Title. Date. Page.

1.1) Classification error

$T_{1,L} \Rightarrow \frac{50}{200} = 0.25$ $\qquad T_{2,L} \Rightarrow 0$

$T_{1,R} \Rightarrow \frac{50}{200} = 0.25$ $\qquad T_{2,R} \Rightarrow \frac{100}{300} = 0.67$

Entropy

$T_{1,L} \Rightarrow -\left(\frac{150}{200}\ln\frac{150}{200} + \frac{50}{200}\ln\frac{50}{200}\right) = 0.562$

$T_{1,R} \Rightarrow -\left(\frac{50}{200}\ln\frac{50}{200} + \frac{150}{200}\ln\frac{150}{200}\right) = 0.562$

$T_{2,L} \Rightarrow -\left(\frac{0}{100}\ln\frac{0}{100} + \frac{100}{100}\ln\frac{100}{100}\right) = 0$

$T_{2,R} \Rightarrow -\left(\frac{200}{300}\ln\frac{200}{300} + \frac{100}{300}\ln\frac{100}{300}\right) = 0.637$

Gini

$T_{1,R} \Rightarrow \left[\frac{150}{200}\left(1-\frac{150}{200}\right) + \frac{50}{200}\left(1-\frac{50}{200}\right)\right] = 0.375$

$T_{1,R} \Rightarrow \left[\frac{50}{200}\left(1-\frac{50}{200}\right) + \frac{150}{200}\left(1-\frac{150}{200}\right)\right] = 0.375$

$T_{2,L} \Rightarrow \frac{0}{100}\left(1-\frac{0}{100}\right) + \frac{100}{100}\left(1-\frac{100}{100}\right) = 0$

$T_{2,R} \Rightarrow \frac{200}{300}\left(1-\frac{200}{300}\right) + \frac{100}{300}\left(1-\frac{100}{300}\right) = 0.44$

1.2) classification error

$L(T_1) = \frac{200}{400}\left(L(T_{1,L})\right) + \frac{200}{400}\left(L(T_{1,R})\right) = 0.25$

$L(T_2) = \frac{100}{400}\left(L(T_{2,L})\right) + \frac{200}{400}\left(L(T_{2,R})\right) = 0.25$

equal

Gini

$T_1 \Rightarrow \frac{200}{400}(0.375) + \frac{200}{400}(0.375) = 0.375$ $\quad$ T_2 is better

$T_2 \Rightarrow \frac{100}{400}(0) + \frac{300}{400}(0.44) = 0.33$

Title. Date. Page.

Conditional Entropy

$T_1 \Rightarrow \frac{200}{400}(0.562) + \frac{200}{400}(0.562) = 0.562$

$T_2 \Rightarrow \frac{100}{400}(0) + \frac{300}{400}(0.637) = 0.478$

T_2 is better

2.1) $f = \epsilon_t(e^{\beta_t} - e^{-\beta_t}) + e^{-\beta_t}$

$$\nabla_{\beta_t} f = \epsilon_t \frac{d}{d\beta_t}\left(e^{\beta_t} - e^{-\beta_t}\right) + \frac{d}{d\beta_t}\left(e^{-\beta_t}\right)$$

$$= \epsilon_t\left(e^{\beta_t} + e^{-\beta_t}\right) - e^{-\beta_t}$$

$$= \epsilon_t e^{\beta_t} + \epsilon_t e^{-\beta_t} - e^{-\beta_t}$$

$$= \epsilon_t e^{\beta_t} + \frac{\epsilon_t - 1}{e^{\beta_t}} = \frac{\epsilon_t e^{2\beta_t}}{e^{\beta_t}} + \frac{\epsilon_t - 1}{e^{\beta_t}} = 0$$

$$\epsilon_t e^{2\beta_t} + \epsilon_t - 1 = 0$$

$$e^{2\beta_t} = \frac{1 - \epsilon_t}{\epsilon_t}$$

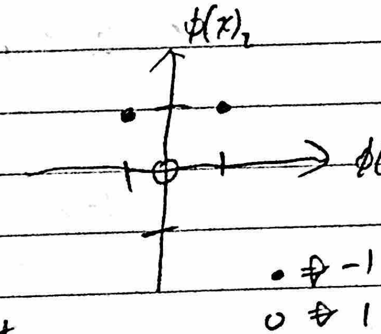
$$2\beta_t = \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

$$\beta_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

Title. Date. Page.

3.1) No because there is no 1D hyperplane (dot) that perfectly separates the points

3.2) $(x_1, y_1) = (-1,-1) \Rightarrow (-1,1,-1)$
$(x_2, y_2) = (1,-1) \Rightarrow (1,1,-1)$
$(x_3, y_3) = (0,1) \Rightarrow (0,0,1)$



yes because there is a line that perfectly separates the points

3.3) $k(x,x') = \phi(x)^T \phi(x') = xx' + x^2 x'^2$

$$\underline{K} = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3) \\ k(x_2,x_1) & k(x_2,x_2) & k(x_2,x_3) \\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3) \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

3.4) $\max_{\{\alpha_n\}} \sum_n \alpha_n - \frac{1}{2} \sum_{m,n} y_m y_n \alpha_m \alpha_n k(x_m, x_n)$ s.t. $\sum_n \alpha_n y_n = 0,\ 0 \le \alpha_n, \forall n$

$= \max_{\{\alpha_n\}} [\alpha_1 + \alpha_2 + \alpha_3] - \frac{1}{2}[2\alpha_1^2 + 2\alpha_2^2]$ st $-\alpha_1 - \alpha_2 + \alpha_3 = 0$,
$0 \le \alpha_n, \forall n$

Title. Date. Page.

3.5) $\alpha_3 = \alpha_1 + \alpha_2 \Rightarrow \max_{\{\alpha_n\}} [\alpha_1 + \alpha_2 + \alpha_1 + \alpha_2] - [\alpha_1^2 + \alpha_2^2]$

$= \max_{\{\alpha_n\}} 2\alpha_1 - \alpha_1^2 + 2\alpha_2 - \alpha_2^2$

$\nabla_{\alpha_1} = 2 - 2\alpha_1 = 0 \qquad \alpha_1 = 1 \Rightarrow \alpha_3 = 2$

$\nabla_{\alpha_2} = 2 - 2\alpha_2 = 0 \qquad \alpha_2 = 1$

$\vec{w}_{opt} = \sum_n \alpha_n^* y_n \vec{\phi}(x_n)$

$= (1)(-1)[-1 \ \ 1]^T + (1)(-1)[1 \ \ 1]^T + (2)(1)[0 \ \ 0]^T$

$= [1 \ -1]^T + [-1 \ -1]^T = [0 \ -2]^T$

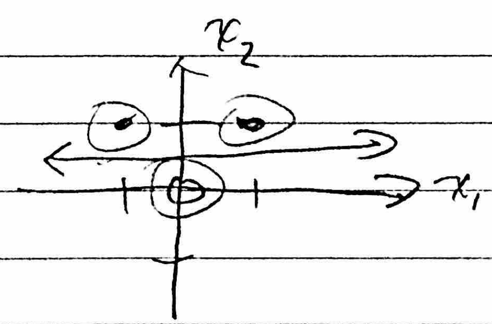
$b^* = y_n - \vec{w}_{opt}^T \vec{\phi}(x_n) = -1 - [0 \ -2]^T[-1 \ \ 1]$ for $n=1$

$= 1$

3.6) $0 = [0 \ -2]^T[x_1 \ \ x_2] + 1$

$0 = -2x_2 + 1$

$x_2 = \frac{1}{2}$



$\alpha_1, \alpha_2, \alpha_3 > 0$