Date ___/___/___

Saathi

AI 534 Written Assignment 1

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

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

$$\prod = \mathcal{L}(W)$$

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

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

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

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

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

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

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

Page No.

Date __/__/__

c. $\underset{W}{\arg\max}\,(l(w)) = \underset{W}{\arg\min}\ \frac{1}{2}\sum_{i=1}^{N} a_i (y_i - W^T x_i)^2$

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

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

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

$\begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{d-1} \end{bmatrix} = W$, $\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} = Y$

~~$L(W) = \|XW - Y\|^2$~~ Let $[a_0, a_1 \cdots a_{n-1}] = A$

$\Rightarrow \underset{W}{\arg\min}\ A\|XW - Y\|^2$

$= \underset{W}{\arg\min}\ A \odot (XW - Y)^2$ $\odot$ = dot operation

Since A is a constant it will not affect the minimum

$\Rightarrow \arg\min\ \|XW - Y\|^2$

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

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

Date __/__/__

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

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

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

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

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

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

$$\cancel{\sum_{k=1}^{K} e^{W_k^T x}} =$$

$$\log(L(W)) = \sum_{i=1}^{N} \log e^{W_{k0}^T x} - \log\left(\sum_{k=1}^{K} e^{W_k^T x} \right)$$

$$= \sum_{i=1}^{N} \left(W_{k0}^T x - \log\left(\sum_{k=1}^{K} e^{W_k^T x} \right) \right)$$

b. For W_{kc} (ie) $I(y_i = k_c) = 1$.

$$\nabla l(W) = \cancel{\sum} \cancel{W_{kc}} \; x - \frac{e^{W_{kc}^T x}}{\sum_{k=1}^{K} e^{W_k^T x}} \cdot x$$

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

Date __/__/__

If $I(y_i = k_c) = 0$ (ie) $W_i \neq W_c$

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

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

$$- \left(1 - I(y_i = k_c)\right) \cdot x \cdot \frac{e^{W_{k_c}^T x}}{\sum_{k=1}^{K} e^{W_k^T x}}$$

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

$$= \theta^{x_i} (1-\theta)^{(1-x_i)}$$

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

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

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

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

Date ___/___/_____

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

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

$\Rightarrow$

$$P(\theta|X) = \frac{\theta^{\alpha - 1 + \sum_{i=1}^{N} x_i} \cdot (1-\theta)^{\beta - 1 + \sum_{i=1}^{N}(1-x_i)}}{B(\alpha, \beta) \cdot \prod P(X_i)}$$

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

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

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

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

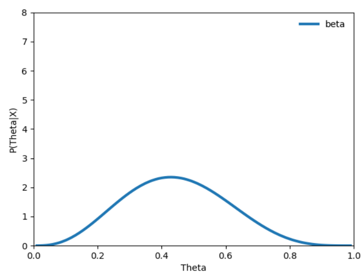
If we observe ~~5H~~ 2 Heads out of 5 tosses:

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

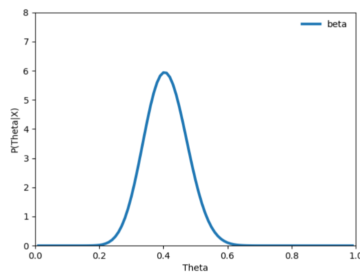
If we observe ~~20~~ 20 Heads out of 50 tosses

$$P(\theta|X) = Beta(22, 32)$$

We see that as the number of observations increases, the value of $P(\theta|X)$ converges to the correct value of 0.4. When the number of observations become infinity, the graph would look like the delta function at $\theta = 0.4$



(a) 2 heads out of 5



(b) 20 heads out of 50

Figure 1: PDF of $P(\theta|X)$