1. Introduction

1.3

$p(a) = \sum\limits_{c \in (r,b,g)} p(a|c)p(c)$
$= p(a|r)p(a) + p(a|b)p(b) + p(a|g)p(g)$
$= \frac{3}{10} \cdot 0.2 + \frac{1}{2} \cdot 0.2 + \frac{3}{10} \cdot 0.6$
$= 0.06 + 0.1 + 0.18 = 0.34$

$p(g|o) = \frac{p(o|g)p(g)}{\sum\limits_{c \in (r,b,g)} p(o|c)p(c)}$
$= \frac{\frac{3}{10} \cdot 0.6}{\frac{4}{10} \cdot 0.2 + \frac{1}{2} \cdot 0.2 + \frac{3}{10} \cdot 0.6}$
$= \frac{0.18}{0.08+0.1+0.18} = \frac{0.18}{0.36} = \frac{1}{2}$

1.12

(1.49) $E[x] = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)xdx = \mu$
(1.50) $E[x^2] = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)x^2dx = \mu^2 + \sigma^2$

$n \neq m$이면, iid이므로
$E[x_n x_m] = E[x_n]E[x_m] = \mu\mu = \mu^2 \leftarrow (1.49)$
$n = m$이면
$E[x_n x_m] = E[x_n^2] = \mu^2 + \sigma^2 \leftarrow (1.50)$

따라서
$E[x_n x_m] = \mu^2 + I_{nm}\sigma^2$

x_n은 iid이므로
$E[\mu_{ML}] = E[\frac{1}{N} \sum\limits_{n=1}^{N} x_n] = \frac{1}{N} \sum\limits_{n=1}^{N} E[x_n] = \frac{1}{N} N\mu = \mu$
$E[\sigma_{ML}^2] = E[\frac{1}{N} \sum\limits_{n=1}^{N} (x_n - \mu_{ML})^2] = \frac{1}{N} E[\sum\limits_{n=1}^{N} (x_n - \frac{1}{N} \sum\limits_{k=1}^{N} x_k)^2]$
$= \frac{1}{N} \sum\limits_{n=1}^{N} E[x_n^2 - \frac{2}{N} x_n \sum\limits_{k=1}^{N} x_k + \frac{1}{N^2} (\sum\limits_{k=1}^{N} x_k)^2]$
$= \frac{1}{N} N[\mu^2 + \sigma^2 - \frac{2}{N}(\mu^2 + \sigma^2 + (N-1)\mu^2) + \frac{1}{N^2}(N(\mu^2 + \sigma^2) + N(N-1)\mu^2)]$
$= \mu^2 + \sigma^2 - 2\mu^2 - \frac{2}{N}\sigma^2 + \frac{1}{N}(\mu^2 + \sigma^2 + (N-1)\mu^2)$
$= \mu^2 + \sigma^2 - 2\mu^2 - \frac{2}{N}\sigma^2 + \mu^2 + \frac{1}{N}\sigma^2$

$= \frac{N-1}{N}\sigma^2$

1.14

$w_{ij}^S = \frac{1}{2}(w_{ij} + w_{ji})$
$w_{ij}^A = \frac{1}{2}(w_{ij} - w_{ji}) = -\frac{1}{2}(w_{ji} - w_{ij}) = -w_{ji}^A$
$w_{ij} = w_{ij}^S + w_{ij}^A = \frac{1}{2}(w_{ij} + w_{ji} + w_{ij} - w_{ji}) = w_{ij}$

$\sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}x_i x_j = \sum_{i=1}^{D}\sum_{j=1}^{D}(w_{ij}^S + w_{ij}^A)x_i x_j$
$= \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^S x_i x_j + \frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}(w_{ij} - w_{ji})x_i x_j$
$= \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^S x_i x_j + \frac{1}{2}(\sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}x_i x_j - \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ji}x_i x_j)$
$= \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^S x_i x_j$

Upper triangular matrix의 값이 정해지면 Lower triangular matrix의 값이 정해지므로 Independent parameter 개수는 Lower triangular matrix parameter 개수가 된다.

