VIVEK RHETAN ML: Assignment 3.

UKK287

1.1. $y(x, w) = w_0 + w_1 x + w_2 x^2 + \cdots + w_m x^m = \sum_{j=0}^{m} w_j x_n^j$ — (1.1)

$E(w) = \frac{1}{2} \sum_{n=1}^{N} \left[y(x_n - w) - t_n \right]^2$ --- (1.2)

$E(w) = \frac{1}{2} \sum_{n=1}^{N} \left\{ \sum_{j=0}^{m} w_j x_n^j - t_n \right]^2$

$\frac{\partial E(w)}{\partial (w_i)} = \sum_{n=1}^{N} \left\{ \sum_{j=0}^{m} w_j x_n^j - t_n \right] \left[\sum_{i=0}^{m} x_n^i \right]$

$= \sum_{n=1}^{N} \left\{ \sum_{j=0}^{m} \sum_{i=0}^{m} w_j x_n^j x_n^i - t_n \sum_{i=0}^{m} x_n^i \right] = 0$

$\Rightarrow \sum_{n=1}^{N} \left[\sum_{j=0}^{m} \left(\sum_{i=1}^{m} x_n^{(i+j)} \right) \cdot w_j \right\} = t_n \sum_{i=0}^{m} x_n^i$

$= \sum_{n=1}^{N} \left\{ \sum_{j=0}^{m} A_{ij} w_j = T_i \right\} = \sum_{n=1}^{N} \{ T_i \} = \sum_{n=1}^{N} \{ T_i \}$.

1.2 $E(w) = \frac{1}{2} \sum_{n=1}^{N} \left\{ y(x_n, w) - t_n \right]^2 + \frac{\lambda}{2} ||w||^2$

$E(w) = \frac{1}{2} \sum_{n=1}^{N} \left[\sum_{j=0}^{m} w_j x_n^j - t_n \right\}^2 + \frac{\lambda}{2} ||w||^2$

$\frac{\partial E}{\partial w} = \sum_{n=1}^{N} \left[\left[\sum_{j=0}^{m} \left(\sum_{i=1}^{m} x_n^{i+j} w_j \right) \right] \right\} - t_n \sum_{i=0}^{m} x_n^i \right\} + \lambda \sum_{j=0}^{m} w_j$

$\Rightarrow A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j}$

$T_i = \sum_{n=1}^{N} (x_n)^i t_n$

$\sum_{j=0}^{m} \left(A_{ij} w_j + \lambda w_j \right) = T_i$

ⓐ : apple ⓞ : orange Ⓛ : lemon.

1.3.

Red box	Blue box	green box
a a a L L L o o o o	a o	a a a o o o L L L L
$p(r) = 0.2$	$p(b) = 0.2$	$p(g) = 0.6$

$$p(\text{apple}) = p\left(\frac{\text{apple}}{\text{red box}}\right) + p\left(\frac{\text{apple}}{\text{blue box}}\right) + p\left(\frac{\text{apple}}{\text{green box}}\right)$$

$$p(a) = p(a/r) + p(a/b) + p(a/g)$$

$$= p(a,r) \cdot p(r) + p(a,b) \cdot p(b) + p(a,g) \cdot p(g)$$

$$= \frac{3}{10}(0.2) + \frac{1}{2}(0.2) + \frac{3}{10}(0.6) = .34$$

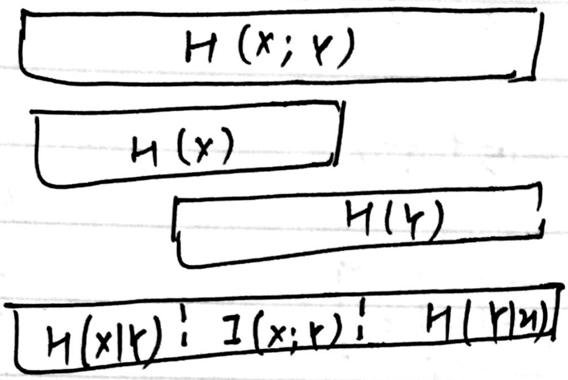
As.

$$p(\text{green}/\text{orange}) = \frac{p(\text{orange}|\text{green}) \cdot p(\text{green box})}{p(\text{orange})}$$

$$= \frac{p(\text{orange}/\text{green}) \cdot p(\text{green})}{p(\text{orange}/\text{green}) \cdot p(\text{green}) + p(\text{orange}/\text{red}) \cdot p(\text{red}) + p(\text{orange}/\text{blue}) \cdot p(\text{blue})}$$

$$= \frac{\frac{3}{10} \cdot (0.6)}{\frac{3}{10}(0.6) + \frac{4}{10}(0.2) + \frac{1}{2}(0.2)} = 0.5$$

1.39 $\quad I[x;y] = H[x] - H[x|y] = H[y] - H[y|x]$



1.40.

