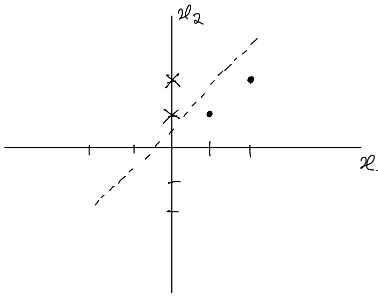
Machine Learning   Homework 1

Daniel Castro - 87644
João Tiago Aparício - 97155

I

$x_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}$ $x_2 = \begin{bmatrix}1\\2\\2\end{bmatrix}$ $x_3 = \begin{bmatrix}1\\0\\1\end{bmatrix}$ $x_4 = \begin{bmatrix}1\\0\\2\end{bmatrix}$

$t_1 = 1$ $t_2 = 1$ $t_3 = -1$ $t_4 = -1$



$w \cdot x = w_0 x_0 + w_1 x_1 + w_2 x_2 \geq 0$
$= -1 + 3x_1 - x_2 \geq 0$
$\Downarrow$
$x_2 \leq 3x_1 - 1$

a) $w(1,1,1)$ $\eta = 1$

$\left.\begin{matrix}1\\-1\end{matrix}\right\} \; o = \text{sgn}(w \cdot x) = \text{sgn}(w_0 x_0 + w_1 x_1 + w_2 x_2)$

1^a reta $\begin{cases} w \cdot x = w_0 x_0 + w_1 x_1 + w_2 x_2 \geq 0 \\ = 1 \cdot 1 + 1 \cdot x_1 + 1 \cdot x_2 \geq 0 \\ = x_2 \geq -x_1 - 1 \quad (1) \end{cases}$

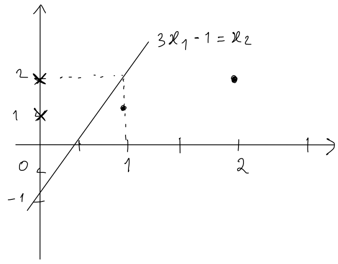
$w_1 = \eta \cdot (t_1 - o_1) \cdot x_1 = 1 \cdot (1 - \text{sgn}(w \cdot x_1)) x_1 =$
$= [\underbrace{1 - \text{sgn}((1,1,1)\cdot(1,1,1))}_{0}] \cdot x_1 = (0,0,0)$
$w = (1,1,1) + (0,0,0) = (1,1,1)$
$w_2 = \eta \cdot (t_2 - o_2) \cdot x_2 = 1 \cdot (1 - \text{sgn}((1,1,1)\cdot(1,2,2))) x_2 =$
$= (0,0,0)$
$w = (1,1,1) + (0,0,0) = (1,1,1)$
$w_3 = \eta \cdot (t_3 - o_3) \cdot x_3 = 1 \cdot (-1 - \text{sgn}(1,1,1)\cdot(1,0,1))) x_3 =$
$= (-2,0,-2)$
$w = (1,1,1) + (-2,0,-2) = (-1,1,-1)$
$w_4 = \eta \cdot (t_4 - o_4) \cdot x_4 = 1 \cdot \underbrace{(-1+1)}_{0} \cdot (1,0,2) = (0,0,0)$
$w = (-1,1,-1) + (0,0,0) = (-1,1,-1)$

$w_1 = \eta \cdot (t_1 - o_1) \cdot x_1 = 1 \cdot (1 - \text{sgn}(w \cdot x_1)) x_1 =$
$= [1 - \text{sgn}((-1,1,-1)\cdot(1,1,1))] \cdot x_1 = (2,2,2)$
$w = (-1,1,-1) + (2,2,2) = (1,3,1)$
$w_2 = \eta \cdot (t_2 - o_2) \cdot x_2 = 1 \cdot (1 - \text{sgn}((1,3,1)\cdot(1,2,2))) x_2 =$
$= (0,0,0)$
$w = (1,3,1) + (0,0,0) = (1,3,1)$
$w_3 = \eta \cdot (t_3 - o_3) \cdot x_3 = 1 \cdot (-1 - \text{sgn}(1,3,1)\cdot(1,0,1))) x_3 =$
$= (-2,0,-2)$
$w = (1,3,1) + (-2,0,-2) = (-1,3,-1)$
$w_4 = \eta \cdot (t_4 - o_4) \cdot x_4 = 1 \cdot (-1 - \text{sgn}((-1,3,-1)) \cdot (1,0,2) = (0,0,0)$
$w = (-1,3,-1) + (0,0,0) = (-1,3,-1)$

