Exame ML 2017/2018

1.

a) False. Since the mixture of p+1 Gaussians requires more parameters to be estimated, then optimal value should be higher, when comparing with p Gaussians.

b) True. Naive Bayes is a special case of the Bayes classifier that assumes that x_i and x_j are conditionally independent given y, for all $i \neq j$: $P(X|Y,Z) = P(X|Z)$.

2. $\boxed{\text{See Springer PDF}}$

a) Primeiro lugar, temos de mapear para o novo feature space: $A_i(p)$ is 1 if $|x_i| > 1$; ~~0~~ otherwise!

Ponto	A_1	A_2	Classe
(-2.4, 0.4)	1	0	—
(-2.1, -0.3)	1	0	—
(-1.7, -1.6)	1	1	—
(-1.6, -1.3)	1	1	—
(-1.5, 1.5)	1	1	—
(-1.2, 1.9)	1	1	—
(-1.1, -2.0)	1	1	—
(-0.5, 0.1)	0	0	+
(0.0, 0.4)	0	0	+
(0.0, 2.0)	0	1	—

Ponto	A_1	A_2	Classe
(0.1, -0.7)	0	0	+
(0.1, -0.6)	0	0	+
(0.1, 0.0)	0	0	+
(0.2, -0.5)	0	0	+
(0.3, -0.5)	0	0	+
(0.4, 0.9)	0	0	+
(0.8, 0.2)	0	0	+
(1.0, 0.1)	0	0	+
(1.7, -1.0)	1	1	−
(2.0, 0.4)	1	0	−

Agora temos de calcular as a priori probabilities:

$$P(C=+) = \frac{10}{20} = \frac{1}{2} \qquad P(C=-) = \frac{10}{20} = \frac{1}{2}$$

$$P(A_1=1 \mid C=+) = \frac{0}{10} = 0 \qquad P(A_2=0 \mid C=+) = \frac{10}{10} = 1$$

$$P(A_1=0 \mid C=+) = \frac{10}{10} = 1 \qquad P(A_2=1 \mid C=+) = \frac{0}{10} = 0$$

$$P(A_2=0 \mid C=-) = \frac{3}{10}$$

$$P(A_1=1 \mid C=-) = \frac{9}{10}$$

$$P(A_2=1 \mid C=-) = \frac{7}{10}$$

$$P(A_1=0 \mid C=-) = \frac{1}{10}$$

The test point is : (0.9, 0.9) → First map A_1 e A_2

So the point is : $A_1 = 0$

$A_2 = 0$

To compute the class label:

For $C = +$:

$$P(C=+) \prod_{j=1}^{2} P(A_j = a_j \mid C=+) = P(C=+) \times P(A_1=0 \mid C=+) \times P(A_2=0 \mid C=+) = \frac{1}{2} \times 1 \times 1 = \frac{1}{2}$$

For $C = -$:

$$P(C=-) \prod_{j=1}^{2} P(A_j = a_j \mid C=-) = P(C=-) \times P(A_1=0 \mid C=-) \times P(A_2=0 \mid C=-) =$$

$$\frac{1}{2} \times \frac{1}{10} \times \frac{7}{10} = \frac{7}{200}$$

Como $\frac{1}{2} > \frac{7}{200}$, o Ponto (0.9, 0.9) seria atribuído à classe $+$.

See lecture 05.pdf

5) A perceptron is part of a family of learning machines, but the standard type consists of a layer of fixed non-linear basis functions followed by a simple linear discriminant function. Generally the perceptron works with values in the range $\{-1, 1\}$, so, a few pre-processing steps would need to be done:

a) Attributes A_1 and A_2 would need to be in range $\{-1, 1\}$.

so
$$A_i(p) = \begin{cases} x_i, & -1 \leq x_i \leq 1 \\ 1, & |x_i| \geq 1 \wedge x_i > 0 \\ -1, & |x_i| \geq 1 \wedge x_i < 0 \end{cases}$$

b) The class can be also put in range $\{-1, 1\}$, so let's say $c_- = -1$ and $c_+ = 1$, but this is not needed since:

$$y = \begin{cases} 1 & \text{if } \sum_{i=0}^{n} w_i x_i > 0 \\ -1, & \text{otherwise} \end{cases}$$

