Perceptron

1) $x^1 = \begin{pmatrix}1\\1\end{pmatrix},\ x^2=\begin{pmatrix}2\\2\end{pmatrix},\ x^3 = \begin{pmatrix}0\\1\end{pmatrix},\ x^4=\begin{pmatrix}0\\2\end{pmatrix}$

$\leftarrow \text{sign}(x) = \begin{cases} +1, & x>0 \\ -1, & x<0\end{cases}$

target = $t^1=1 \quad t^2=1,\ t^3=-1,\ t^4=-1$

w_0 = bias = 1

$$W = \begin{array}{c} \begin{pmatrix}1\\1\\1\end{pmatrix} \end{array}\begin{array}{l} w_0\\ w_1\\ w_2\end{array} \quad x^1=\begin{pmatrix}1\\1\\1\end{pmatrix},\ x^2=\begin{pmatrix}1\\2\\2\end{pmatrix},\ x^3=\begin{pmatrix}1\\0\\1\end{pmatrix},\ x^4=\begin{pmatrix}1\\0\\2\end{pmatrix}$$

$O = \text{sign}(w\cdot x)$

$W_{new} = W_{old} + \eta(t-O)\cdot x$

$\eta = 1$

↳ learning rate

a)

1st epoch

$O_1 = \text{sign}((1,1,1)\cdot(1,1,1)) = \text{sign}(3) = +1 = t_1$ ✓

$O_2 = \text{sign}((1,1,1)\cdot(1,2,2)) = \text{sign}(5) = +1 = t_2$ ✓

$O_3 = \text{sign}((1,1,1)\cdot(1,0,1)) = \text{sign}(2) = +1 \neq t_3 \Rightarrow$ update w

$w = (1,1,1) + (-1-(1))(1,0,1) = (1,1,1) + (-2,0,-2) = (-1,1,-1)$

$O_4 = \text{sign}((-1,1,-1)(1,0,2)) = \text{sign}(-1+0-2) = \text{sign}(-3) = -1$ ✓

2nd epoch

$O_1 = \text{sign}((-1,1,-1)\cdot(1,1,1)) = \text{sign}(-1+1-1) = \text{sign}(-1) = -1$ ✗ $\Rightarrow$ u

$w = (-1,1,-1) + (1-(-1))\cdot(1,1,1) = (-1,1,-1) + (2,2,2) = (1,3,1)$

$O_2 = \text{sign}(\overbrace{(1,3,1)\cdot(1,2,2)}^{>0}) = 1$ ✓

$O_3 = \text{sign}((1,3,1)\cdot(1,0,1)) = \text{sign}(2) = +1$ ✗ $\Rightarrow$ update w

$w = (1,3,1) + (-1-(1))(1,0,1) = (1,3,1) + (-2,0,-2) = (-1,3,-1)$

$O_4 = \text{sign}((-1,3,-1)\cdot(1,0,2)) = \text{sign}(-1+0+(-2)) = \text{sign}(-3)$ ✓

3rd epoch

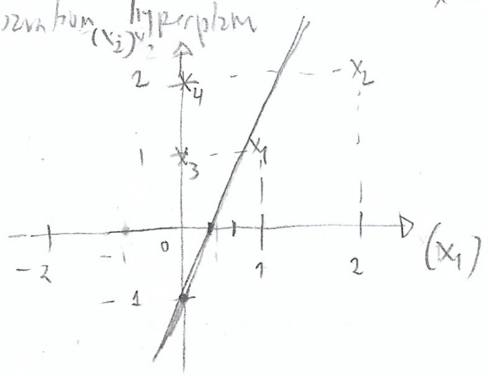
$O_1 = \text{sign}((-1,3,-1)\cdot(1,1,1)) = \text{sign}(-1+3-1) = \text{sign}(1) = 1$ ✓

$O_2 = \text{sign}((-1,3,-1)\cdot(1,2,2)) = \text{sign}(-1+6-2) = \text{sign}(3) = 1$ ✓

$O_3 = \text{sign}((-1,3,-1)\cdot(1,0,1)) = \text{sign}(-1+0-1) = \text{sign}(-2) = -1$ ✓

