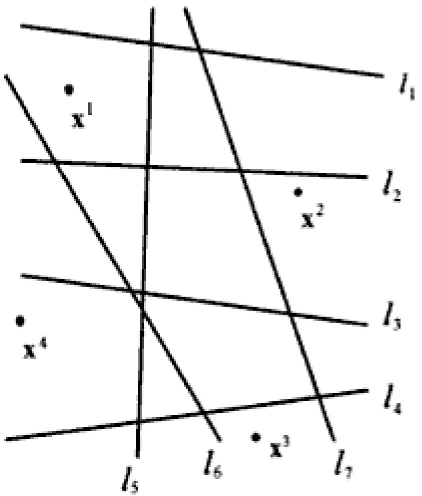
Solution 4

Prof. Krishna R. Pattipati

Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: krishna@engr.uconn.edu (860) 486-2890

ECE 6141
Neural Networks for Classification and Optimization

- Problem 1: For N samples in a p-dimensional feature space, the number of possible dichotomies (linearly separable groupings) is computed as follows.



Consider N points in general position $X = (\underline{x}^1, \underline{x}^2, \ldots, \underline{x}^N)$ in a p dimensional space.Let $O(N, p)$ be the number of linear dichotomies (X^+, X^-) of X. Suppose, we add a new point $\underline{x}^{N+1}$ such that the set of $(N+1)$ points is in a general position. Now some of the general dichotomies pass through $\underline{x}^{N+1}$. Let D be the number of linear dichotomies that pass through $\underline{x}^{N+1}$. For each of these dichotomies, there will be two new dichotomies $(X^+ \cup \underline{x}^{N+1}, X^-)$ $(X^+, X^- \cup \underline{x}^{N+1})$ by infinitismally perturbing the hyperplane passing through $\underline{x}^{N+1}$. Thus, there are $2D$ linear dichotomies. For each of the remaining $O(N, p) - D$ dichotomies, either $(X^+ \cup \underline{x}^{N+1}, X^-)$ or $(X^+, X^- \cup \underline{x}^{N+1})$ is separable. Thus,

$$O(N+1, p) = 2D + O(N, p) - D = O(N, p) + D$$

Each line gives two dichotomies. The Classes can be on either side (1,2) or (2,1). Number of possible dichotomies: 14 = 2(1+3+3)

- ## Problem 1:

Since a hyperplane passing through $\underline{x}^{N+1}$ has dimension (p-1), $D = O(N, p-1)$.

$So, O(N+1, p) = O(N, p) + O(N, p-1)$

If $N = 1,\ O(1, p)$=2; $O(N, 0) = 0$

$$O(N, p) = 2\sum_{l=0}^{p}\binom{N-1}{l} = 2[1 + (N-1) + (N-1)(N-2)/2 + ... + \binom{N-1}{p}]$$

$Why?$ Take z-transform wrt $N \Rightarrow zO(z, p) - 2z.z^{-1} = O(z, p) + O(z, p-1)$

$$\Rightarrow O(z, p) = \frac{1}{z-1}O(z, p-1) + \frac{2}{z-1} \Rightarrow O(z, p) = 2\sum_{l=0}^{p}\frac{1}{(z-1)^{l+1}}$$

$$\frac{1}{(z-1)^{l+1}} \Leftrightarrow \frac{(N-1)(N-2).....(N-l)}{l!} = \binom{N-1}{l}$$..... see Lathi, pp.498

$$\Rightarrow O(N, p) = 2\sum_{l=0}^{p}\binom{N-1}{l}$$

Total number of groupings: 2^N

$If\ \ N \le (p+1), O(N, p) = 2^N$

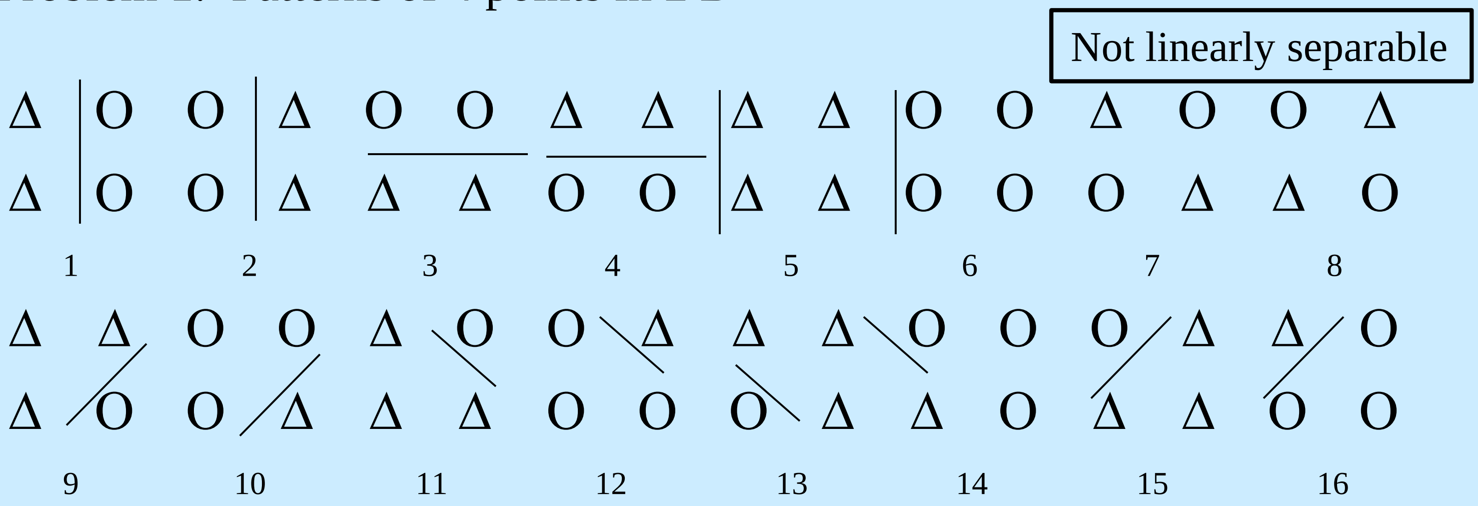
$so,$

$$f(N, p) = \begin{cases} 1\ if\ \ N \le p+1 \\ \frac{2}{2^N}\sum_{i=0}^{p}\binom{N-1}{i}\ if\ \ N > p+1 \end{cases}$$