$$I[x,y] = KL(p(x,y)\,||\,p(x)p(y))$$
$$= -\iint p(x,y)\ \ln\left(\frac{p(x)\cdot p(y)}{p(x,y)}\right)\cdot dx\,dy$$

if $p(x,y) = p(y|x)\cdot p(x)$

$$I(x:y) = -\iint p(x,y)\ln p(y)\,dx\,dy + \iint p(x,y)\ln p(y|x)\,dx\,dy$$
$$= -\int p(y)\ln p(y)\,dy + \iint p(x,y)\ln p(y|x)\,dx\,dy$$
$$= H(y) - H(y|x)$$

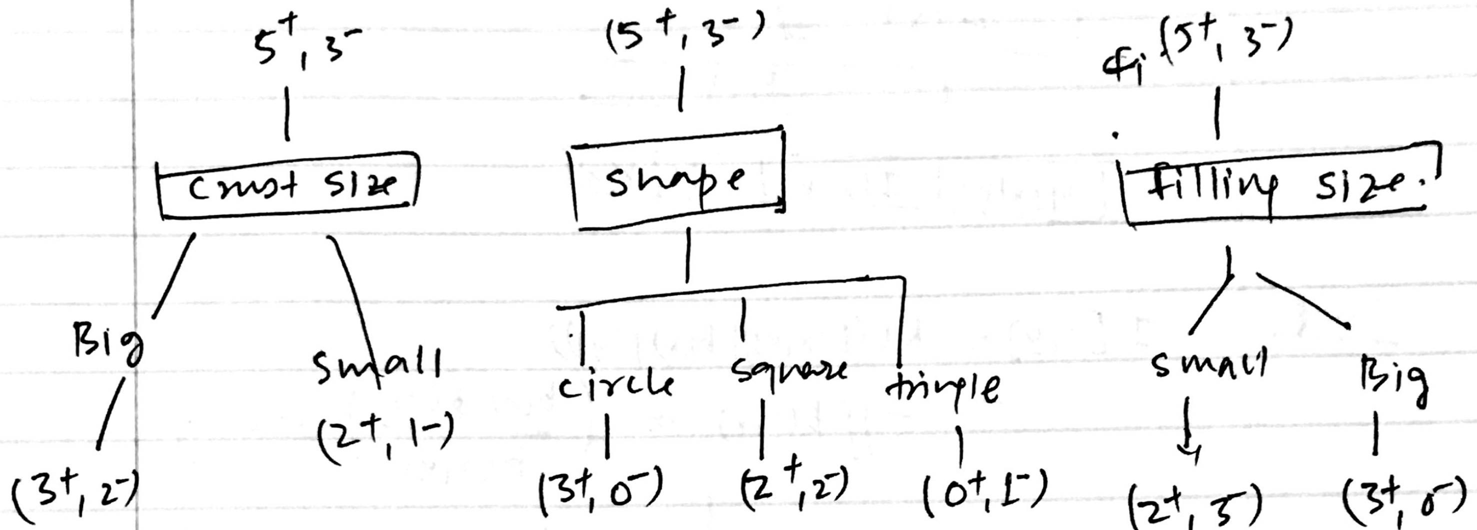
if $p(x,y) = p(x/y)\cdot p(y)$

$$= -\iint p(x,y)\ln p(x)\,dx\,dy + \iint p(x,y)\ln p(x|y)\,dx\,dy$$
$$= -\int p(x)\ln p(x)\,dx + \iint p(x,y)\ln p(x|y)\,dx\,dy$$
$$= H[x] - H[x/y]$$

$$\Rightarrow I[x;y] = H[y] - H[y|x] = H[x] - H[x|y]$$

Decision Tree Question

$$F_n = -\frac{3}{8}\log\frac{3}{8} - \frac{5}{8}\log\frac{5}{8} = 0.287$$



$$I(n, \text{crust size}) = F_n - \left[\frac{5}{8}\left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right) + \frac{3}{8}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)\right]$$

$$= F_n - (-0.1 - 0.18) = 0.287 - (0.281) = 0.007$$

$$I(n, \text{shape}) = F_n - \left[\frac{3}{8}(0) + \frac{-4}{8}\left(\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log\frac{2}{4}\right) + \frac{1}{8}(0)\right]$$

$$= 0.287 - 0.150 = 0.137$$

$$I(n, \text{filling size}) = F_n - \left[-5/8\left(\frac{2}{5}\log\frac{2}{5} + \frac{3}{5}\log\frac{3}{5}\right) + \frac{3}{8}(0)\right]$$

$$= 0.287 - 0.182 = 0.105$$

Maximum Information gain in selecting the Shape node.