$O_4 = \text{sign}((-1,3,-1)\cdot(1,0,2)) = \text{sign}(-3) = -1$ ✓ $\Rightarrow$ convergency

b) separation hyperplane



$x^1 = \begin{pmatrix}1\\1\end{pmatrix},\ x^2=\begin{pmatrix}2\\2\end{pmatrix}\ x^3=\begin{pmatrix}0\\1\end{pmatrix}\ x^4=\begin{pmatrix}0\\2\end{pmatrix}$ $\quad W = (\underset{w_0}{-1}, \underset{w_1}{3}, \underset{w_2}{-1})$

$w_0 + w_1x_1 + w_2x_2 = 0$

$\Leftrightarrow w_2x_2 = -w_1x_1 - w_0$

$\Leftrightarrow x_2 = \dfrac{-w_1x_1 - w_0}{w_2}$

$\Leftrightarrow x_2 = \dfrac{-(3)x_1 - (-1)}{-1} = \dfrac{-3x_1+1}{-1} = 3x_1 - 1$

$x_2 = 3x_1 - 1$

$x_1 = 0 \Rightarrow x_2 = -1$

$x_2 = 0 \quad 0 = 3x_1 - 1 \Leftrightarrow 1 = 3x_1 \Leftrightarrow x_1 = \frac{1}{3}$

II

a) $E(D) = \sum -p_i \log_2 p_i = (-1/5 \log_2 1/5) \times 3 - 2/5 \log_2 2/5$
$= -3/5 \times (-2.32) - 2/5(-1.32) = 1.39 + 0.53 = 1,92$

$E(F_1) = 3/5 \, \overset{A}{E(F_1|a)} + \underline{2/5 \, \overset{C}{E(F_1|c)}}$

$E(F_1|a) = E(1/3, 2/3) = - 1/3 \log_2 1/3 - (\ldots) = 0,92$

$E(F_1|c) = E(1/2, 1/2) = 1$

proportion of the attribute values

Constrained by the attr value, Entropy of the resulting dataset

$E(F_1) = 3/5 \times 0,92 + 2/5 \times 1$
$IG(F_1) = E(D) - E(F_1)$
$= 1,92 - 0,952 = 0,968$

$E(F_2) = 2/5 \, \underset{A}{E(1/2, 1/2)} + 3/5 \, \underset{B}{E(1/3, 1/3, 1/3)}$

$IG(F_2) = 0,57$

$E(F_3) = 3/5 \, \underset{B}{E(1/3, 1/3, 1/3)} + 2/5 \, \underset{C}{E(1/2, 1/2)}$

$IG(F_3) = 0,57$

$E(F_4) = 3/5 \, \underset{A}{E(2/3, 1/3)} + 2/5 \, \underset{X}{E(1/2, 1/2)}$

$IG(F_4) = 0,97$

A: Either F_1 or F_4 can be choosen as root.

b) Picking F_1 as root

$E(D_{F_1/a}) = 0,92$

$E(F_2) = 1/3 \, \overset{A}{\cancel{E(1)}} + 2/3 \, \overset{B}{E(1/2, 1/2)}$

$E(F_3) = 1/3 \, \overset{C}{\cancel{E(1)}} + 2/3 \, \overset{B}{E(1/2, 1/2)}$

$E(F_4) = 1 \, \overset{A}{E(1/3, 2/3)}$

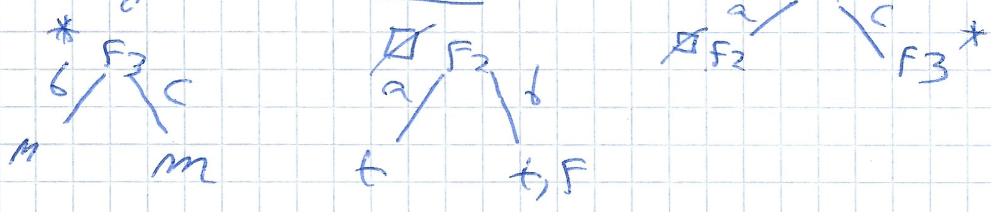
choose either F_2, F_3

$E(D_{F_1/c}) = 1$

$E(F_2) = 1/2 \, E(1) + 1/2 \, E(1) = 0$
$E(F_3) = 0$
$E(F_4) = E(1/2, 1/2) = 1$

choose either F_2, F_3

Looking at the table



III)

Perceptron can be used if relationship between defect output and (length, width) is linear.
Decision tree can be used also if attributes can be discretised into binary splits by an algorithm other than ID3.

