \frac{f(x)}{y\|w\|} = \frac{yf(x)}{\|w\|} \quad (y \text{ is a sign indicator})$$

2. (a) $Ax = \lambda x$ $(A - \lambda I)x = 0$

$$|A - \lambda I| = 0 \quad \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = 0 \quad (3-\lambda)^2 - 1 = 0$$

$\lambda_1 = 2$ $\lambda_2 = 4$

$$\lambda_1 = 2 \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \quad x_1 + x_2 = 0 \quad \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\lambda_2 = 4 \quad \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \quad x_1 = x_2 \quad \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

(b) Given $Ax = \lambda_i x$ for $i = 1 \cdots n$

We want to find λ that satisfies $A^k x = \lambda x$

By associativity, $A^k x = A^{k-1}(Ax) = A^{k-1}(\lambda x) = \lambda A^{k-1} x \cdots = \lambda^k x$

where $\lambda = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$

Given A^k and eigenvector λ_i^k and the eigenvector $Ax = \lambda_i x$

$$A^k x = A^{k-1}(Ax) = A^{k-1}(\lambda_i x) = \lambda_i (A^{k-1} x) = \cdots$$
$$= \lambda_i^k x$$

therefore each eigenvector of A is still an eigenvector of A^k

3. (a) $\hat{p} = \arg\max_p P(k \mid n, p)$

$$= \arg\max_p C_n^k p^k (1-p)^{n-k}$$

$$= \arg\max_p \ln C_n^k + k \ln p + (n-k)\ln(1-p)$$

$$= \arg\max_p \underbrace{(k \ln p + (n-k)\ln(1-p))}_{f(p)}$$

$f'(p) = \frac{k}{p} - \frac{n-k}{1-p}$ when $f'(p) = 0$ $f(p)$ max

$$p = \frac{k}{n}$$

(b) We want to maximize $P(n_j, n)$ subject to $\sum_{i=1}^{m} p_i = 1$

Introduce the λ lagrange Multiplier

$$P(n_j, n) = \frac{n!}{\prod n_i} \prod p_i^{n_i}$$

$$\log P(n_j, n) = \log n! - \sum \log n_i + \sum n_i \log p_i$$

$$L(n_j, n) = \log P(n_j, n) - \lambda(\sum p_i - 1)$$

take partial derivative on every p_i, when $\hat{p}_i = \frac{n_i}{\lambda}$ max

then $\lambda = n_1 + \cdots + n_m = n$

$$\hat{p}_i = \frac{n_i}{n}$$

4. (a) $$\frac{du}{dx} = \frac{d(1+e^{-x})^{-1}}{dx} = -(1+e^{-x})^{-2}\cdot e^{-x}\cdot(-1) = e^{-x}(1+e^{-x})^{-2}$$

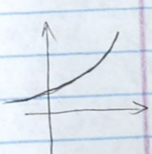
$$u(1-u) = \frac{1}{1+e^{-x}}\cdot\frac{e^{-x}}{1+e^{-x}} = e^{-x}(1+e^{-x})^{-2}$$

$$\therefore \frac{du}{dx} = u(1-u)$$

(b) $$\nabla l = \left(\frac{\partial l}{\partial w_0}, \frac{\partial l}{\partial w_1} \cdots \frac{\partial l}{\partial w_m}\right)^T$$

Take the ith element for example

$$\frac{\partial l}{\partial w_i} = \frac{\partial l}{\partial u}\cdot\frac{\partial u}{\partial z}\cdot\frac{\partial z}{\partial w_i}$$



$$= \left(\frac{y}{u} - \frac{(1-y)}{1-u}\right)(1+e^{-z})^{-2}e^{-z}\cdot x_i$$

$$= (y-u)x_i$$ every other element is just the same

(c) $$H_{ij} = \frac{\partial^2 l}{\partial w_i \partial w_j} = \frac{\partial}{\partial w_j}(y-u)x_i = -x_i\frac{\partial u}{\partial w_j} = -x_i\frac{\partial u}{\partial z}\frac{\partial z}{\partial w_j}$$

$$= -x_i x_j \frac{e^{-z}}{(1+e^{-z})^2}$$ (symmetrix)

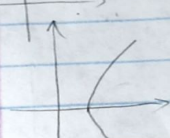
First proof:

(d) Logistic Regression $t = \beta_0 + \beta x$

$$L(\beta_0, \beta) = \prod_{i=1}^{n} \sigma(t_i)^{y_i}(1-\sigma(t_i))^{1-y_i}$$

$$l(\beta_0,\beta) = \sum y_i \log \sigma(t_i) + (1-y_i)\log(1-\sigma(t_i))$$

$$= \sum y_i \log\frac{\sigma(t_i)}{1-\sigma(t_i)} + \sum \log(1-\sigma(t_i))$$



$$= \sum y_i t - \sum \log(1+e^t)$$

$$= \underbrace{\sum y_i(\beta_0+\beta x)}_{\text{affine function}} - \sum \underbrace{\log}_{\text{concave}}(1+\underbrace{e^{\overbrace{t}^{\text{affine function (concave)}}}}_{\text{exponential (convex)}})$$

every component is concave