Note:

$$\boxed{\text{Perceptron: } y(x, w) = \left(\sum_{j=1}^{M} w_j \phi(x) \right)}$$

3. See lecture 02!

a) We have a linear regression problem. Here, it would be useful to use a normal equation approach: $w = (X^T X)^{-1} X^T y$

$$\boxed{X^T X w = X^T y}$$

The model is: $y_1 = w_{11} x_1 + w_{21} x_2$

$$w = \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix}, \quad X = \begin{bmatrix} -1.5 & -2 \\ -1 & -0.5 \\ 0 & 0 \\ 1 & 1 \\ 1.5 & 2 \end{bmatrix}, \quad y_1 = \begin{bmatrix} -5.1 \\ -2.5 \\ 0.05 \\ 3.2 \\ 4.9 \end{bmatrix}$$

Ento:

$$X^T X w = X^T y_1$$

$$\underbrace{\begin{bmatrix} -1.5 & -1 & 0 & 1 & 1.5 \\ -2 & -0.5 & 0 & 1 & 2 \end{bmatrix}}_{2\times 5} \underbrace{\begin{bmatrix} -1.5 & -2 \\ -1 & -0.5 \\ 0 & 0 \\ 1 & 1 \\ 1.5 & 2 \end{bmatrix}}_{5\times 2} \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix} = X^T y_1$$

$$= \underbrace{\begin{bmatrix} 6.5 & 7.5 \\ 7.5 & 9.25 \end{bmatrix}}_{2\times 2} \underbrace{\begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix}}_{2\times 1} = \underbrace{\begin{bmatrix} -1.5 & -1 & 0 & 1 & 1.5 \\ -2 & -0.5 & 0 & 1 & 2 \end{bmatrix}}_{2\times 5} \underbrace{\begin{bmatrix} -5.1 \\ -2.5 \\ 0.05 \\ 3.2 \\ 4.9 \end{bmatrix}}_{5\times 1}$$

$$= \begin{bmatrix} 6.5w_{11} + 7.5w_{21} \\ 7.5w_{11} + 9.25w_{21} \end{bmatrix} = \begin{bmatrix} 20.7 \\ 24.45 \end{bmatrix},$$

2×1

convertendo em sistema de equações:

$$\begin{cases} 6.5w_{11} + 7.5w_{21} = 20.7 \\ 7.5w_{11} + 9.25_{21} = 24.45 \end{cases} \Leftrightarrow \begin{cases} w_{11} = 2,0903 \\ w_{21} = 0,9483 \end{cases}$$

Convalidei com casio fx-CG

b) The wanted model is: $y_2 = w_{12} x_1 x_2 + w_{22} x_2$

$$W = \begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix}, \quad X = \begin{array}{cc} x_1x_2 & x_2 \\ \hline \end{array} \begin{bmatrix} 3 & -2 \\ 0.5 & -0.5 \\ 0 & 0 \\ 1 & 1 \\ 3 & 2 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 2.7 \\ -1.05 \\ 0.03 \\ 3.9 \\ 8.7 \end{bmatrix}$$

Com as normal equations

$$X^T X W = X^T y_2$$

$$\begin{bmatrix} 3 & 0.5 & 0 & 1 & 3 \\ -2 & -0.5 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 0.5 & -0.5 \\ 0 & 0 \\ 1 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix} = X^T y_2$$

$$(1)\ \underset{2\times 2}{\begin{bmatrix} 19.25 & 0.75 \\ 0.75 & 9.25 \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix}} = \underset{2\times 5}{\begin{bmatrix} 3 & 0.5 & 0 & 1 & 3 \\ -2 & -0.5 & 0 & 1 & 2 \end{bmatrix}} \underset{5\times 1}{\begin{bmatrix} 2.7 \\ -1.05 \\ 0.01 \\ 3.9 \\ 8.7 \end{bmatrix}}$$

$$\Leftrightarrow \begin{bmatrix} 19.25\,w_{12} + 0.75\,w_{22} \\ 0.75\,w_{12} + 9.25\,w_{22} \end{bmatrix} = \begin{bmatrix} 37.575 \\ 16.425 \end{bmatrix}$$