따라서
$\sum_{n=1}^{D} n = \frac{(D+1)D}{2}$

1.20

$\epsilon \ll 1$이므로 Shall에 대해서 적분하면
$\int_{shall} p(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{(2\pi\sigma^2)^{D/2}}exp(-\frac{r^2}{2\sigma^2})S_D r^{D-1}\epsilon$
$= \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{D/2}}exp(-\frac{r^2}{2\sigma^2})\epsilon = p(r)\epsilon$

$p(r)$을 r에 대해 미분하고 0이라고 정리한 뒤 Large D 조건을 활용하면
$(D-1)\frac{S_D r^{D-2}}{(2\pi\sigma^2)^{D/2}}exp(-\frac{r^2}{2\sigma^2}) + (-\frac{r}{\sigma^2})\frac{S_D r^{D-1}}{(2\pi\sigma^2)^{D/2}}exp(-\frac{r^2}{2\sigma^2}) = 0$
$(D-1)r^{D-2} + (-\frac{r}{\sigma^2})r^{D-1} = 0$
$\hat{r} \simeq \sqrt{D}\sigma$

$\epsilon \ll \hat{r}$이라 하면
$p(\hat{r}+\epsilon) = \frac{S_D(\hat{r}+\epsilon)^{D-1}}{(2\pi\sigma^2)^{D/2}}exp(-\frac{(\hat{r}+\epsilon)^2}{2\sigma^2})$
$= \frac{S_D}{(2\pi\sigma^2)^{D/2}}exp(-\frac{(\hat{r}+\epsilon)^2}{2\sigma^2} + \ln(\hat{r}+\epsilon)^{D-1})$
$= \frac{S_D}{(2\pi\sigma^2)^{D/2}}exp(-\frac{\hat{r}^2+2\hat{r}\epsilon+\epsilon^2}{2\sigma^2} + (D-1)\ln(\hat{r}(1+\frac{\epsilon}{\hat{r}})))$

$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} exp(-\frac{\hat{r}^2}{2\sigma^2} - \frac{2\hat{r}\epsilon+\epsilon^2}{2\sigma^2} + (D-1)\ln\hat{r} + (D-1)\ln(1+\frac{\epsilon}{\hat{r}}))$
$\simeq \frac{S_D \hat{r}^(D-1)}{(2\pi\sigma^2)^{D/2}} exp(-\frac{\hat{r}^2}{2\sigma^2}) exp(-\frac{2\hat{r}\epsilon+\epsilon^2}{2\sigma^2} + D(\frac{\epsilon}{\hat{r}} - \frac{\epsilon^2}{2\hat{r}^2})) \because \ln(1+x) = x - \frac{x^2}{2} + O(x^3)$
$= \frac{S_D \hat{r}^(D-1)}{(2\pi\sigma^2)^{D/2}} exp(-\frac{\hat{r}^2}{2\sigma^2}) exp(-\frac{2\hat{r}\epsilon+\epsilon^2}{2\sigma^2} + \frac{\hat{r}^2}{\sigma^2}(\frac{2\hat{r}\epsilon-\epsilon^2}{2\hat{r}^2}))$
$= p(\hat{r}) exp(-\frac{\epsilon^2}{\sigma^2})$

마지막으로 비율을 비교하면
$p(0)/p(\hat{r}) = exp(-\frac{0}{2\sigma^2})/exp(-\frac{\hat{r}^2}{2\sigma^2}) = exp(\frac{\hat{r}^2}{2\sigma^2})$
$= exp(\frac{D\sigma^2}{2\sigma^2}) = exp(\frac{D}{2})$

1.21

$a, b > 0$ 이므로
$a \le b \Rightarrow a^2 \le ab \Rightarrow a \le (ab)^{1/2}$

오분류 확률을 계산하기 위한 R_n을 C_n인 구역이라고 하면
$p(\text{오분류}) = \int_{R_1} p(C_2|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} + \int_{R_2} p(C_1|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$