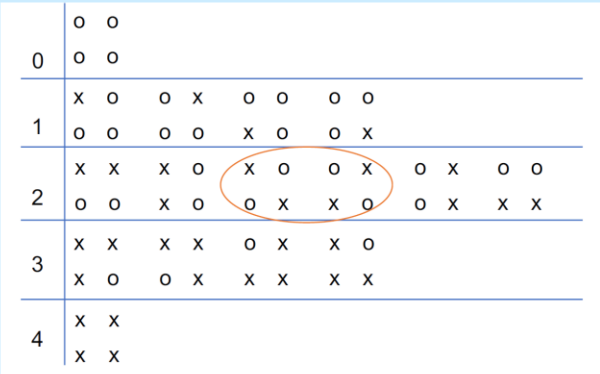
$\Rightarrow$ Binomial distribution with (N -1) flips resulting in p or fewer heads.

$f(N, p) \to 0\ as\ N \to \infty.$

- Problem 1: Patterns of 4 points in 2 D



2 out of 16 are not linearly separable ⇒ 7/8 are linearly separable



Good way to visualize

- Problem 2

$$\phi_m = \frac{e^{t_m}}{\sum_{k=1}^{C} e^{t_k}}; m = 1, 2, \ldots, C$$

$$\frac{\partial \phi_m}{\partial t_m} = \frac{e^{t_m}}{\sum_{k=1}^{C} e^{t_k}} - \frac{\left(e^{t_m}\right)^2}{\left(\sum_{k=1}^{C} e^{t_k}\right)^2} = \frac{e^{t_m}}{\sum_{k=1}^{C} e^{t_k}}\left[1 - \frac{e^{t_m}}{\sum_{k=1}^{C} e^{t_k}}\right] = \phi_m(1-\phi_m)$$

$$\frac{\partial \phi_m}{\partial t_k} = -\frac{e^{t_m} e^{t_k}}{\left(\sum_{k=1}^{C} e^{t_k}\right)^2} = -\phi_m \phi_k$$

Thus,

$$\frac{\partial \phi_m}{\partial t_k} = \phi_m(\delta_{mk} - \phi_k); m = 1, 2 \ldots, C; k = 1, 2, \ldots, C$$

where

$$\delta_{mk} = \begin{cases} 1, \text{if } m = k \\ 0, \text{otherwise} \end{cases}$$

- Problem 3

$$z(\underline{x},\underline{w}) = \frac{1}{(1+\|\underline{x}-\underline{w}\|^2)}$$

$$\nabla_{\underline{w}} z(\underline{x},\underline{w}) = \frac{2(\underline{x}-\underline{w})}{(1+\|\underline{x}-\underline{w}\|^2)^2} \propto [z(\underline{x},\underline{w})]^2(\underline{x}-\underline{w})$$

$$Evidently, \underline{w}^{(m+1)} = \underline{w}^{(m)} - \eta\,[z(\underline{x},\underline{w})]^2\,(\underline{x}-\underline{w})$$

- Problem 3

$$a)\sum_{k=1}^{\infty}\eta_k = \sum_{k=1}^{100}1 + \frac{1}{1000}\sum_{k=101}^{500}k + \sum_{k=501}^{\infty}\frac{1}{200+k} \approx 100 + 120.2 + \ln(200+x)\,|_{501}^{\infty} = \infty$$

$$\sum_{k=1}^{\infty}\eta_k^2 = \sum_{k=1}^{100}1 + \frac{1}{(1000)^2}\sum_{k=101}^{500}k^2 + \sum_{k=501}^{\infty}\frac{1}{(200+k)^2} \approx 100 + 41.4534 - \frac{1}{(200+x)}\,|_{501}^{\infty} = 141.452 < \infty$$

$$b)\ \eta_k = \eta_0\frac{(k/K)+1}{(k/K)^2+1};\ 0<\eta_0<1$$

$$\sum_{k=1}^{\infty}\eta_k = \eta_0\sum_{k=1}^{\infty}\frac{(k/K)+1}{(k/K)^2+1} = \eta_0[\int_1^{\infty}\frac{(x/K)}{(x/K)^2+1}dx + \int_1^{\infty}\frac{1}{(x/K)^2+1}dx = \eta_0[K\ln(1+(x/K)^2)\,|_1^{\infty} + K\frac{\pi}{4}] = \infty$$

$$\sum_{k=1}^{\infty}\eta_k^2 = \eta_0^2\sum_{k=1}^{\infty}(\frac{(k/K)+1}{(k/K)^2+1})^2 = \eta_0^2[\int_1^{\infty}\frac{[(x/K)+1]^2}{[(x/K)^2+1]^4}dx = \eta_0^2K(\frac{\pi}{4}+1) < \infty$$