Sistema de Equações:

$$\begin{cases} 19.25\,w_{12} + 0.75\,w_{22} = 37.575 \\ 0.75\,w_{12} + 9.25\,w_{22} = 16.425 \end{cases} \Leftrightarrow \begin{cases} w_{12} = 1,8887 \\ w_{22} = 1.6225 \end{cases}$$

(see next page!)

c) doing everything:

$$W = \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}, \quad X = \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_1x_2 & x_2 \end{array} \\ \begin{bmatrix} -1.5 & -2 & 3 & -2 \\ -1 & -0.5 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1.5 & 2 & 3 & 2 \end{bmatrix} \\ 5\times 4 \end{array}, \quad Y = \begin{bmatrix} -5.1 & 2.7 \\ -2.5 & -1.05 \\ 0.05 & 0.01 \\ 3.2 & 3.9 \\ 4.9 & 8.7 \end{bmatrix}$$

Resolvendo:

(See next page!)

$$X^T X W = X^T Y$$

$$\underset{4\times 5}{\begin{bmatrix} -1.5 & -1 & 0 & 1 & 1.5 \\ -2 & -0.5 & 0 & 1 & 2 \\ 3 & 0.5 & 0 & 1 & 3 \\ -2 & -0.5 & 0 & 1 & 2 \end{bmatrix}} \underset{5\times 4}{\begin{bmatrix} -1.5 & -2 & 3 & -2 \\ -1 & -0.5 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1.5 & 2 & 3 & 2 \end{bmatrix}} \quad W = X^T Y$$

e) Para resolver esta, podemos aproveitar o sistema de equações anterior e resolver impondo as condições de a) e b)

$$\begin{cases} 6.5\,w_{11} + 7.5\,w_{21} = 20.7 \quad (1) \\ 7.5\,w_{11} + 9.25\,w_{21} = 24.45 \quad (2) \\ 19.25\,w_{12} + 0.75\,w_{22} = 37.575 \quad (3) \\ 0.75\,w_{12} + 9.25\,w_{22} = 16.425 \quad (4) \end{cases}$$

Se $w_{11} = w_{12}$, podemos somar duas equações para ficar com um sistema de 3 equações e três incógnitas:

Somando (2) e (3):

$$\begin{cases} 6.5\,w_{11} + 7.5\,w_{21} = 20.7 \\ 26.75\,w_{11} + 9.25\,w_{21} + 0.75\,w_{22} = 62.025 \\ 0.75\,w_{11} + 9.25\,w_{22} = 16.425 \end{cases}$$

Temos agora um S.E com três equações e três incógnitas:

Resolvendo o sistema (casio fx-CG20, máquina de calcular):

$$\begin{cases} w_{11} = 1.8831 \\ w_{21} = 1.279 \\ w_{22} = 1.6229 \end{cases}$$

, $w_{11} = w_{12}$, logo $w_{12} = 1.8831$

4. see a3-solution.pdf

Cases $K=5$, $\varepsilon=0.1$ | $K=20$, $\varepsilon=0.4$

$K=5$, $\varepsilon=0.4$ | $K=20$, $\varepsilon=0.1$

Error formula, based on K and ε:

a) In order for the majority vote classifier to make a mistake, more than half of K classifiers must fail. Since each classifier fails independently with Bernoulli (ε), the probability that more than $K/2$ out of K trials of independent Bernoulli (ε) variables are 1 gives the desired probability:

$$\varepsilon_{majority} = \sum_{n=\lfloor K/2 \rfloor +1}^{K} \binom{K}{n} \varepsilon^{n} (1-\varepsilon)^{K-n}$$

Were

$\binom{K}{n} \varepsilon^{n} (1-\varepsilon)^{K-n} \to P($exactly n hypothesis make an error$)$

and

$$\binom{K}{n} = \frac{K!}{n!(K-n)!},$$ where "$x!$" is x factoriell. This is "K choose n", a number of distinct ways of choosing n distinct objects from a set of K distinct objects.