여기서 오분류를 하였으므로 R_1에서 $p(C_2|\boldsymbol{x}) \le p(C_1|\boldsymbol{x})$이고
R_2에서 $p(C_1|\boldsymbol{x}) \le p(C_2|\boldsymbol{x})$ 이며 확률은 음수가 아니므로
$p(\text{오분류}) \le \int_{R_1} (p(C_2|\boldsymbol{x})p(C_1|\boldsymbol{x}))^{1/2} p(\boldsymbol{x})d\boldsymbol{x} + \int_{R_2} (p(C_1|\boldsymbol{x})p(C_2|\boldsymbol{x}))^{1/2} p(\boldsymbol{x})d\boldsymbol{x}$
$= \int (p(C_1|\boldsymbol{x})p(C_2|\boldsymbol{x}))^{1/2} p(\boldsymbol{x})d\boldsymbol{x}$
$= \int (p(\boldsymbol{x}, C_1)p(\boldsymbol{x}, C_2))^{1/2} d\boldsymbol{x}$

1.27

$E[L_q] = \int\int |y(\boldsymbol{x}) - t|^q p(\boldsymbol{x}, t) d\boldsymbol{x}dt$
$= \int\int |y(\boldsymbol{x}) - t|^q p(t|\boldsymbol{x})p(\boldsymbol{x}) d\boldsymbol{x}dt$
$= \int [\int |y(\boldsymbol{x}) - t|^q p(t|\boldsymbol{x})dt] p(\boldsymbol{x}) d\boldsymbol{x}$

$y(\boldsymbol{x})$ 는 개별 $\boldsymbol{x}$와 독립적으로 선택될 수 있으므로
$\int |y(\boldsymbol{x}) - t|^q p(t|\boldsymbol{x})dt$의 최소화 조건을 찾는다.

이를 위해 아래 수식을 전개하고 정리하면
$\int |y(\boldsymbol{x}) - t|^q p(t|\boldsymbol{x})dt = \int_{-\infty}^{y(\boldsymbol{x})} (y(\boldsymbol{x}) - t)^q p(t|\boldsymbol{x})dt + \int_{y(\boldsymbol{x})}^{\infty} (t - y(\boldsymbol{x}))^q p(t|\boldsymbol{x})$
$\Rightarrow \int_{-\infty}^{y(\boldsymbol{x})} q(y(\boldsymbol{x}) - t)^{q-1} p(t|\boldsymbol{x})dt - \int_{y(\boldsymbol{x})}^{\infty} q(t - y(\boldsymbol{x}))^{q-1} p(t|\boldsymbol{x})dt = 0$
$\Rightarrow \int_{-\infty}^{y(\boldsymbol{x})} |y(\boldsymbol{x}) - t|^{q-1} p(t|\boldsymbol{x})dt = \int_{y(\boldsymbol{x})}^{\infty} |y(\boldsymbol{x}) - t|^{q-1} p(t|\boldsymbol{x})dt$

$q=1$인 경우,
$\int_{-\infty}^{y(\boldsymbol{x})} p(t|\boldsymbol{x})dt = \int_{y(\boldsymbol{x})}^{\infty} p(t|\boldsymbol{x})dt$
이므로 $y(\boldsymbol{x})$는 t의 conditional median이 된다.

$q \to 0$인 경우,
$|y(\boldsymbol{x})-t|^q$는 $t=y(\boldsymbol{x})$ 근처의 neighborhood에서만 0이 되고 다른곳에서는 1이 된다.
따라서 최소값을 가지려면 $y(\boldsymbol{x})$와 t는 근접한 부분이 많아야하고, 이에 따라 t가 가장 많이 발생하는 곳인 $p(t)$가 가장 큰 값을 가지는 위치에 $y(\boldsymbol{x})$가 있어야 한다.
따라서 이 값은 conditional mode가 된다.