- Problem 4

$y(\underline{x},\underline{w}) = P(t=1|\underline{x})$ **output**

t **true label**

z **measured label**

$$Z^N = [z^1, z^2, ..., z^N]$$

$$\underline{X}^N = [\underline{x}^1, \underline{x}^2, ..., \underline{x}^N]$$

$$P(z=1|t=1) = P(z=0|t=0) = 1-\varepsilon$$

$$P(z=1|t=0) = P(z=0|t=1) = \varepsilon$$

$$P(z=1|\underline{x}) = P(z=1|t=1)\,P(t=1|\underline{x}) + P(z=1|t=0)\,P(t=0|\underline{x})$$

$$= (1-\varepsilon)y(\underline{x},\underline{w}) + \varepsilon(1-y(\underline{x},\underline{w})) = \varepsilon + (1-2\varepsilon)y(\underline{x},\underline{w})$$

$$P(z=0|\underline{x}) = (1-\varepsilon) - (1-2\varepsilon)y(\underline{x},\underline{w})$$

$$P(z|\underline{x},\underline{w}) = [\varepsilon + (1-2\varepsilon)y(\underline{x},\underline{w})]^z[(1-\varepsilon) - (1-2\varepsilon)y(\underline{x},\underline{w})]^{1-z}$$

$$P(Z^N|X^N) = \sum_{T^N} P(Z^N|T^N)P(T^N|X^N) = \sum_{T^N}\prod_{n=1}^{N} P(z_n|t_n)P(t_n|x_n)$$

$$= \prod_{n=1}^{N}\sum_{t_n=0}^{1} P(z_n|t_n)P(t_n|x_n)$$

$$P(Z^N|\underline{X}^N,\underline{w}) = \prod_{n=1}^{N}[\varepsilon + (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})]^{z^n}[(1-\varepsilon) - (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})]^{1-z^n}$$

$$E(\underline{w}) = -\ln P(Z^N|\underline{X}^N,\underline{w}) = -\sum_{n=1}^{N}\{z^n \ln[\varepsilon + (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})] + (1-z^n)\ln[(1-\varepsilon) - (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})]\}$$

$$= -\sum_{n=1}^{N}\{z^n \ln\left(\frac{\varepsilon + (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})}{(1-\varepsilon) - (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})}\right) + \ln[(1-\varepsilon) - (1-2\varepsilon)y^n(\underline{x}^n,\underline{w})]\}$$

when $\varepsilon = 0,\ E(\underline{w})$ **simplifies to (5.21).**

- Problem 5

$$t^n = \begin{cases} N/N_1 \ \textbf{for class } C_1 \\ -N/N_2 \ \textbf{for class } C_2 \end{cases} \Rightarrow \frac{1}{N}\sum_{n=1}^{N} \underline{x}^n = \frac{1}{N}(N_1\underline{m}_1 + N_2\underline{m}_2); \sum_{n=1}^{N} t^n = 0; \sum_{n=1}^{N} t^n \underline{x}^n = N(\underline{m}_1 - \underline{m}_2)$$

$$\sum_{n=1}^{N} \underline{x}^n \underline{x}^{n^T} = \sum_{n\in C_1} (\underline{x}^n - \underline{m}_1)(\underline{x}^n - \underline{m}_1)^T + \sum_{n\in C_2} (\underline{x}^n - \underline{m}_2)(\underline{x}^n - \underline{m}_2)^T + N_1\underline{m}_1\underline{m}_1^T + N_2\underline{m}_2\underline{m}_2^T$$

$$= S_W + N_1\underline{m}_1\underline{m}_1^T + N_2\underline{m}_2\underline{m}_2^T$$

$$E = \frac{1}{2}\sum_{n=1}^{N} (\underline{w}^T \underline{x}^n + w_0 - t^n)^2$$

$$\nabla_{w_0} E = \sum_{n=1}^{N} (\underline{w}^T \underline{x}^n + w_0 - t^n) = 0 \Rightarrow w_0 = -\frac{\underline{w}^T (N_1\underline{m}_1 + N_2\underline{m}_2)}{N}$$

$$\nabla_{\underline{w}} E = \sum_{n=1}^{N} (\underline{w}^T \underline{x}^n + w_0 - t^n)\underline{x}^n = \underline{0}$$

$$\Rightarrow \left[\sum_{n=1}^{N} \underline{x}^n \underline{x}^{n^T} - \frac{(N_1\underline{m}_1 + N_2\underline{m}_2)}{N}(N_1\underline{m}_1 + N_2\underline{m}_2)^T\right]\underline{w} = \sum_{n=1}^{N} t^n \underline{x}^n = N(\underline{m}_1 - \underline{m}_2)$$

$$\Rightarrow \left(S_W + \frac{N_1 N_2}{N} S_B\right)\underline{w} = N(\underline{m}_1 - \underline{m}_2); S_B = (\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^T$$

$$\Rightarrow \underline{w} \propto S_W^{-1}(\underline{m}_1 - \underline{m}_2)$$

- Problem 6

$$p(\underline{x}_n \mid \underbrace{\{\underline{p},\underline{q},\{\underline{\mu}\}_{j=1}^{m},\{\sigma_k^2\}_{k=1}^{l}\}}_{\underline{\theta}}) = \sum_{j=1}^{m}\sum_{k=1}^{l} p_j q_k N(\underline{x}_n \mid \underline{\mu}_j, \sigma_k^2)$$

Define latent variables

$$z_{nj} = \begin{cases} 1 \text{ if } j^{th} \text{ component of mean is active} \\ 0, \text{ otherwise} \end{cases} ; n = 1,2,..,N;\ j = 1,2,..,m$$

$$y_{nk} = \begin{cases} 1 \text{ if } k^{th} \text{ component of covariance is active} \\ 0, \text{ otherwise} \end{cases} ; n = 1,2,..,N;\ k = 1,2,..,l$$

Then

$$p(\underline{x}_n \mid \{\underline{p},\underline{q},\{\underline{\mu}\}_{j=1}^{m},\{\sigma_k^2\}_{k=1}^{l}\} = \prod_{j=1}^{m}\prod_{k=1}^{l} p_j^{z_{nj}} q_k^{y_{nk}} \left[N(\underline{x}_n \mid \underline{\mu}_j, \sigma_k^2)\right]^{z_{nj}y_{nk}}$$