Portanto:

	$k=5$	$k=20$
$\varepsilon=0.1$	4.6×10^{-4}	$7{,}0886\times10^{-7}$
$\varepsilon=0.4$	$0{,}08704$	0.1275

CA.

$k=5,\ \varepsilon=0.1$ $\quad \sum_{n=\frac{5}{2}+1}^{5} \binom{k}{n} \varepsilon^{n} (1-\varepsilon)^{k-n}$ $\quad , n=\frac{5}{2}+1=2.5+1=3.5=4$

Então:

$$\varepsilon_{majority}=\frac{5!}{4!(5-4)!}(0.1)^4(1-0.1)^{5-4}+\frac{5!}{5!(5-5)!}(0.1)^5(1-0.1)^{5-5}$$

$$=\boxed{4.6\times10^{-4}}$$

$k=5,\ \varepsilon=0.4$,

$$\varepsilon_{majority}=\frac{5!}{4!(5-4)!}(0.4)^4(1-0.4)^{5-4}+\frac{5!}{5!(5-5)!}(0.4)^5(1-0.4)^{5-5}=$$

$$=0{,}08704$$

c A : $K = 20$, $\varepsilon = 0.1$

$$\varepsilon_{majority} = \sum_{n=11}^{20} \binom{20}{n} \varepsilon^{n} (1-\varepsilon)^{20-n}$$

$n = \frac{20}{2} + 1 = 11$

Enter:

$$\varepsilon_{majority} = \frac{20!}{11!(20-11)!}(0.1)^{11}(1-0.1)^{20-11} +$$

$$\frac{20!}{12!(20-12)!}(0.1)^{12}(1-0.1)^{20-12} +$$

$$\frac{20!}{13!(20-13)!}(0.1)^{13}(1-0.1)^{20-13} +$$

$$\frac{20!}{14!(20-14)!}(0.1)^{14}(1-0.1)^{20-14} +$$

$$\frac{20!}{15!(20-15)!}(0.1)^{15}(1-0.1)^{20-15} +$$

$$\frac{20!}{16!(20-16)!}(0.1)^{16}(1-0.1)^{20-16} +$$

$$\frac{20!}{17!(20-17)!}(0.1)^{17}(1-0.1)^{20-17} +$$

$$\frac{20!}{18!(20-18)!}(0.1)^{18}(1-0.1)^{20-18} +$$

$$\frac{20!}{19!(20-19)!}(0.1)^{19}(1-0.1)^{20-19} +$$

$$\frac{20!}{20!(20-20)!}(0.1)^{20}(1-0.1)^{20-20} = 7{,}0886 \times 10^{-7}$$

Result:

$7{,}0886 \times 10^{-7}$

Utilizando as fórmulas anteriores, para $K = 20$ e $\varepsilon = 0.4$:

$\varepsilon_{majority} = 0,1275$

See a3-solution.pdf

5) If the independence assumption is removed, the ensemble error can be worse than ε. Consider the case where we have 3 classifiers A, B and C, and consider the following scenario where ✓ is a correct prediction and X is an incorrect prediction:

A	B	C	Majority
X	X	✓	X
✓	X	X	X
X	✓	X	X
✓	✓	✓	✓

So, if the above pattern continues, we get:

- $\varepsilon_{majority} = \frac{3}{4}$; $\varepsilon = \frac{6}{12} = \frac{1}{2}$

5. SVMs (ver ap14_final-solution.pdf)

The kernel function is a $k(x, z) = \phi(x) \cdot \phi(z)$

Portanto:

$$k(\bar{x}, \bar{y}) = (\bar{x} \cdot \bar{y} + a)^2, \text{ tal que } k(\bar{x}, \bar{y}) = \phi(\bar{x}) \cdot \phi(\bar{y})$$

$$= \cancel{(a^2 + \bar{x} \cdot \bar{y} \; \bar{x} \bar{y} + 2a \bar{x} \bar{y})} =$$

$$= \left((\bar{x} \cdot \bar{y})^2 + 2a(\bar{x} \cdot \bar{y}) + a^2 \right) =$$

$$= \left(\bar{x} \bar{y} \cdot \bar{y}^T \bar{x} + 2a \bar{y} \cdot \bar{x} + a^2 \right) =$$

$\Rightarrow$ Logo,

$$\bar{y} \bar{y}^T = A \quad ; \quad 2a \bar{y} = \bar{v} \quad ; \quad a^2 = c$$

Para as dimensões baterem certo:

$$\bar{x}^T A \bar{x} + \bar{v}^T \bar{x} + c = 0 \text{ , será a equação do}$$

classification boundary.