1.30

(1.46) $N(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$
(1.104) $H[\mathbf{x}] = -\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$
(1.110) $H[x] = \frac{1}{2}[1+\ln(2\pi\sigma^2)], x \sim N(x|\mu,\sigma^2)$
(1.113) $KL(p\|q) = -\int p(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}$
$= -\int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x} + \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$

$KL(p\|q) = -\int p(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}$
$= -\int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x} + \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$
$= -\int p(x) \ln[\frac{1}{(2\pi s^2)^{1/2}} exp(-\frac{1}{2s^2}(x-m)^2)] d\mathbf{x} - H[x]$
$= -\int p(x)[-\frac{1}{2} ln(2\pi s^2) - \frac{1}{2s^2}(x^2 - 2mx + m^2)] d\mathbf{x} - H[x]$
$= \frac{1}{2} ln(2\pi s^2) + \frac{1}{2s^2}(\mu^2 + \sigma^2 - 2m\mu + m^2) - \frac{1}{2}(1 + \ln 2\pi\sigma^2)$
$= \frac{1}{2}[ln(\frac{s^2}{\sigma^2}) + \frac{\mu^2+\sigma^2-2m\mu+m^2}{s^2} - 1]$

1.33

$$H[y|x] = -\sum_i \sum_j p(x_i, y_j) \ln p(y_j|x_i) = -\sum_i \sum_j p(y_j|x_i) p(x_i) \ln p(y_j|x_i)$$

문제 조건에 따라 $H[y|x]$는 0이 된다.
이때 확률은 $[0,1]$ 사이 값을 가지므로 $p(x) > 0$인 x에 대해
$-p(y_j|x_i) \ln p(y_j|x_i)$ 값은 $[0,\infty)$ 사이의 값을 가지게된다.
이때 $H[y|x]=0$ 이어야 하므로 $-p(y_j|x_i) \ln p(y_j|x_i) = 0$이 되어야하는데
$\sum_j p(y_j|x_i) = 1$이므로 $p(y_j|x_i)$가 모두 0일 수는 없고
이에 따라 $\ln p(y_j|x_i) = 0 \Rightarrow p(y_j|x_i) = 1$ 이어야한다.

따라서 모든 x_i에 대해 특정 y_j인 경우 $p(y_j|x_i) = 1$이 성립하므로,
개별 x에 대해 $p(y|x) \neq 0$인 y값은 하나뿐임이 증명된다.

1.39

$$(a) H[x] = -\sum_i p(x_i) \ln p(x_i) = -(\tfrac{2}{3} \ln \tfrac{2}{3} + \tfrac{1}{3} \ln \tfrac{1}{3}) = \ln 3 - \tfrac{2}{3} \ln 2$$

$$(b) H[y] = -\sum_i p(y_i) \ln p(y_i) = -(\tfrac{1}{3} \ln \tfrac{1}{3} + \tfrac{2}{3} \ln \tfrac{2}{3}) = \ln 3 - \tfrac{2}{3} \ln 2$$

$$(c) H[y|x] = -\sum_i \sum_j p(x_i, y_j) \ln p(y_j|x_i)$$

$$= -(\tfrac{1}{3} \ln \tfrac{1}{2} + \tfrac{1}{3} \ln \tfrac{1}{2} + \tfrac{1}{3} \ln 1) = \tfrac{2}{3} \ln 2$$

$$(d) H[x|y] = -\sum_i \sum_j p(x_i, y_j) \ln p(x_i|y_j)$$

$$= -(\tfrac{1}{3} \ln 1 + \tfrac{1}{3} \ln \tfrac{1}{2} + \tfrac{1}{3} \ln \tfrac{1}{2}) = \tfrac{2}{3} \ln 2$$

$$(e) H[x, y] = -\sum_i \sum_j p(x_i, y_j) \ln p(x_i, y_j)$$

$$= -(\tfrac{1}{3} \ln \tfrac{1}{3} + \tfrac{1}{3} \ln \tfrac{1}{3} + \tfrac{1}{3} \ln \tfrac{1}{3}) = \ln 3$$

$$(f) I[x, y] = KL(p(x, y)) \| p(x)p(y))$$

$$= -\sum_x \sum_y p(x, y) \ln \tfrac{p(x)p(y)}{p(x,y)}$$

$$= -\sum_x \sum_y p(x, y) \ln p(x) - \sum_x \sum_y p(x, y) \ln \tfrac{p(y)}{p(x,y)}$$

$$= -\sum_x p(x) \ln p(x) + \sum_x \sum_y p(x, y) \ln p(x|y)$$

$$= H[x] - H[x|y] = H[y] - H[y|x] = \ln 3 - \tfrac{4}{3} \ln 2$$