Complete data likelihood :

$$L(\{\underline{x}_n, \underline{z}_n, \underline{y}_n\}_{n=1}^{N} \mid \{\underline{p},\underline{q},\{\underline{\mu}\}_{j=1}^{m},\{\sigma_k^2\}_{k=1}^{l}\}) = \prod_{n=1}^{N}\prod_{j=1}^{m}\prod_{k=1}^{l} p_j^{z_{nj}} q_k^{y_{nk}} \left[N(\underline{x}_n \mid \underline{\mu}_j, \sigma_k^2)\right]^{z_{nj}y_{nk}}$$

NLL :

$$J = -\ln L = \sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{k=1}^{l}\{-z_{nj}\ln p_j - y_{nk}\ln q_k + \frac{z_{nj}y_{nk}}{2}[p\ln 2\pi + p\ln\sigma_k^2 + \frac{(\underline{x}_n-\underline{\mu}_j)^T(\underline{x}_n-\underline{\mu}_j)}{\sigma_k^2}]$$

- Problem 6

$NLL:$

$$J = -\ln L = \sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{k=1}^{l}\{-z_{nj}\ln p_j - y_{nk}\ln q_k + \frac{z_{nj}y_{nk}}{2}[p\ln 2\pi + p\ln\sigma_k^2 + \frac{(\underline{x}_n-\underline{\mu}_j)^T(\underline{x}_n-\underline{\mu}_j)}{\sigma_k^2}]$$

$$s.t. \sum_{j=1}^{m} p_j = 1; p_j \geq 0; \sum_{j=1}^{m} q_j = 1; q_j \geq 0$$

$E - step:$

$$Q(\underline{\theta},\underline{\theta}^{old}) = \sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{k=1}^{l}\{-E\{z_{nj}\}\ln p_j - E\{y_{nk}\}\ln q_k + \frac{E\{z_{nj}y_{nk}\}}{2}[p\ln 2\pi + p\ln\sigma_k^2 + \frac{(\underline{x}_n-\underline{\mu}_j)^T(\underline{x}_n-\underline{\mu}_j)}{\sigma_k^2}]$$

$$E\{z_{nj}\} = \gamma_{nj} = P(z_{nj}=1|\underline{x}_n,\underline{\theta}^{old}) = \frac{p_j^{old}\left[\sum_{k=1}^{l} q_k^{old} N(\underline{x}_n|\underline{\mu}_j^{old},(\sigma_k^2)^{old})\right]}{\sum_{i=1}^{m} p_i^{old}\sum_{k=1}^{l} q_k^{old} N(\underline{x}_n|\underline{\mu}_i^{old},(\sigma_k^2)^{old})}$$

$$E\{y_{nk}\} = \delta_{nk} = P(y_{nk}=1|\underline{x}_n,\underline{\theta}^{old}) = \frac{q_k^{old}\left[\sum_{j=1}^{m} p_j^{old} N(\underline{x}_n|\underline{\mu}_j^{old},(\sigma_k^2)^{old})\right]}{\sum_{j=1}^{m} p_j^{old}\sum_{i=1}^{l} q_i^{old} N(\underline{x}_n|\underline{\mu}_j^{old},(\sigma_i^2)^{old})}$$

$$E\{z_{nj}y_{nk}\} = \pi_{njk} = P(z_{nj}=1 \,\&\, y_{nk}=1|\underline{x}_n,\underline{\theta}^{old}) = \frac{p_j^{old} q_k^{old} N(\underline{x}_n|\underline{\mu}_j^{old},(\sigma_k^2)^{old})}{\sum_{r=1}^{m} p_r^{old}\sum_{i=1}^{l} q_i^{old} N(\underline{x}_n|\underline{\mu}_r^{old},(\sigma_i^2)^{old})}$$

- Problem 6

$M - step:$

$$Q(\underline{\theta},\underline{\theta}^{old}) = \sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{k=1}^{l} -\gamma_{nj}\ln p_j - \delta_{nk}\ln q_k + \frac{\pi_{njk}}{2}[p\ln 2\pi + p\ln\sigma_k^2 + \frac{(\underline{x}_n - \underline{\mu}_j)^T(\underline{x}_n - \underline{\mu}_j)}{\sigma_k^2}]$$

$$s.t. \sum_{j=1}^{m} p_j = 1; p_j \geq 0; \sum_{j=1}^{m} q_j = 1; q_j \geq 0$$

$$p_j^{new} = \frac{N_j}{N}; N_j = \sum_{n=1}^{N}\gamma_{nj} = \sum_{n=1}^{N}\sum_{k=1}^{l}\pi_{njk}$$

$$q_k^{new} = \frac{L_k}{N}; L_k = \sum_{n=1}^{N}\delta_{nk} == \sum_{n=1}^{N}\sum_{j=1}^{m}\pi_{njk}$$

$$\frac{\partial Q}{\partial \underline{\mu}_j} = 0 \Rightarrow \underline{\mu}_j^{new} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{l}\frac{\pi_{njk}}{\sigma_k^{2,new}}\underline{x}_n}{\sum_{n=1}^{N}\sum_{k=1}^{l}\frac{\pi_{njk}}{\sigma_k^{2,new}}}$$

$$\frac{\partial Q}{\partial \sigma_k} = 0 \Rightarrow \sigma_k^{2,new} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{m}\pi_{njk}(\underline{x}_n - \underline{\mu}_j^{new})^T(\underline{x}_n - \underline{\mu}_j^{new})}{p\sum_{n=1}^{N}\sum_{j=1}^{m}\pi_{njk}} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{m}\pi_{njk}(\underline{x}_n - \underline{\mu}_j^{new})^T(\underline{x}_n - \underline{\mu}_j^{new})}{pL_k}$$

You could also just do
a gradient step for M-step

iterate

Diagnostic Tree – Example (Problem 6)

Patterns with binary valued attributes

Z=1	Z=0
1100	1100
0000	1111
1010	1110
0011	0111

$H(z) = H_b(1/2) = 1; H_b(p) = \textit{Binary entropy with probability } p$

$H(z|T_1) = (5/8)H_b(2/5) + (3/8)H_b(2/3) = 0.9512 \Rightarrow IG(z,T_1) = 0.0488$