6. K-Means

Centróis: μ_1, μ_2, μ_3

Pontos: x_1, x_2, x_3, x_4, x_5, x_6

1ª Iteração: Distância Euclidiana: $d_{Euc}(p_1, p_2) = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$

Ponto \ cluster	$\mu_1(-2.5, 0)$	$\mu_2(0,0)$	$\mu_3(2.5, 0)$	
(-4, 1)	1,803 μ_1	4.123	6.576	
(-4, -1)	1,803 μ_1	4.123	6.576	
(-1, 0)	1.5	1 μ_2	3.5	
(1, 0)	3.5	1 μ_2	1.5	
(4, 1)	6.576	4.123	1.803	μ_3
(4, -1)	6.576	4.123	1.803	μ_3

Novos vector prototypes:

$$\mu_1 = \left(\frac{-4-4}{2}, \frac{1-1}{2}\right) = (-4, 0)$$

$$\mu_2 = \left(\frac{-1-1}{2}, \frac{0+0}{2}\right) = (-1, 0)$$

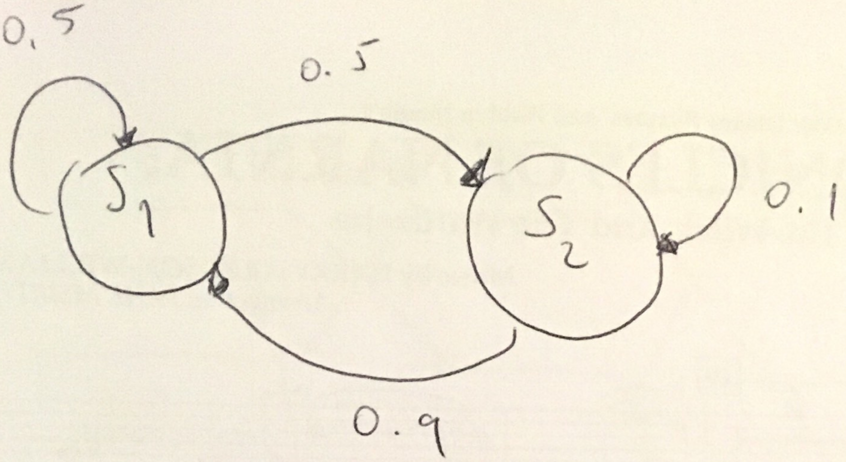
$$\mu_3 = \left(\frac{4+4}{2}, \frac{1-1}{2}\right) = (4, 0)$$

Pontos \ clusters	$\mu_1(-4,0)$	$\mu_2(-1,0)$	$\mu_3(4,0)$
$(-4,1)$	1 μ_1	3.162	8.062
$(-4,-1)$	1 μ_1	3.162	8.062
$(-1,0)$	3	0 μ_2	5
$(1,0)$	5	2 μ_2	3
$(4,1)$	8.062	3.162	1 μ_3
$(4,-1)$	8.062	3.162	1 μ_3

Dado que os pontos que foram atribuídos a cada vector prototype não mudam, isso significa que o algoritmo convergiu para o ótimo.

