3. (c) $\hat{p}(x) \triangleq \frac{1}{N}\sum_{i=1}^{N} I[x=x_i] = \frac{1}{N}\sum_{i=1}^{N}\delta(x-x_i)$

$D_{KL}(\hat{p}\|q) = -\int \hat{p}(x)\ln\frac{q(x|\theta)}{\hat{p}(x)}dx = -\int\hat{p}(x)\ln q(x|\theta)dx + \int\hat{p}(x)\ln\hat{p}(x)dx.$

$\underset{\theta}{\arg\min}\, D_{KL}(\hat{p}\|q) = \underset{\theta}{\arg\max}\int\hat{p}(x)\ln q(x|\theta)dx$

$\int\hat{p}(x)\ln q(x|\theta)dx = \frac{1}{N}\int\sum_{i=1}^{N} N\delta(x-x_i)\ln q(x|\theta)dx = \underline{\frac{1}{N}\sum_{i=1}^{N}\ln p(x_i|\theta)}.$

$\Rightarrow$ The minimum of $D_{KL}(\hat{p}\|q)$ is obtained by the maximum likelihood estimation #

(d) To maxmize $-\int p(x)\ln p(x)dx$. we consider $F(p,\lambda_1,\lambda_2,\lambda_3)$

$= -\int p(x)\ln|x)dx + \lambda_1(\int p(x)dx - 1) + \lambda_2(\int xp(x)dx - \mu) + \lambda_3(\int(x-\mu)^2p(x)dx - \sigma^2)$

$\frac{\partial F}{\partial p} = -1 - \ln p + \lambda_1 + \lambda_2 x + \lambda_3(x-\mu)^2 = 0 \Rightarrow p = \exp(\lambda_1 - 1 + \lambda_2 x + \lambda_3(x-\mu)^2)$

For $\int p(x)dx$ to be finite requires $\lambda_2 = 0$ and $\lambda_3 < 0$.

$\Rightarrow P(x) = e^a e^{-b(x-\mu)^2}$. where $a = \lambda_1 - 1$. $b = -\lambda_3 > 0$.

$\int p(x)dx = e^a\sqrt{\frac{\pi}{b}} = 1 \Rightarrow P(x) = \sqrt{\frac{b}{\pi}}e^{-b(x-\mu)^2}$. $\int(x-\mu)^2p(x)dx = \frac{1}{2b} = \sigma^2$

$\Rightarrow b = \frac{1}{2\sigma^2} \Rightarrow \underline{p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(\frac{-1}{2\sigma^2}(x-\mu)^2)}$. #

4. (a) $L(\beta, b) = L(w) = \sum_{i=1}^{n} c_i(y_i - \beta^T x_i - b)^2 = (Y - Xw)^T C(Y - Xw)$

where $Y \in \mathbb{R}^n$. $X \in \mathbb{R}^{n\times m}$. $w \in \mathbb{R}^m$. $C \in \mathbb{R}^{n\times n}$. $C_{ij} = c_i$, $i=j$; $C_{ij} = 0$, $i \neq j$.

$\frac{\partial L}{\partial w} = 2(X^TCXw - X^TCY) = 0 \Rightarrow \underline{w = (X^TCX)^{-1}X^TCY}$ #

$\epsilon_i = y_i - x_i w$. $\Rightarrow l(w) = \frac{-n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i w)^2$

$\Rightarrow \underset{w}{\arg\min}\sum_{i=1}^{n}(y_i - x_i w) = \underset{w}{\arg\max}\, l(w)$.

(b) $l(w) = \sum_{i=1}^{n}\frac{-1}{2}\ln(\sigma_i^2) - \frac{1}{2\sigma_i^2}(y_i - x_i w)^2$

$\Rightarrow \underset{w}{\arg\min}\sum_{i=1}^{n} c_i(y_i - x_i w) = \underset{w}{\arg\max}\, l(w)$. where $c_i = \frac{1}{\sigma_i^2}$

5. (a) $t^{(i)}(w^Tx^{(i)} + b) \geq 1 - \xi_i \Rightarrow \xi_i \geq 1 - t^{(i)}(w^Tx^{(i)} + b)$ and $\xi_i \geq 0$.

$\therefore \min_{w,b}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \iff \min_{w,b}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\max(0, 1 - t^{(i)}(w^Tx^{(i)} + b))$