$H(z|T_2) = (5/8)H_b(1/5) + (3/8)H_b(0) = 0.4512 \Rightarrow IG(z,T_2) = 0.5488$

$H(z|T_3) = (5/8)H_b(2/5) + (3/8)H_b(2/3) = 0.9512 \Rightarrow IG(z,T_3) = 0.0488$

$H(z|T_4) = (3/8)H_b(1/3) + (5/8)H_b(3/5) = 0.9512 \Rightarrow IG(z,T_4) = 0.0488$

Select Test T_2

$T_2 = 0 \Rightarrow z = 1$

when $T_2 = 1$, *the reduced cases are* →

Z=1	Z=0
1100	1100
	1111
	1110
	0111

$H(z_1) = H_b(1/5) = 0.7219$

$H(z_1|T_1) = (4/5)H_b(1/4) + (1/5)H_b(0) = 0.6490 \Rightarrow IG(z_1,T_1) = 0.0729$

$H(z_1|T_3) = (3/5)H_b(1) + (2/5)H_b(1/2) = 0.40 \Rightarrow IG(z_1,T_3) = 0.3219$

$H(z_1|T_4) = (2/5)H_b(1) + (3/5)H_b(1/3) = 0.5510 \Rightarrow IG(z_1,T_4) = 0.1709$

Select Test T_3

$T_3 = 1 \Rightarrow z = 0$

$T_3 = 0 \Rightarrow z = 0 \textit{ or } z = 1$ →

Z=1	Z=0
1100	1100

Diagnostic Tree – Example (Problem 6) Alternate Way of Computing IG

Patterns with binary valued attributes

Z=1	Z=0
1100	1100
0000	1111
1010	1110
0011	0111

$H(T_1) = H(T_2) = H(T_3) = H_b(5/8) = H_b(3/8) = 0.9544;$

$H(T_1 \mid z) = (1/2)H_b(1/2) + (1/2)H_b(3/4) = 0.9056 \Rightarrow IG(z, T_1) = 0.0488$

$H(T_2 \mid z) = (1/2)H_b(1/4) + (1/2)H_b(1) = 0.4056 \Rightarrow IG(z, T_2) = 0.5488$

$H(T_3 \mid z) = (1/2)H_b(1/2) + (1/2)H_b(3/4) = 0.9056 \Rightarrow IG(z, T_3) = 0.0488$

$H(T_4 \mid z) = (1/2)H_b(1/4) + (1/2)H_b(1/2) = 0.9056 \Rightarrow IG(z, T_4) = 0.0488$

Select Test T_2

$T_2 = 0 \Rightarrow z = 1$

when $T_2 = 1$, *the reduced cases are* →

Z=1	Z=0
1100	1100
	1111
	1110
	0111

$H(T_1) = H_b(4/5) = 0.7219; H(T_2) = 0\ (as\ it\ should);$

$H(T_3) = H(T_4) = H_b(3/5) = 0.9709$

$H(T_1 \mid z_1) = (1/5)H_b(1) + (4/5)H_b(3/4) = 0.6490 \Rightarrow IG(z_1, T_1) = 0.0729$

$H(T_3 \mid z_1) = (1/5)H_b(1) + (4/5)H_b(3/4) = 0.6490 \Rightarrow IG(z_1, T_3) = 0.3219$

$H(T_4 \mid z_1) = (1/5)H_b(1) + (4/5)H_b(1/2) = 0.80 \Rightarrow IG(z_1, T_4) = 0.1709$

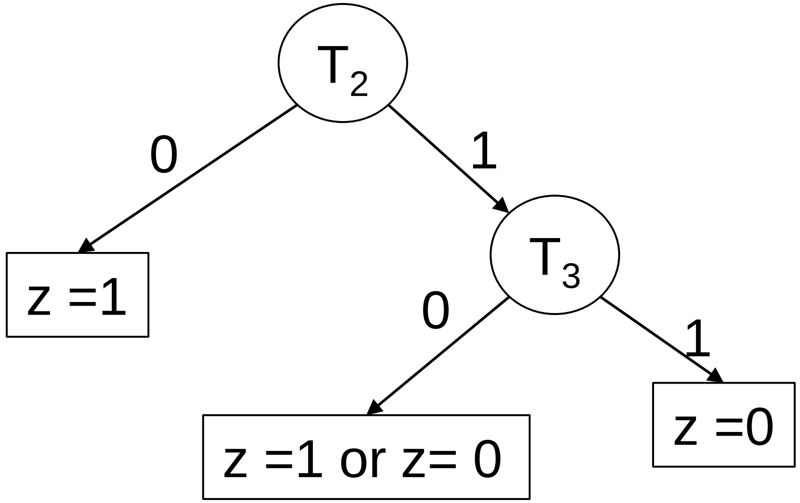
Select Test T_3

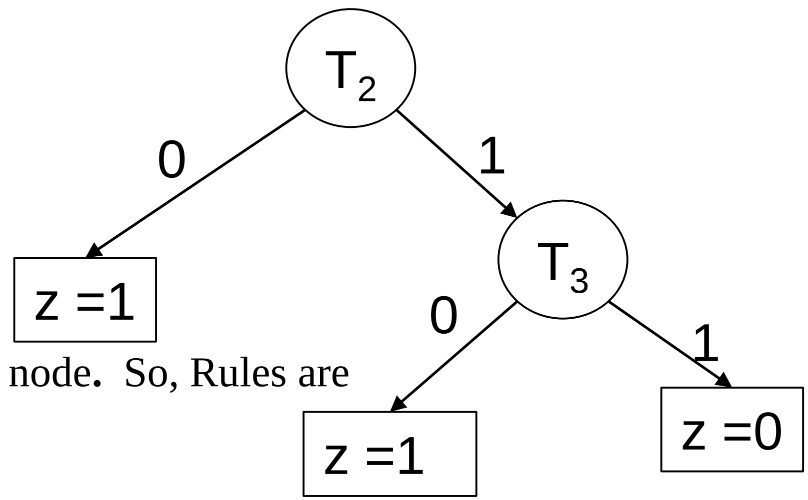
$T_3 = 1 \Rightarrow z = 0$

$T_3 = 0 \Rightarrow z = 0\ or\ z = 1$ →

Z=1	Z=0
1100	1100

Diagnostic Tree – Example (Problem 6)