Logo, os vector prototype finais são:

$$\mu_1 = [-4,\ 0]^T \qquad \mu_2 = [-1,\ 0]^T \qquad \mu_3 = [4,\ 0]^T$$



Initial Probabilities:

$$pi0 = \begin{bmatrix} 0.75 & 0.25 \end{bmatrix}$$

Transition Matrix

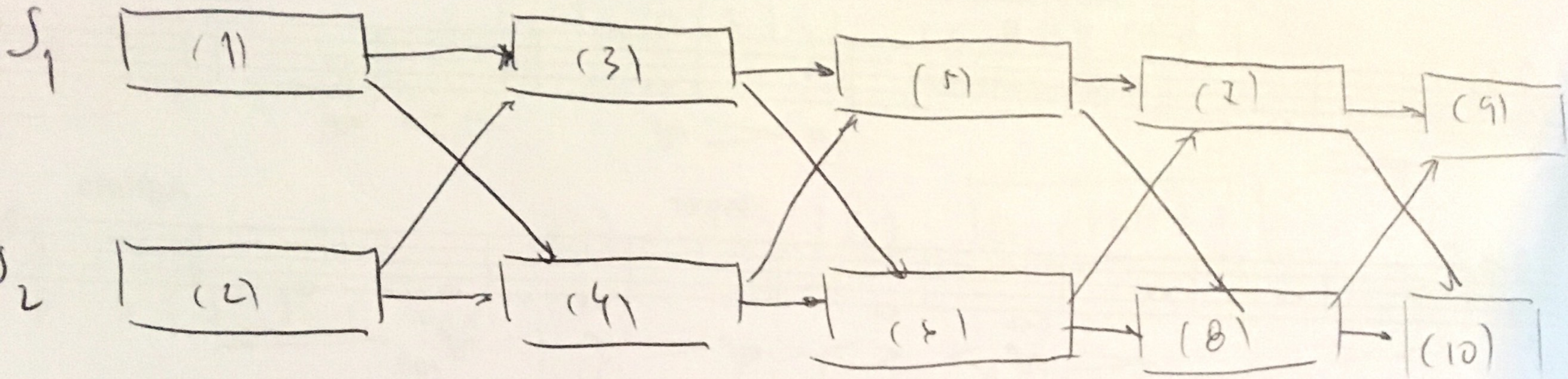
$$\begin{array}{c|c} & \begin{matrix} S_1 & S_2 \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \end{matrix} & \begin{bmatrix} 0.5 & 0.5 \\ 0.9 & 0.1 \end{bmatrix} \end{array}$$

Emission Matrix:

$$\begin{array}{c|c} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \end{matrix} & \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \end{array}$$

a) Running the feedforward algorithm given in class,

the probability for {AAB} is: 0.125



	A	A	B	?	?
	1	2	3	4	5

$(1) = 0.75 \times p(A \mid s_1) = 0.375$

$(2) = 0.25 \times p(A \mid s_2) = 0.125$

$(3) \quad p(A \mid s_1) \times \max \begin{cases} (1) \times p(s_1 \mid s_1) = \boxed{0.1875} \\ (2) \; p(s_1 \mid s_2) = 0.1125 \end{cases}$

$= 0.09375$

$(4) \quad p(A \mid S_2) \times \max \begin{cases} (1) \times p(S_2 \mid S_1) = \boxed{0.1875} \\ (2) \times p(S_2 \mid S_2) = 0.0125 \end{cases}$

$= 0.09375$

(5)

$$p(B \mid S_1) \times \max \begin{cases} (3) \times p(S_1 \mid S_1) = 0.046875 \\ (4) \times p(S_1 \mid S_2) = \boxed{0.084375} \end{cases}$$

$= 0.0421875$

(6)

$$p(B \mid S_2) \times \max \begin{cases} (3) \times p(S_2 \mid S_1) = \boxed{0.046875} \\ (4) \times p(S_2 \mid S_2) = 9.375E-03 \end{cases}$$

$= 0.0234375$

(7) I+A:

$$p(A \mid S_1) \times \max \begin{cases} (5) \times p(S_1 \mid S_1) = 0.02109375 \\ (6) \times p(S_1 \mid S_2) = 0.02109375 \end{cases}$$

$= 0.010546875$

I+B:

$$p(B \mid S_1) \times \max \begin{cases} (5) \times p(S_1 \mid S_1) = 0.02109375 \\ (6) \times p(S_1 \mid S_2) = '' \end{cases}$$

$= \boxed{0.010546875}$

(8)