7.4. for a given support vector

$$w^T x_i + b = d_i \quad \text{—— (1)}$$

we know that $\quad w = \sum_{j=1}^{m} \alpha_i d_i x_i$

and also $\quad \sum_{j=1}^{m} \alpha_i d_i = 0$

(1) will give

$$b = d_i - w^T x_i$$

$$b = d_i - \sum_{j=1}^{m} \alpha_j d_j x_j^T x_i \quad \text{— (2)}$$

Multiply ep 2 by $\sum_{j=1}^{m} \alpha_i d_i$

$$\sum_{j=1}^{m} \alpha_i d_i b = \sum_{j=1}^{m} \alpha_i d_i^2 - \sum_{j=1}^{m} \alpha_i d_i \sum_{j=1}^{m} \alpha_j d_j x_j^T x_j$$

The term on the LHS is zero and $d_i^2 = 1$ so:

$$0 = \sum_{j=1}^{m} \alpha_i - \sum_{j=1}^{m} \alpha_i d_i \sum_{j=1}^{m} \alpha_j d_j x_j^T x_i$$

Now using ep 1.

$$0 = \sum_{j=1}^{m} \alpha_i - w^T w.$$

But $\quad w^T w = \frac{1}{\rho^2}$

$$\Rightarrow \quad \frac{1}{\rho^2} = \sum_{n=1}^{N} a_n.$$

11.11. For repvired distribution $p(z)$ to be Invariance is to choose the transition probabilities to satisfy the property of detailed balance

$$p^*(z).\ T(z,z') = p^*(z').\ T(z',z) \quad \text{—(11.40)}$$

$$p^*(z) = \sum_{z_1} T(z',z).\ p^*(z') \quad \text{—(11.39)}$$

$$\Rightarrow \sum_{x} p^*(z').T(z',z) = \sum_{z_1} p^*(z).\ T(z,z')$$

$$= p^*(z).\ \sum_{z_1} p(z'|z) = p^*(z) \quad \text{—(11.41)}$$

$$T(z',z) = \sum_{K=1}^{K} \alpha_K B_K(z',z) \quad \Big| \ \sum_K \alpha_K = 1.$$

using 11.41.

$$p^*(z).\ T(z,z') = p^*(z').\ T(z',z) = \sum_{z} p^*(z).T(z,z')$$

$$= p^*(z).\ \sum_{z_1} p^*(z'|z)$$

$$= p^*(z_K|\{z_i\}_{i\neq K}).\ p^*(\{z_i\}_{i\neq K}).\ p^*(z_K'|\{z\}_{i\neq K})$$

$$= p^*(z_K|\{z_i\}_{i\neq K}).\ p^*(\{z_i'\}_{i\neq K}).\ p^*(z_K'|\{z'\}_{i\neq K})$$

$$= p^*(z_K|\{z_i'\}_{i\neq K}).\ p^*(z_K', \{z\}_{i\neq K}\}$$

$$= p^*(z').\ T(z',z).$$

1.40. $f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \le \sum_{i=1}^{M} \lambda_i f(\lambda_i)$ — for convex f.

as $f(x) = \ln x$  concave function

Jensen's inequality $f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \ge \sum_{i=1}^{M} \lambda_i f(x_i)$

$$= \ln\left(\sum_{i=1}^{M} (\lambda_i x_i)\right) = \sum_{i=1}^{M} \lambda_i \ln(x_i)$$

$$\Rightarrow \ln\left(\sum_{i=1}^{M} (\lambda_i x_i)\right) = \ln\left(\frac{\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_M x_M}{\lambda_1 + \lambda_2 + \dots \lambda_M}\right); \quad \lambda_1 + \lambda_2 + \dots \lambda_M = 1. \quad \text{— (i)}$$

$$\Rightarrow \sum_{i=1}^{M} \lambda_i \ln(x_i) = \sum_{i=1}^{M} \ln(x_i)^{\lambda_i} = \ln(x_1)^{\lambda_1} + \ln(x_2)^{\lambda_2} + \dots + \ln(x_M)^{\lambda_M}$$

$$= \ln\left(x_1^{\lambda_1} \cdot x_2^{\lambda_2} \cdot x_3^{\lambda_3} \dots x_M^{\lambda_M}\right) = \ln\left(x_1^{\lambda_1} \cdot x_2^{\lambda_2} \dots x_M^{\lambda_M}\right)^{(\lambda_1 + \dots \lambda_M)}$$

$$\Rightarrow \ln\left(\frac{\lambda_1 x_1 + \lambda_2 x_2 + \dots \lambda_M x_M}{\lambda_1 + \lambda_2 + \dots \lambda_M}\right) \ge \ln\left(x_1^{\lambda_1} \cdot x_2^{\lambda_2} \dots x_M^{\lambda_M}\right)^{(\lambda_1 + \dots \lambda_M)}$$

given $\sum_{i=1}^{M} \lambda_i = 1$

$\because$ ln is monotonically increasing

$\therefore$ if $f(a) > f(b) \Rightarrow a > b$

$$\Rightarrow \frac{\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_M x_M}{\lambda_1 + \lambda_2 + \dots + \lambda_M} \ge \left(x_1^{\lambda_1} \cdot x_2^{\lambda_2} \dots \lambda_M^{\lambda_M}\right)^{(\lambda_1 + \lambda_2 \dots}$$

$$\Rightarrow \boxed{A.M. \ge G.M.}$$