$w_1 = \eta \cdot (t_1 - o_1) \cdot x_1 = 1 \cdot (1 - \text{sgn}(w \cdot x_1)) x_1 =$
$= [1 - \text{sgn}((-1,1,-1)\cdot(1, \ ,1))] \cdot x_1 = (0,0,0)$
$w = (-1,3,-1) + (0,0,0) = (-1,3,-1)$
$w_2 = \eta \cdot (t_2 - o_2) \cdot x_2 = 1 \cdot (1 - \text{sgn}((-1,3,-1)\cdot(1,2,2))) x_2 =$
$= (0,0,0)$
$w = (-1,3,-1) + (0,0,0) = (-1,3,-1)$
$w_3 = \eta \cdot (t_3 - o_3) \cdot x_3 = 1 \cdot (-1 - \text{sgn}(-1,3,-1)\cdot(1,0,1))) x_3 =$
$= (0,0,0)$
$w = (-1,3,-1) + (0,0,0) = (-1,3,-1)$
$w_4 = \eta \cdot (t_4 - o_4) \cdot x_4 = 1 \cdot (-1 - \text{sgn}((-1,3,-1)) \cdot (1,0,2) = (0,0,0)$
$w = (-1,3,-1) + (0,0,0) = (-1,3,-1)$

In this iteration the values of the weights converged.

b) $x_2 \leq 3x_1 - 1$



II

F_1	F_2	F_3	F_4	S Out
a, c	a, b	b, c	a, x	f, m, m, t

F_1	F_2	F_3	F_4	O
c	a	b	x	m
a	a	c	a	t
a	b	b	a	t
c	b	c	x	m
a	b	b	a	f

a) $\text{Gain}(S,A) = G(S,A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{\#S_v}{\#S} \cdot \text{Entropy}(S_v)$

$\text{Entropy}(S) = \sum_{i=1} -p_i \log_2 p_i$

$\text{Gain}(S, F_1)$

$E(S) = 3\left(-\frac{1}{5}\log_2\frac{1}{5}\right) - \frac{2}{5}\log_2\frac{2}{5} = 1{,}922$

$E(F_{1a}) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0{,}918$

$E(F_{1c}) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$

$G(S,F_1) = 1{,}922 - \frac{3}{5}\cdot 0{,}918 - \frac{2}{5}\cdot 1 = 0{,}971$

$E(F_{2a}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$

$E(F_{2b}) = -\frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} = 3\cdot\left(\frac{1}{3}\log\frac{1}{3}\right) = 1{,}585$

$G(S,F_2) = 1{,}922 - \frac{2}{5}\cdot 1 - \frac{3}{5}\cdot 1{,}585 = 0{,}571$

$E(F_{3b}) = 3\left(-\frac{1}{3}\log\frac{1}{3}\right) = 1{,}585$

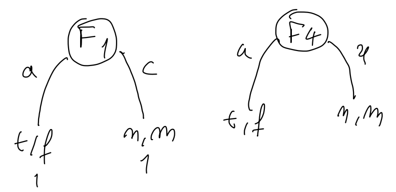
$E(F_{3c}) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1$

$G(S,F_3) = 0{,}571$

$E(F_{4a}) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0{,}918$

$E(F_{4x}) = 1$

$G(S,F_4) = 1{,}922 - \frac{2}{5}\cdot 0{,}918 = 0{,}971$



b)

F_2	F_3	F_4	out
a	c	a	t
b	b	a	t
b	b	a	f

$E(S_a) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0{,}918$

$S_{F_{1a}}$

$E(F_{2a}) = 1\log 1 = 0$

$E(F_{2b}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$

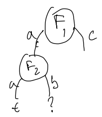
$G(S_a, F_2) = 0{,}918 - \frac{1}{3}\cdot 0 - \frac{2}{3}\cdot 1 = 0{,}251$