I+A

$$p(A \mid S_2) \times \max \begin{cases} (5) \times p(S_2 \mid S_1) = \boxed{0.02109375} \\ (6) \times p(S_2 \mid S_2) = 2.34375E-3 \end{cases}$$

$= 0.010546875$

I+B = I+A = $\boxed{0.010546875}$

(9) I_t A

$$P(A|S_1) \times \max \begin{cases} (7) \times p(S_1|S_1) = 5.273 \times 10^{-3} \\ (8) \times p(S_1|S_2) = \cancel{5.273 \times 10^{-3}}\ \boxed{9.492 \times 10^{-3}} \end{cases}$$

$= 4.74 \times 10^{-3}$

$\times 2.6\cancel{37} \times 10^{-3}$

I_t B

$$P(B|S_1) \times \max \begin{cases} (7) \times p(S_1|S_1) = \text{"} \\ (8) \times p(S_1|S_2) = \text{"} \end{cases}$$

$= 4.74 \times 10^{-3}$

(10) I_t A

$$P(A|S_2) \times \max \begin{cases} (7) \times p(S_2|S_1) = \boxed{5.273 \times 10^{-3}} \\ (8) \times p(S_2|S_2) = 1.0546 \end{cases}$$

$= 2.6365 \times 10^{-3}$

$I_t B = I_t A = 2.6365 \times 10^{-3}$

Nesta última fase: Avaliando os cálculos

Intervidos

Verifica-se

que existem probabilidades similares de ser A ou B,

mas igual pelo que se o threshold for $p > 0.5$ então

A, poderia ser

escolhido A.

(c)

(1) Basicamente este algoritmo tem de se modelar para ser um algoritmo de predição já que o output é observado no nosso podemos dizer que o output $x_{t1} \dots x_{t+n}$, sempre nos time step t_{1-1}, e assim fazemos de conta que o output pode ser com a observação do futuro. Ou então, em cada iteração ele deve avaliar a probabilidade de ambos os estados anteriores, e com já tem no modelo.

$C_{(2)}$ (see lecture 33.pdf)

Let's assume a stochastic matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

If $a_{11} \wedge a_{12} \wedge a_{21} \wedge a_{22} \geq 0$, all values nonnegative

e $a_{11} + a_{21} \wedge a_{12} + a_{22} = 1$, ~~all~~ each ~~column~~ row sums to 1.

Then this is a stochastic matrix

We wants to compute that $\begin{bmatrix} a_{11} & \cdots & a_{12} \\ a_{21} & \cdots & a_{22} \\ \vdots & & \\ a_{31} & & a_{37} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

$= \begin{bmatrix} a_{11} + a_{12} + \cdots a_{31} \\ a_{21} + a_{22} + a_{23} \\ a_{31} + a_{32} + a_{37} \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

This computation shows that 1 is an eigenvector of A an $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ is an eigenvector, compea to the eigenvalue 1.

Let λ be an eigenvalue of the stochastic matrix A and let v be a corresponding eigenvector:

So, $A v = \lambda v$

Comparing the i-th row of the both sides, we obtain:

$a_{1 v_1} + a_{2v} v_2 + \cdots + a_{in} v_n = \lambda v_i$

for $i = 1, \ldots, n$

Let $|v_k| = \max \{ |v_1|, |v_2|, \ldots, |v_n| \}$, v_k is the entry of v that has the maximal absolute value

Note that $|v_k| > 0$, since otherwise we have $v = 0$ and this contradicts that an eigenvector is a nonzero vector.

Then, from (*) with $i = k$, we have:

$$|\lambda| \cdot |v_k| = |a_{k1} v_1 + a_{k2} v_2 + \dots + a_{kn} v_n|$$

Triangle inequality!

$a_{ij} \geq 0$

since $|v_k|$ is max!

$$\leq a_{k1}|v_1| + a_2 |v_2| + \dots + a_{kn} |v_n|$$

$$\leq a_{k1}|v_k| + a_2 |v_k| + \dots + a_{kn} |v_k|$$

$$= (a_{k1} + a_{k2} + \dots + a_{kn}) |v_k| = |v_k|$$

Since $|v_k| > 0$, it follows that $\lambda \leq 1$, as required.