Sea un modelo de Markov de conjunto de estados Q = {1, 2, F} y conjunto de símbolos Σ = {a, b}. Se pide:

a) Sean el vector de probabilidades iniciales (π), matriz de transición entre estados (A) y matriz de generación de símbolos (B):

π	1	2
	0.6	0.4

A	1	2	F
1	0.6	0.3	0.1
2	0.3	0.4	0.3

B	a	b
1	0.3	0.7
2	0.8	0.2

Realiza una traza del algoritmo de Viterbi para la cadena y = aab obteniendo la mejor secuencia de estados.

b) Sean las tres cadenas de símbolos: y1 = bbaa, y2 = abab y y3 = aabbb. Al aplicar el algoritmo de Viterbi con un cierto modelo de Markov M , se obtienen, respectivamente, las siguientes secuencias óptimas de estados: 1122F , 2121F y 22111F . A partir de dichas cadenas y sus respectivas se uen ias óptimas de estados, re-estima las probabilidades iniciales (π), de transición (A) y de emisión (B) de M (del mismo modo que se hace en una iteración del algoritmo de
re-estimación de Viterbi).

a) Traza del algoritmo de Viterbi para la cadena $y = aab$:

V	a	a	b	
1	$0.6 \cdot 0.3 = 0.18$	$0.18 \cdot 0.6 \cdot 0.3 = 0.0324$ $0.32 \cdot 0.3 \cdot 0.3 = 0.0288$ $0.0324 > 0.028$ (de 1)	$0.0324 \cdot 0.6 \cdot 0.7 = 0.0136$ $0.1024 \cdot 0.3 \cdot 0.7 = 0.0215$ $0.0136 < 0.0215$ (de 2)	
2	$0.4 \cdot 0.8 = 0.32$	$0.18 \cdot 0.3 \cdot 0.8 = 0.0432$ $0.32 \cdot 0.4 \cdot 0.8 = 0.1024$ $0.0432 < 0.1024$ (de 2)	$0.0324 \cdot 0.3 \cdot 0.2 = 0.0019$ $0.1024 \cdot 0.4 \cdot 0.2 = 0.0082$ $0.0019 < 0.0082$ (de 2)	
F	–	–	–	$0.0215 \cdot 0.1 = 0.0022$ $0.0082 \cdot 0.3 = 0.0025$ $0.0022 < 0.0025$ (de 2)

La secuencia óptima de estados es: 222F

b) La estimación de π, A y B para las cadenas de entrenamiento $y_1 = bbaa$, $y_1 = abab$ y $y_3 = aabbb$ es

π: El estado 1 se ha utilizado una vez como estado inicial y el estado 2 dos veces.

A:
- La transición 1-1 3 veces, la 1-2 2 veces, la 1-F dos veces
- La transición 2-1 3 veces, la 2-2 2 vez, la 2-F una vez

B:
- El símbolo a se ha emitido 0 vez en el estado 1 y 6 veces en el estado 2
- El símbolo b se ha emitido 7 veces del estado 1 y 0 veces del estado 2

Normalizando

	1	2
π	$\frac{1}{3}$	$\frac{2}{3}$

A	1	2	F
1	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{2}{7}$
2	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

B	a	b
1	0.0	1.0
2	1.0	0.0

Sea un problema de clasificación en C clases para objetos representados mediante una característica de tipo contador, $x \in \{0, 1, 2, \ldots\}$. Para toda clase c, suponemos dadas:

- Su probabilidad a priori, $P(c)$.
- Su función de (masa de) probabilidad condicional, $P(x \mid c)$, la cual es Poisson(λ_c) con λ_c conocida.

Se pide:

1. (0.5 puntos) Sea el caso particular: $C = 2$, $P(c = 1) = P(c = 2) = \frac{1}{2}$, $\lambda_1 = 1$, $\lambda_2 = 2$ y $x = 2$. Determina la probabilidad *incondicional* de ocurrencia de $x = 2$, $P(x = 2)$.
2. (0.5 puntos) En el caso particular anterior, halla la probabilidad a posteriori $P(c = 2 \mid x = 2)$, así como la probabilidad de error si $x = 2$ se clasifica en la clase $c = 2$.
3. (0.5 puntos) Más generalmente, para cualquier número de clases C y cualesquiera probabilidades a priori, considera el caso en el que, dado un cierto $\tilde{\lambda} \in \mathbb{R}^+$, $\lambda_c = \tilde{\lambda}$ para todo c. En tal caso, existe una clase que no depende de x, c^*, en la que se puede clasificar todo x con mínima probabilidad de error. Determínala.
4. (0.5 puntos) En el caso general, prueba que el clasificador de Bayes para este problema puede expresarse como un clasificador basado en funciones discriminantes lineales como sigue (ln indica logaritmo natural):

$$c^*(x) = \arg\max_c \; g_c(x) \quad \text{con} \quad g_c(x) = w_c\, x + w_{c0}, \quad w_c = \ln \lambda_c \quad \text{y} \quad w_{c0} = \ln p(c) - \lambda_c$$