$E(F_{3c}) = 0$; $E(F_{3b}) = 1$

$G(S_a, F_a) =$

$E(S_{F_{4c}}) = -\frac{1}{3}\log\frac{1}{3}\cdot\frac{2}{3}\log\frac{2}{3} = 0{,}918$

$G(S_a, F_4) = 0{,}918 - 0{,}918 = 0$



$E(S_{F_{1c}}) = -\frac{1}{2}\log\frac{1}{2} \cdot \frac{1}{2}\log\frac{1}{2} = 1$

$E(F_{2a}) = 6$

$E(F_{2b}) = 0$

$G(S_c, F_2) = 1 - \frac{1}{2}\cdot 0 - \frac{1}{2}\cdot 0 = 1$

F_2	F_3	F_4	out
a	b	x	m
b	c	x	m

$E(F_{3b}) = 0$

$E(F_{3c}) = 0$

$G(S_c, F_3 = 1$

$E(F_4, x) = 1$

$G(S_c, F_4) = 0$

F_3	F_4	Out
b	a	t
b	a	f

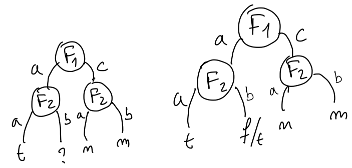
$E(S_{F_{4c}} F_{2b}) = 1$

$E(F_{3b}) = 1$

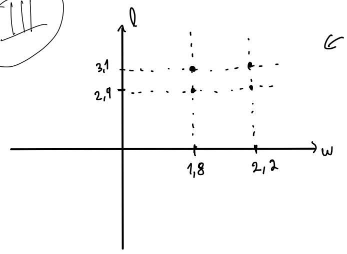
$G(S, F_3 b) = 0$

$E(F_4, a) = 1$

$G(S, F_{4a}) = 0$

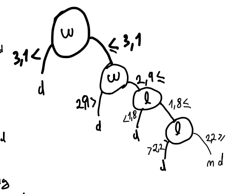


III



l	n	output
		defective d
		non-defective nd

the data given cannot be divided by a linear expression such as the one produced by a Perceptron But can be separated using a decision tree like such:



IV

$P(x|\mu,\sigma^2) = N(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\cdot\pi}\cdot\sigma} \cdot \exp\left(-\frac{1}{2\sigma^2}\cdot(x-\mu)^2\right)$

a)

x_1	x_2	Class
0	10	A
0	20	A
20	10	A
20	20	A
30	30	B
30	40	B
50	30	B
50	40	B

$x = \begin{bmatrix}10\\10\end{bmatrix}$ $P(C=A) = \frac{1}{2}$ $P(C=B) = \frac{1}{2}$

$\mu_{x_1 \, C=A} = \frac{0\cdot 2 + 20\cdot 2}{4} = 10$

$\mu_{x_2 \, C=A} = \frac{10\cdot 2 + 20\cdot 2}{4} = 15$

$\mu_{x_1 \, C=B} = \frac{30\cdot 2 + 50\cdot 2}{4} = 40$

$\mu_{x_2 \, C=B} = \frac{30\cdot 2 + 40\cdot 2}{4} = 35$

C = A: $\mu = (10, 15)$ C = B: $\mu = (40, 35)$

$\sigma_{x_1 \, C=A} = \left(\frac{1}{4-1}\cdot\left(2\cdot(0-10)^2 + 2\cdot(20-10)^2\right)\right)^{\frac{1}{2}}$

$= 11.547 \Rightarrow \sigma^2_{x_1 \, C=2} = 133{,}33$

C = A: $\sigma^2_{x_1} = 133{,}33$, $\sigma^2_{x_2} = 33{,}33$

C = B: $\sigma^2_{x_1} = 133{,}33$, $\sigma^2_{x_2} = 33{,}33$

Probability of being class A

$P(C=A|x) = \frac{P(C=A)\,P(x|C=A)}{P(x)} = \frac{P(C=A)\,P(x_1=10|C=A)\,P(x_2=10|C=A)}{P(x)} =$