3 IV

a) $P(C=A) = 1/2 \quad P(C=B) = 1/2$

$p(X_1 | C=A)$

$\mu = 10 \quad \sigma^2 = [(0-10)^2 + (0-10)^2 + (20-10)^2 + (20-10)^2]/3 = 133.3$
$\sigma = 11.55$

$p(X_2 | C=A)$
$\sigma^2 = [(10-15)^2 + (20-15)^2 + (10-15)^2 + (20-15)^2]/3 = 33.3$
$\mu = 15 \quad \sigma = 5{,}77$

$p(X_1 | C=B)$ $\qquad$ $p(X_2 | C=B)$

$\mu = 40 \quad \sigma^2 = 133.3$ $\qquad$ $\mu = 35 \quad \sigma^2 = 33.3$
$\sigma = 11.55$ $\qquad$ $\sigma = 5{,}77$

$$P(C=A | X_1=10, X_2=10) = \frac{P(C=A)\, P(X_1=10|C=A)\, P(X_2=10|C=A)}{P(X_1=10, X_2=10)}$$
[...]

Priors and denominator are the same so we just need to know which likelihood is bigger. (product of the conditionals)

$N(10 | \mu=10, \sigma=11.55) \qquad N \sim \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$= \frac{1}{28.95} \cancel{e^{-1/2 \times 0}} = 0{,}035$

$N(10 | \mu=15, \sigma=5{,}77) = \frac{1}{14.46} e^{-1/2 \times (0{,}86)^2} = \frac{0.686}{14.46} = 0{,}047$

$\rightarrow P(C=A | \ldots) = 0{,}035 \times 0{,}047 = 1{,}67 \times 10^{-3}$

Now, get PDF for x_1, x_2 in class B

$N(10 | \mu=40, \sigma=11.55)$
$= 0{,}001$

$N(10 | \mu=35, \sigma=5{,}77) = 5{,}86 \times 10^{-6}$

$\longrightarrow P(C=B | \ldots) = 0{,}001 \times 5{,}86 \times 10^{-6}$

Most likely class will be A.

b) $\vec{\mu}_{C=A} = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$ $\vec{\mu}_{C=B} = \begin{bmatrix} 40 \\ 35 \end{bmatrix}$

$$\mathrm{cov}(x_1, x_2)/_{C=A} = [(0-10)(10-15) + (0-10)(20-15) + (20-10)(10-15) + (20-10)(20-15)]/3$$

$$= [50 - 50 - 50 + 50]/3 = \cancel{33.3} \ \phi$$

$$\mathrm{cov}(x_1, x_2)/_{C=B} = [\dots] = \phi$$

COV MATRIX $\Sigma_{C=A} = \begin{bmatrix} \mathrm{cov}(x_1, x_1) & \mathrm{cov}(x_1, x_2) \\ \mathrm{cov}(x_2, x_1) & \underbrace{\mathrm{cov}(x_2, x_2)}_{= \sigma^2_{x_2}} \end{bmatrix} = \begin{bmatrix} 133.33 & \phi \\ \phi & 33.33 \end{bmatrix}$

Variances are equal for vars x_1, x_2 constrained by any of the two $C = A, B$, so: $\Sigma_{C=A} = \Sigma_{C=B} = \Sigma$

$$\det \Sigma = ad - bc = 133.33 \times 33.33 = 4443,9$$

$$\Sigma^{-1} = \frac{1}{4443,9} \times \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} 2,25 \times 10^{-4} \times 33.33 & 0 \\ 0 & 2,25 \times 10^{-4} \times 133.33 \end{bmatrix}$$

$$= \begin{bmatrix} 7,5 \times 10^{-3} & 0 \\ 0 & 0,029 \end{bmatrix}$$

Again, prior probabilities for A, B are the same, and we also ignore the denominator, so we're interested in calculating only the distribution.

$$N([10,10], C=A) = \frac{e^{-1/2 M}}{2\pi \sqrt{\det \Sigma}}$$

$$M = [0 \ -5] \begin{bmatrix} 7,5 \times 10^{-3} & 0 \\ 0 & 0,029 \end{bmatrix} \begin{bmatrix} 0 \\ -5 \end{bmatrix} = [0 \ -0,145] \begin{bmatrix} 0 \\ -5 \end{bmatrix} = 0,725$$

$$N_{C=A} = \frac{e^{-0,363}}{418.85} = 1,66 \times 10^{-3}$$

$$N([10,10], C=B) = [\dots] = 6,92 \times 10^{-9}$$

sample belongs to class A