- Same tree for part b.
- Rules: $z = 1 : \overline{T}_2 \vee (T_2 \overline{T}_3 \ wp\ 1/2)$

  $z = 0 : T_2 T_3 \vee (T_2 \overline{T}_3 \ wp\ 1/2)$

- For part c, make decision $z = 1$ at ambiguity node. So, Rules are

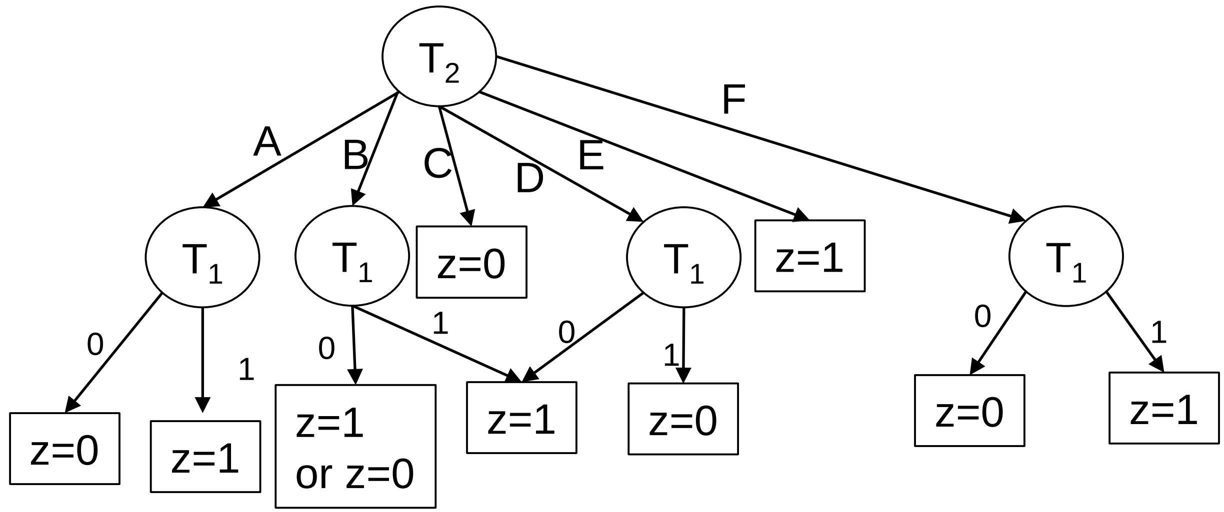
  $z = 1 : \overline{T}_2 \vee T_2 \overline{T}_3$

  $z = 0 : T_2 T_3$

Problem 7.a

Patterns

Z=1	Z=0
1A	0A
0E	0C
0B	1C
1B	0F
1F	0B
0D	1D



$$H(z) = H_b(1/2) = 1; H_b(p) = \textit{Binary entropy with probability } p$$

$$H(z \mid T_1) = (5/12)H_b(3/5) + (7/12)H_b(3/7) = 0.4046 + 0.5747 = 0.9793 \Rightarrow IG(z, T_1) = 0.0207$$

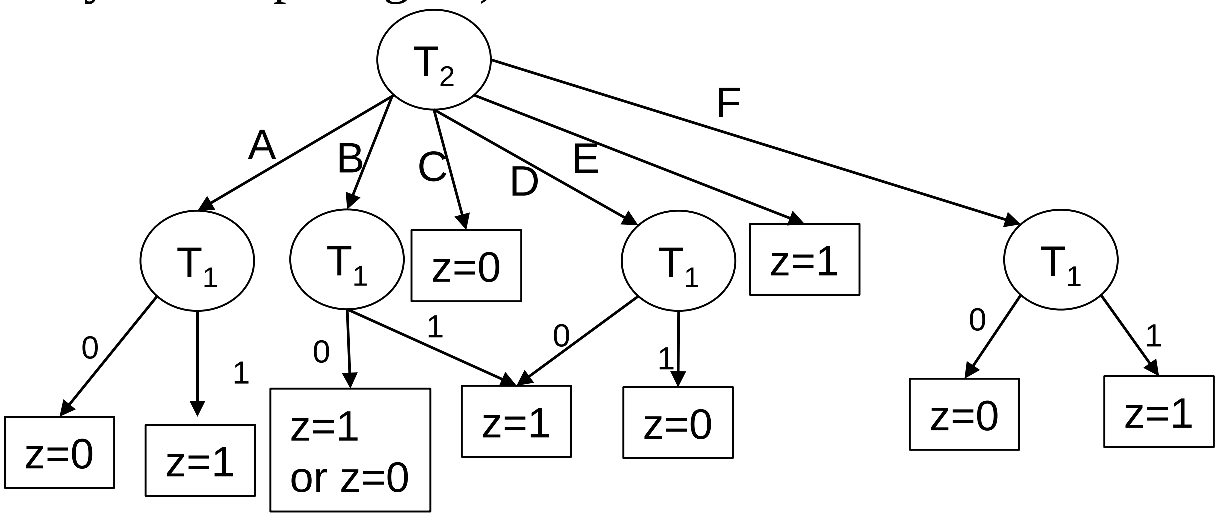
$$H(z \mid T_2) = (1/6)H_b(1/2) + (1/4)H_b(2/3) + (1/6)H_b(1) + (1/6)H_b(1/2) + (1/12)H_b(1) + (1/6)H_b(1/2)$$
$$= 0.7296 \Rightarrow IG(z, T_2) = 0.2704$$

Select Test T_2 and then T_1 as needed.

