

Date ___/___/___

AI534 Written Assignment 1

1a. $\epsilon_i \sim N(0, \sigma_i^2) \Rightarrow y_i = W^T x_i + \epsilon_i$

$$P(Y|X;W) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i - W^T x_i)^2}{2\sigma_i^2}}$$

* $\mathcal{H} = \mathcal{L}(W)$

$$\ell(W) = \log(\mathcal{L}(W)) = \log \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i - W^T x_i)^2}{2\sigma_i^2}} \right]$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^N \log\left(\frac{1}{\sigma_i}\right) - \frac{(y_i - W^T x_i)^2}{2\sigma_i^2}$$

b. $\underset{W}{\operatorname{argmax}} \ell(W) = \underbrace{\frac{1}{\sqrt{2\pi}} \sum_{i=1}^N \log\left(\frac{1}{\sigma_i}\right)}_{\text{constant}} - \underbrace{\frac{(y_i - W^T x_i)^2}{2\sigma_i^2}}_{\text{constant}}$

$$= \underset{W}{\operatorname{argmax}} \sum_{i=1}^N -\frac{(y_i - W^T x_i)^2}{2\sigma_i^2}$$

$$= \underset{W}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i} (W^T x_i - y_i)^2$$

$$= \frac{1}{2} \sum_{i=1}^N a_i (W^T x_i - y_i)^2$$

where $a_i = \frac{1}{\sigma_i}$

$$c. \operatorname{argmax}_N (l(w)) = \operatorname{argmin}_W \frac{1}{2} \sum_{i=1}^N a_i (y_i - w^T x_i)^2$$

$$\nabla_W L = \sum_{i=1}^N a_i (y_i - w^T x_i) \cdot (-x_i)$$

$$= \sum_{i=1}^N a_i (w^T x_i - y_i) x_i = \sum_{i=1}^N a_i (\hat{y}_i - y_i) x_i$$

d. ~~matrix~~ Let $\begin{bmatrix} x_{00} & x_{01} & \dots & x_{0,d-1} \\ x_{10} & x_{11} & \dots & x_{1,d-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1,0} & \dots & \dots & x_{n-1,d-1} \end{bmatrix} = X$

$$\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{d-1} \end{bmatrix} = W \quad , \quad \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} = Y$$

~~Let~~ Let $[a_0, a_1, \dots, a_{n-1}] = A$

$$\Rightarrow \operatorname{argmin}_W A \|XW - Y\|^2$$

$$= \operatorname{argmin}_W A \odot (XW - Y)^2 \quad \odot = \text{dot operation}$$

Since A is a constant it will not affect the minimum

$$\Rightarrow \operatorname{argmin} \|XW - Y\|^2$$

$$= (W^T X^T X W - 2W^T X^T Y + Y^T Y)$$

$$\Rightarrow W = (X^T X)^{-1} X^T Y.$$

Date / /

2. a. $L(W) = \prod_{i=1}^N \prod_{k=1}^K p(y_i = k | x_i)^{I(y_i = k)}$

$$\log(L(W)) = \sum_{i=1}^N \left(\sum_{k=1}^K \log P(y_i = k | x_i)^{I(y_i = k)} \right)$$

$$= \sum_{i=1}^N \left(\sum_{k=1}^K I(y_i = k) \log P(y_i = k | x_i) \right)$$

$$= \sum_{i=1}^N \left(\sum_{k=1}^K I(y_i = k) \cdot \log \left(\frac{e^{W_k^T x}}{\sum_{k=1}^K e^{W_k^T x}} \right) \right)$$

$$= \sum_{i=1}^N \sum_{k=1}^K I(y_i = k) \left(\log(e^{W_k^T x}) - \log \left(\sum_{k=1}^K e^{W_k^T x} \right) \right)$$