$= \frac{P(C=A)\,N(x_1|\mu_{x_1 \, C=A}, \sigma^2_{x_1 \, C=A})\,N(x_2|\mu_{x_2 \, C=A}, \sigma^2_{x_2 \, C=B})}{P(x)} =$

$= \frac{\frac{1}{2}\cdot 3{,}45498\cdot 10^{-2}\cdot 4{,}74915\cdot 10^{-2}}{P(x)} = \frac{8{,}20411\cdot 10^{-4}}{P(x)}$

Probability of being class B

$P(C=B|x) = \frac{P(C=B)\,P(x|C=B)}{P(x)} = \frac{P(C=B)\,P(x_1=10|C=B)\,P(x_2=10|C=B)}{P(x)} =$

$= \frac{P(C=B)\,N(x_1|\mu_{x_1 \, C=B}, \sigma^2_{x_1 \, C=B})\,N(x_2|\mu_{x_2 \, C=B}, \sigma^2_{x_2 \, C=B})}{P(x)} =$

$= \frac{\frac{1}{2}\cdot 1{,}18213\cdot 10^{-3}\cdot 5{,}85564\cdot 10^{-6}}{P(x)} = \frac{3{,}464107\cdot 10^{-9}}{P(x)}$

Given that $P(C=A|X) > P(C=B|X)$ the classification for $x = \begin{bmatrix}10\\10\end{bmatrix}$ would be class = A.

C.A.

C = A

$N(x_1|\mu_{x_1 \, C=A}, \sigma^2_{x_1 \, C=A}) =$
$= \frac{1}{\sqrt{2\cdot\pi}\cdot\sqrt{133{,}33}}\cdot\exp\left(-\frac{1}{2\cdot 133{,}33}\cdot(10-10)^2\right)$
$= 3{,}45498\cdot 10^{-2}$

$N(x_2|\mu_{x_2 \, C=A}, \sigma^2_{x_2 \, C=A}) =$
$= \frac{1}{\sqrt{2\cdot\pi}\cdot\sqrt{33{,}33}}\cdot\exp\left(-\frac{1}{2\cdot 33{,}33}\cdot(10-15)^2\right)$
$= 4{,}74915\cdot 10^{-2}$

C = B

$N(x_1|\mu_{x_1 \, C=A}, \sigma^2_{x_1 \, C=A}) =$
$= \frac{1}{\sqrt{2\cdot\pi}\cdot\sqrt{133{,}33}}\cdot\exp\left(-\frac{1}{2\cdot 133{,}33}\cdot(10-40)^2\right)$
$= 1{,}18213\cdot 10^{-3}$

$N(x_2|\mu_{x_2 \, C=A}, \sigma^2_{x_2 \, C=A}) =$
$= \frac{1}{\sqrt{2\cdot\pi}\cdot\sqrt{33{,}33}}\cdot\exp\left(-\frac{1}{2\cdot 33{,}33}\cdot(10-35)^2\right)$
$= 5{,}85564\cdot 10^{-6}$

$x = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$

b) $P(x \mid \mu, \Sigma) = \mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2 \cdot \pi)^{D/2}} \cdot \frac{1}{|\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot (x-\mu)^T \Sigma^{-1} \cdot (x-\mu)\right)$

	$P(X_1, X_2 \mid C=A)$	$P(X_1, X_2 \mid C=B)$
μ	$\begin{bmatrix} 10 \\ 15 \end{bmatrix}$	$\begin{bmatrix} 40 \\ 35 \end{bmatrix}$
Σ	$\begin{bmatrix} 133,33 & 0 \\ 0 & 33,33 \end{bmatrix}$	$\begin{bmatrix} 133,33 & 0 \\ 0 & 33,33 \end{bmatrix}$

$$P(C=A \mid X_1=10, X_2=10) = \frac{P(C=A)\, P(X_1=10, X_2=10 \mid C=A)}{P(X_1=10, X_2=10)} =$$