Problem 7.a (Another way of computing IG)

Patterns

Z=1	Z=0
1A	0A
0E	0C
0B	1C
1B	0F
1F	0B
0D	1D



$$H(T_1) = H_b(5/12) = 0.9799$$

$$H(T_2) = -[(1/6)\log_2(1/6) + (1/4)\log_2(1/4) + (1/6)\log_2(1/6) + (1/6)\log_2(1/6) + (1/12)\log_2(1/12) + (1/6)\log_2(1/6)] = 2.5221$$

$$H(T_1 \mid z) = (1/2)H_b(1/2) + (1/2)H_b(1/3) = 0.9592 \Rightarrow IG(z,T_1) = 0.0207$$

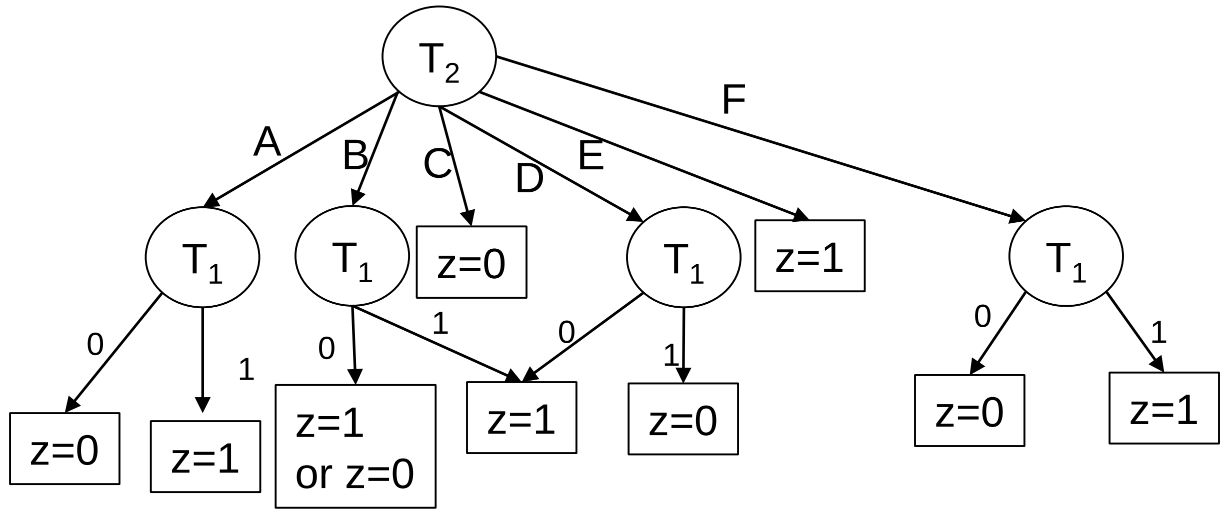
$$H(T_2 \mid z) = -(1/2)[(4/6)\log_2(1/6) + (1/3)\log_2(1/3)] - (1/2)[(4/6)\log_2(1/6) + (1/3)\log_2(1/3)] = 2.2517 \Rightarrow IG(z,T_2) = 0.2704$$

Select Test T_2 *and then* T_1 *as needed.*

Problem 7.b

Patterns

Z=1	Z=0
1A	0A
0E	0C
0B	1C
1B	0F
1F	0B
0D	1D



$$Let\ G_b(p) = 1 - p^2 - (1-p)^2 = 2p(1-p)$$

$$Gini(z) = 1/2$$

$$G(z \mid T_1) = (5/12)G_b(3/5) + (7/12)G_b(3/7) = 0.4857 \Rightarrow G(z,T_1) = 0.0143$$

$$G(z \mid T_2) = (1/6)G_b(1/2) + (1/4)G_b(2/3) + (1/6)G_b(1) + (1/6)G_b(1/2) + (1/12)G_b(1) + (1/6)G_b(1/2)$$

$$= 0.3611 \Rightarrow IG(z,T_2) = 0.1389$$

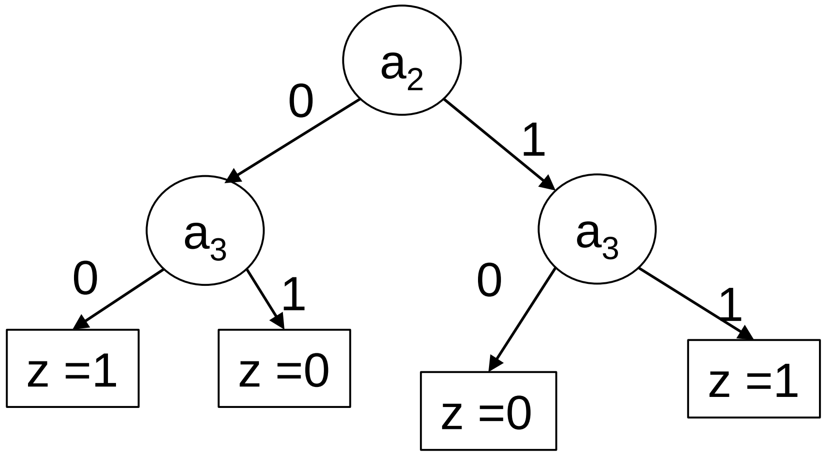
Select Test T_2 *and then* T_1 *as needed.*

Same Tree as IG

Problem 8

Attribute/ Sample	1	2	3	Class
1	1	1	1	1
2	1	0	0	1
3	1	1	0	0
4	0	0	1	0

Optimal strategy: a_2 *followed by* a_3 *for both outcomes.*



Rule: $a_2 = a_3 \Rightarrow z = 1$;
else z=0

Problem 8

$H(z) = H_b(1/2) = 1; H_b(p) = \text{Binary entropy with probability } p$

$H(z \mid a_1) = (3/4)H_b(2/3) + (1/4)H_b(1) = 0.6887 \Rightarrow IG(z, a_1) = 0.3113$