Solución:

1. $P(x = 2 \mid c = 1) = \frac{1}{2e} = 0.1839 \qquad P(x = 2 \mid c = 2) = \frac{2}{e^2} = 0.2707.$
   $P(x = 2) = 0.5 \cdot 0.1839 + 0.5 \cdot 0.2707 = 0.2273.$
2. $P(c = 2 \mid x = 2) = \frac{P(c=2) \cdot P(x=2 \mid c=2)}{P(x=2)} = \frac{0.5 \cdot 0.2707}{0.2273} = 0.5955.$
   $P(c \neq 2 \mid x = 2) = 1 - P(c = 2 \mid x = 2) = 0.4045.$
3. $c^*(x) = \arg\max_c P(c)\, P(x \mid c) = \arg\max_c P(c)$ Poisson(λ) $= \arg\max_c P(c) \to c^* = \arg\max_c P(c).$
4.
$$\begin{aligned} c^*(x) &= \arg\max_c \; \ln P(c) + \ln P(x \mid c) \\ &= \arg\max_c \; \ln P(c) - \lambda_c + x \ln \lambda_c - \ln x! \\ &= \arg\max_c \; x \ln \lambda_c + (\ln P(c) - \lambda_c) \end{aligned}$$

Sea M un modelo de Markov de conjunto de estados $Q = \{1, 2, F\}$; alfabeto $\Sigma = \{a, b, c\}$; probabilidades iniciales $\pi_1 = \frac{1}{2}, \pi_2 = \frac{1}{2}$; y probabilidades de transición entre estados y de emisión de símbolos:

A	1	2	F
1	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$
2	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

B	a	b	c
1	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$

Sea x ="ac". Se pide:

1. (0,75 puntos) Haz una traza del algoritmo *Forward* para hallar la probabilidad $P_M(x)$ de que M genere x.
2. (0,75 puntos) Realiza una traza del algoritmo de *Viterbi* para obtener la secuencia de estados más probable, $\tilde{q}_M(x)$, con la que M genera x.
3. (0,50 puntos) Con base en los resultados obtenidos en los apartados anteriores, podemos afirmar que M genera x con probabilidad $P_M(x)$, siguiendo la secuencia de estados $\tilde{q}_M(x)$. ¿Cierto o falso? Razona brevemente la respuesta.
4. (1 punto) A partir de las cadenas de entrenamiento x y "cb", y sabiendo que $\tilde{q}_M(\text{cb})$ = "21F", reestima los parámetros de M mediante el algoritmo de reestimación por Viterbi (hasta convergencia).

1.

α_{qt}	a	c
1	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}+\frac{1}{8}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{64}+\frac{1}{64}=\frac{2}{64}$
2	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{8}\cdot\frac{1}{4}\cdot\frac{1}{2}=\frac{1}{16}+\frac{1}{64}=\frac{5}{64}$

$$P_M(x)=\frac{2}{64}\cdot\frac{1}{4}+\frac{5}{64}\cdot\frac{1}{4}=\frac{7}{256}$$

2.

V_{qt}	a	c
1	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	$\text{máx}\left(\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4},\frac{1}{8}\cdot\frac{1}{2}\cdot\frac{1}{4}\right)=\text{máx}\left(\frac{1}{64},\frac{1}{64}\right)=\frac{1}{64}$
2	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\text{máx}\left(\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{2},\frac{1}{8}\cdot\frac{1}{4}\cdot\frac{1}{2}\right)=\text{máx}\left(\frac{1}{16},\frac{1}{64}\right)=\frac{1}{16}$

$$\tilde{P}_M(x)=\text{máx}\left(\frac{1}{64}\cdot\frac{1}{4},\frac{1}{16}\cdot\frac{1}{4}\right)=\frac{1}{64}$$

$$\tilde{q}_M(x)=\text{"}12F\text{"}$$

3. Falso, M genera x con probabilidad $P_M(x)$, siguiendo la secuencia de estados “12F” ($\tilde{q}_M(x)$), “11F”, “21F” o “22F”. Más precisamente, M genera x mediante “12F” con probabilidad $\tilde{P}_M(x)$; pero también puede generar x mediante una secuencia distinta de “12F”, con probabilidad $P_M(x)-\tilde{P}_M(x)=\frac{7}{256}-\frac{1}{64}=\frac{3}{256}$.