$$= \frac{\frac{1}{2} \cdot \mathcal{N}\left(\begin{bmatrix} 10 \\ 10 \end{bmatrix} \mid \mu = \begin{bmatrix} 10 \\ 15 \end{bmatrix}, \Sigma_{C=A} = \begin{bmatrix} 133,33 & 0 \\ 0 & 33,33 \end{bmatrix}\right)}{P(x)} =$$

$$= \frac{\frac{1}{2} \cdot \frac{1}{(2 \cdot \pi)^{2/2}} \cdot \frac{1}{(4443,8889)^{1/2}} \cdot e^{\overbrace{-\frac{1}{2} \cdot \left(\begin{bmatrix} 10 \\ 10 \end{bmatrix} - \begin{bmatrix} 10 \\ 15 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 0,0007500187 & 0 \\ 0 & 0,03000300 \end{bmatrix} \cdot \left(\begin{bmatrix} 10 \\ 10 \end{bmatrix} - \begin{bmatrix} 10 \\ 15 \end{bmatrix}\right)}^{= -3,75038 \cdot 10^{-1}}}}{P(x)} =$$

$$= \frac{1,6408224948 \cdot 10^{-3}}{2 P(x)} = \frac{8,20411 \cdot 10^{-4}}{P(x)}$$

$$P(C=B \mid X_1=10, X_2=10) = \frac{P(C=B)\, P(X_1=10, X_2=10 \mid C=B)}{P(X_1=10, X_2=10)} =$$

$$= \frac{\frac{1}{2} \cdot \mathcal{N}\left(\begin{bmatrix} 10 \\ 10 \end{bmatrix} \mid \mu = \begin{bmatrix} 40 \\ 35 \end{bmatrix}, \Sigma_{C=A} = \begin{bmatrix} 133,33 & 0 \\ 0 & 33,33 \end{bmatrix}\right)}{P(x)} =$$

$$= \frac{\frac{1}{2} \cdot \frac{1}{(2 \cdot \pi)^{2/2}} \cdot \frac{1}{(4443,8889)^{1/2}} \cdot e^{\overbrace{-\frac{1}{2} \cdot \left(\begin{bmatrix} 10 \\ 10 \end{bmatrix} - \begin{bmatrix} 40 \\ 35 \end{bmatrix}\right)^T \cdot \begin{bmatrix} 0,0007500187 & 0 \\ 0 & 0,03000300 \end{bmatrix} \cdot \left(\begin{bmatrix} 10 \\ 10 \end{bmatrix} - \begin{bmatrix} 40 \\ 35 \end{bmatrix}\right)}^{= -12,75102197}}}{P(x)}$$

$$= \frac{6.9221348 \cdot 10^{-9}}{2 \cdot P(x)} = \frac{3,6106741827 \cdot 10^{-9}}{P(x)}$$

$P(C=A \mid X) > P(C=B \mid X)$ logo X seria classificado com $C=A$

CA

$C=A$

$\Sigma_{00} = \frac{1}{4-1} \cdot (2 \cdot (0-10)^2 + 2 \cdot (20-10)^2) = 133,33$

$\Sigma_{11} = 33,33$

$\Sigma_{01} = \frac{1}{4-1}((0-10)(10-15)+(0-10)(20-15)+(20-10)(10-15)+(20-10)(20-15)) = 0$

$\Sigma_{10} = \frac{1}{4-1}((0-10)(10-15)+(0-10)(20-15)+(20-10)(10-15)+(20-10)(20-15)) = 0$

$C=B$

$\Sigma_{00} = 133,33$

$\Sigma_{11} = 33,33$

$\Sigma_{01} = cov(x_1, x_2) = 0$

$\Sigma_{10} = cov(x_1, x_2) = 0$

$$\begin{bmatrix} 133,33 & 0 \\ 0 & 33,33 \end{bmatrix}^{-1} = \begin{bmatrix} 7,50019 \cdot 10^{-3} & 0 \\ 0 & 3.0003 \cdot 10^{-2} \end{bmatrix}$$

$\det(\Sigma_{C=A}) = 4443,8889$