$H(z \mid a_2) = (1/2)H_b(1/2) + (1/2)H_b(1/2) = 1 \Rightarrow IG(z, a_2) = 0.$

$H(z \mid a_3) = (1/2)H_b(1/2) + (1/2)H_b(1/2) = 1 \Rightarrow IG(z, a_3) = 0.$

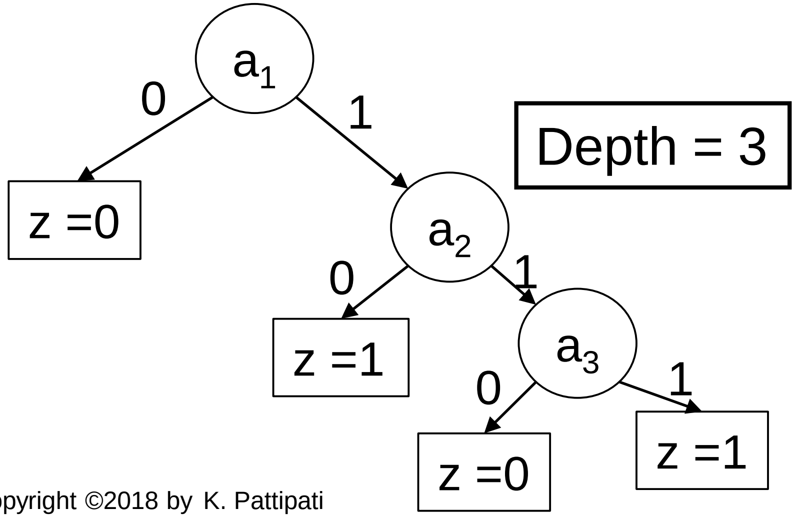
Select attribute a_1 *first*; $a_1 = 0 \Rightarrow z = 0; a_1 = 1 \Rightarrow$ *need to split samples* $1-3$.

$H(z_1) = H_b(2/3) = 0.9183$

$H(z_1 \mid a_2) = (2/3)H_b(1/2) + (1/3)H_b(1) = 2/3 \Rightarrow IG(z_1, a_2) = 0.2516$

$H(z_1 \mid a_3) = (1/3)H_b(1) + (2/3)H_b(1/2) = 2/3 \Rightarrow IG(z_1, a_3) = 0.2516$

Select attribute a_2; $a_2 = 0 \Rightarrow z = 1; a_2 = 1 \Rightarrow$ *need to split samples* $1 \& 3$. *Select attribute* a_3 *to split them.*



Attribute/ Sample	1	2	3	Class
1	1	1	1	1
2	1	0	0	1
3	1	1	0	0
4	0	0	1	0

Problem 8 (Another way of computing IG)

$H(a_1) = H_b(3/4) = 0.8113; H(a_2) = H(a_3) = H_b(1/2) = 1$

$H(a_1 | z) = (1/2)H_b(1) + (1/2)H_b(1/2) = 0.5 \Rightarrow IG(z, a_1) = 0.3113$

$H(a_2 | z) = (1/2)H_b(1/2) + (1/2)H_b(1/2) = 1 \Rightarrow IG(z, a_2) = 0.$

$H(z | a_3) = (1/2)H_b(1/2) + (1/2)H_b(1/2) = 1 \Rightarrow IG(z, a_3) = 0.$

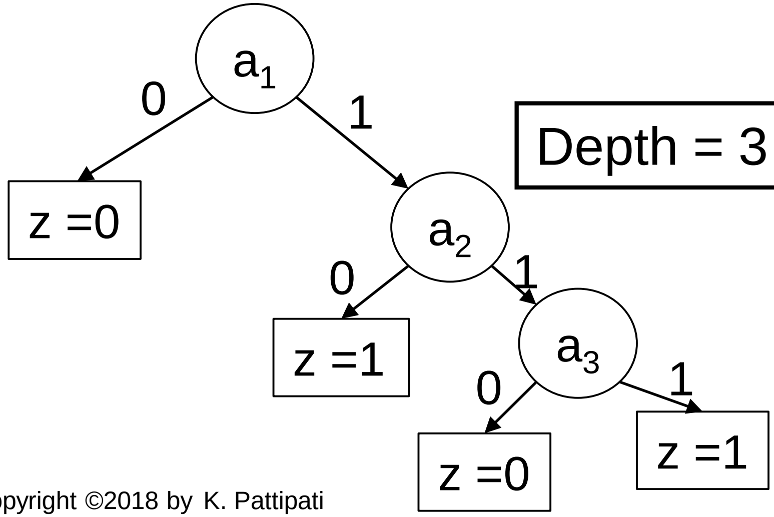
Select attribute a_1 *first*; $a_1 = 0 \Rightarrow z = 0; a_1 = 1 \Rightarrow$ *need to split samples* $1-3.$

$H(a_1) = 0; H(a_2) = H(a_3) = H_b(2/3) = 0.9183$

$H(a_2 | z_1) = (2/3)H_b(1/2) + (1/3)H_b(1) = 2/3 \Rightarrow IG(z_1, a_2) = 0.2516$

$H(a_3 | z_1) = (1/3)H_b(1) + (2/3)H_b(1/2) = 2/3 \Rightarrow IG(z_1, a_3) = 0.2516$

Select attribute a_2; $a_2 = 0 \Rightarrow z = 1; a_2 = 1 \Rightarrow$ *need to split samples* $1 \& 3.$ *Select attribute* a_3 *to split them.*



Depth = 3

Attribute/ Sample	1	2	3	Class
1	1	1	1	1
2	1	0	0	1
3	1	1	0	0
4	0	0	1	0