Since $I(y_i = k)$ is 1 only for a particular value of $k = k_0$, we can say

~~$$\sum_{k=1}^K I(y_i = k) W_k^T x =$$~~

$$\log(L(W)) = \sum_{i=1}^N \log e^{W_{k_0}^T x} - \log \left(\sum_{k=1}^K e^{W_k^T x} \right)$$

$$= \sum_{i=1}^N \left(W_{k_0}^T x - \log \left(\sum_{k=1}^K e^{W_k^T x} \right) \right)$$

b. For W_{k_0} (i.e.) $I(y_i = k_0) = 1$.

$$\nabla l(W) = \sum_{k=1}^K W_k X - \frac{e^{W_{k_0}^T x}}{\sum_{k=1}^K e^{W_k^T x}} \cdot X$$

$$= X \left[\frac{\sum_{k=1}^K e^{W_k^T x} - e^{W_{k_0}^T x}}{\sum_{k=1}^K e^{W_k^T x}} \right]$$

Date ____ / ____ / ____

If $I(y_i = k_c) > 0$ (ie) $w_i \neq w_c$

$$\nabla l(w) = \frac{-e^{w_{k_c}^T x}}{\sum_{k=1}^K e^{w_k^T x}} \cdot x$$

$$\Rightarrow \nabla l(w) = I(y_i = k) \cdot x \left(1 - \frac{e^{w_{k_c}^T x}}{\sum_{k=1}^K e^{w_k^T x}} \right)$$

$$- (1 - I(y_i = k_c)) \cdot x \cdot \frac{e^{w_{k_c}^T x}}{\sum_{k=1}^K e^{w_k^T x}}$$

3. $p(x_i | \theta) = \theta^{x_i} (1-\theta)^{1-x_i}$ $k_i = 1$ if $x_i = 1$
 $= \theta^{x_i} (1-\theta)^{(1-x_i)}$

$$P(\theta) = \frac{\theta^{(\alpha-1)} \cdot (1-\theta)^{(\beta-1)}}{B(\alpha, \beta)}$$

$$P(\theta | x_i) = \frac{P(\theta) \cdot P(x_i | \theta)}{P(x_i)}$$

$$\Rightarrow P(\theta | x_1, \dots, x_n) = \frac{P(\theta) \cdot P(x_1 | \theta) \cdot P(x_2 | \theta) \cdot \dots \cdot P(x_n | \theta)}{P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_n)}$$

$$= \frac{\theta^{(\alpha-1)} \cdot (1-\theta)^{(\beta-1)}}{B(\alpha, \beta)} \cdot \frac{\prod_{i=1}^n \theta^{x_i} \cdot (1-\theta)^{(1-x_i)}}{\prod_{i=1}^n P(x_i)}$$

Now since $P(\theta | x_1, \dots, x_n)$ is a PDF

$$\int P(\theta | x_1, \dots, x_n) = 1$$

$$\Rightarrow \frac{\theta^{\alpha-1+\sum_{i=1}^N x_i} (1-\theta)^{\beta-1+\sum_{i=1}^N (1-x_i)}}{B(\alpha, \beta) \cdot \prod_{i=1}^N P(x_i)}$$

Comparing this to the Beta Function for $P(\theta)$ we see

$$P(\theta | x) = \text{Beta}\left(\alpha + \sum_{i=1}^N x_i, \beta + \sum_{i=1}^N (1-x_i)\right)$$

and $\prod_{i=1}^N P(x_i) = \frac{B(\alpha + \sum_{i=1}^N x_i, \beta + \sum_{i=1}^N (1-x_i))}{B(\alpha, \beta)}$

b- $P(\theta) = \text{Beta}(2, 2)$

If we observe ~~5~~ 2 Heads out of 5 tosses:

$$P(\theta | x) = \text{Beta}(2+2, 2+3) \\ = \text{Beta}(4, 5)$$

If we observe ~~20~~ 20 Heads out of 50 tosses

$$P(\theta | x) = \text{Beta}(22, 32)$$