4. En la primera iteración, debemos hallar la secuencia de estados más probable con la que M genera “ac”, así como la secuencia de estados más probable con la que M genera “cb”. La primera, obtenida en el apartado segundo, es "$12F$". La segunda, dada en el enunciado, es "$21F$". A partir de los pares (“ac”,"$12F$") y (“cb”,"$21F$"), obtenemos:

$$\pi_1=\frac{1}{2},\quad \pi_2=\frac{1}{2}$$

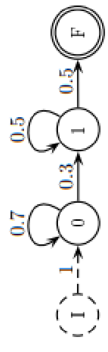
A	1	2	F
1	0	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{2}$	0	$\frac{1}{2}$

B	a	b	c
1	$\frac{1}{2}$	$\frac{1}{2}$	0
2	0	0	$\frac{2}{2}$

En la segunda iteración, partimos de un modelo en el que los símbolos “a” y “b” sólo se emiten en el estado 1, mientras que “c” sólo se emite en el 2. Por tanto, “ac” sólo puede generarse por el camino "$12F$", y “cb” sólo por "$21F$". Esto es, obtenemos los mismos pares (cadena-de-entrenamiento, camino-más-probable) que en la primera iteración, por lo que la segunda iteración termina con el mismo modelo que la primera y el algoritmo de re-estimación acaba.

Se tiene un problema de clasificación en dos clases, A y B, de objetos representados mediante cadenas de símbolos en el alfabeto $\Sigma=\{a,b\}$. Las probabilidades a priori de las clases son $P(A)=0.6$ y $P(B)=0.4$. La función de probabilidad condicional de la clase A viene caracterizada por un cierto modelo de Markov M_A; esto es, $P(x\mid A)=P_{M_A}(x)$. Asimismo, la de la clase B viene dada el modelo de Markov M_B, $P(x\mid B)=P_{M_B}(x)$:

$$\begin{bmatrix}a\\b\end{bmatrix}\qquad\begin{bmatrix}0.9\\0.1\end{bmatrix}\qquad\begin{bmatrix}0.2\\0.8\end{bmatrix}$$



Sea $x=$“aab”. Se sabe que $P_{M_A}(x)=0.063872$. Se pide:

1. (1.5 puntos) Calcula $P_{M_B}(x)$ mediante el algoritmo *Forward*.
2. (0.5 puntos) Sea $\tilde{P}_{M_B}(x)$ la aproximación de Viterbi a $P_{M_B}(x)$. Sabemos que, en general, la aproximación de Viterbi no es mayor que la probabilidad exacta, si bien pueden coincidir. En el caso que nos ocupa ($\tilde{P}_{M_B}(x)$ y $P_{M_B}(x)$) y a la vista de la representación gráfica de M_B, sin necesidad de calcular $\tilde{P}_{M_B}(x)$, ¿se puede afirmar que no coinciden, esto es, que $\tilde{P}_{M_B}(x)$ es *estrictamente* menor que $P_{M_B}(x)$?
3. (0.5 puntos) Halla $P(A\mid x)$ y $P(B\mid x)$.
4. (0.5 puntos) Clasifica x por mínima probabilidad de error.

1.

α_{qt}	a	a	b
0	$1\cdot .9=.9$	$.9\cdot .7\cdot .9+0=.567$	$.567\cdot .7\cdot .1+.054\cdot 0\cdot .1=.03969$
1	$0\cdot .2=0$	$.9\cdot .3\cdot .2+0=.054$	$.567\cdot .3\cdot .8+.054\cdot .5\cdot .8=.15768$

$$P_{M_B}(x)=.03969\cdot 0+.15768\cdot .5=.07884$$

2. En general, la aproximación de Viterbi es *estrictamente* menor que la probabilidad exacta siempre que haya dos o más caminos que generen la cadena con probabilidad no nula. Esto es así en el caso de x ya que M_B puede generarla con probabilidad no nula por dos caminos: $001F$ y $011F$.

3.

$$P(x)=P(A)\,P_{M_A}(x)+P(B)\,P_{M_B}(x)=.6\cdot .063872+.4\cdot .07884=.0383232+.031536=.0698592$$

$$P(A\mid x)=\frac{P(A)\,P_{M_A}(x)}{P(x)}=\frac{.0383232}{.0698592}=.5486$$

$$P(B\mid x)=\frac{P(B)\,P_{M_B}(x)}{P(x)}=\frac{.031536}{.0698592}=.4514$$

4.

$$c(x)=\underset{c\in\{A,B\}}{\arg\max}\ P(c\mid x)=A$$