Tema3.-Modelos-ocultos-de-Markov...

DesastreTotal

Biotecnología

3º Grado en Ingeniería de la Salud

Escuela Técnica Superior de Ingeniería Informática
Universidad de Sevilla

Vídeo explicación del que hemos sacado el ejemplo:

BIOTECNOLOGÍA

TEMA 3: Modelos ocultos de Markov

El modelo oculto de Markov es un modelo probabilístico usado en la predicción de procesos aleatorios. Se denomina oculto porque no sabemos los estados intermedios, sino que simplemente partimos de probabilidades generales a partir de observaciones y vemos cómo se relacionan todas esas observaciones entre sí.

Este modelo se utiliza para describir sistemas complejos que evolucionan con el tiempo. Están compuestos por una serie de estados discretos que cambian a lo largo del tiempo, pero no podemos observar directamente en qué estado se encuentra en un momento dado.

Para entender mejor este concepto, vamos a poner un ejemplo:

Nuestro amigo Marco vive en una ciudad en la que puede haber tres tipos de tiempo: lluvia, niebla y sol. Según el estado de ánimo que tenga nuestro amigo, habrá una probabilidad u otra de que haga sol, llueva o haya niebla.

Marco no nos dice el tiempo que hace en su ciudad, pero sí nos cuenta cuál es su estado de ánimo cada día. Por tanto, para poder averiguar el tiempo que hace en la ciudad donde vive Marco vamos a usar lo que sí conocemos que es su estado de ánimo.

De esta manera, planteamos un modelo oculto de Markov en el que tenemos:

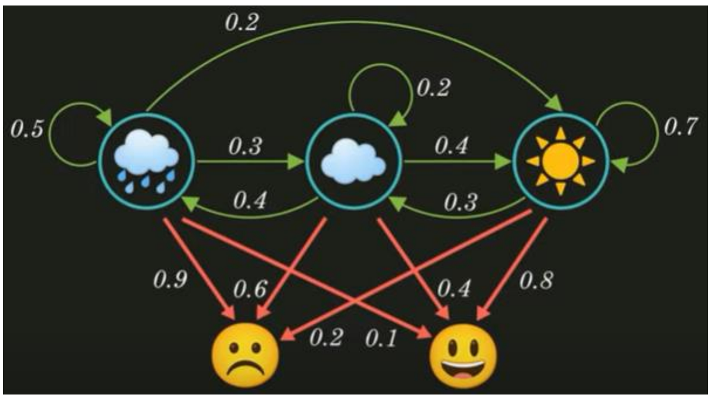
- **Variables Observables**: aquello que sí conocemos, en este caso, el estado de ánimo de nuestro amigo. Digamos que Marco es un chaval de extremos y solo sabe estar muy feliz o muy triste, las variables observables serán: {Yi} = triste, feliz.

  ->Probabilidades de emisión (en rojo)

- **Estados ocultos**: aquellas variables que NO conocemos. En este ejemplo, el tiempo que hace en la ciudad donde vive Marco: {Xi} = lluvia, niebla, sol. -> Esta sería la cadena de Markov normal si no estuviera oculta.

  ->Probabilidades de transición (en verde).

Veamos nuestro ejemplo de forma más visual:



Por tanto, los elementos que definen un modelo de Markov son:

- Dos conjuntos de variables infinitas:
    - {Xi}, por ejemplo, el tiempo que hace en una ciudad.
    - {Yi}, por ejemplo, el estado de ánimo de una persona.
- Los Dominios de estos dos conjuntos de variables:
    - {Xi} = {lluvia, sol, niebla}
    - {Yi} = {feliz, triste}
- Las propiedades que asumimos:
    - Las variables están idénticamente distribuidas:

      $$P(T_{k+i} = t \mid T_k = t') = P(T_{i+1} = t \mid t')$$

    - Las variables están condicionadas a una región:

      $$P(H_k=n \mid T_k=t) = P(Ti=n \mid Ti=t)$$

    - Tener algún conocimiento adicional sobre la secuencia no modifica los valores de la misma

      $P(T_{k+i}=t \mid T_k=t') = P(T_{k+i}=t \mid T_k=t' \text{ y } X=x)$ donde x es cualquier conocimiento sobre la secuencia en posiciones anteriores a k.

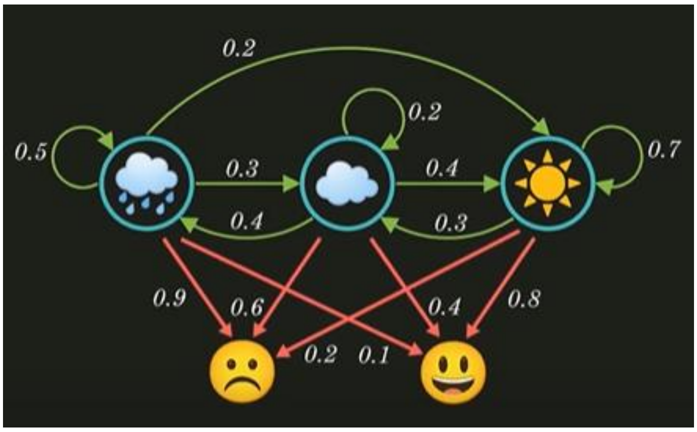
    - El valor de la probabilidad del tipo de región a la que pertenezca el símbolo posterior, no se ve afectado por ningún otro conocimiento sobre los símbolos. -> El modelo asume que las observaciones están relacionadas solo con el estado actual del sistema y no dependen de estados anteriores o futuros.

Ya conocemos el planteamiento general de los modelos ocultos de Markov. Ahora bien, dependiendo de qué es lo que queramos averiguar tendremos que realizar un procedimiento u otro.

- **Filtrado**: para actualizar la distribución de probabilidad de los estados ocultos a medida que se reciben nuevas observaciones.
- **Decodificación**: hallar probabilidades de secuencias
    - Para calcular la probabilidad conjunta de una secuencia de observaciones y una secuencia de estados ocultos -> **Algoritmo de avance** (forward algorithm)
    - Para encontrar la secuencia de estados ocultos más probable dada una secuencia de valores observables -> **Viterbi**
    - Calcular la probabilidad de observar una secuencia de observaciones a partir de un estado particular en el modelo oculto de Markov -> **Algoritmo de retroceso** (backward algorithm)
- **Aprendizaje**: a partir de un conjunto de secuencias de símbolos y los posibles estados que pueden darse, estimamos diferentes probabilidades sin conocer previamente la relación entre las secuencias y los estados.
    - **Algoritmo Baum-Welch** (no lo vamos a dar)

Antes de comenzar a explicar los tipos de algoritmos es importante entender cómo se obtienen las matrices A, B y pi:

Si tenemos el siguiente dibujo:



Tal y como hemos explicado antes, el tiempo son los estados ocultos y el estado de ánimo son los estados observables.

La matriz A está formada por las probabilidades que encontramos en las flechas verdes de los estados ocultos.

La matriz B está formada por las probabilidades que encontramos en las flechas rojas de los estados observables.

$$\begin{array}{c|ccc|cc} & \text{🌧} & \text{☁} & \text{☀} & \text{☹} & \text{😃} \\ \hline \text{🌧} & 0.5 & 0.3 & 0.2 & 0.9 & 0.1 \\ \text{☁} & 0.4 & 0.2 & 0.4 & 0.6 & 0.4 \\ \text{☀} & 0.0 & 0.3 & 0.7 & 0.2 & 0.8 \end{array}$$

La matriz pi nos dice la probabilidad de que una secuencia empiece en el estado 1 lluvia o que empiece en el estado 2 sol.

Vídeo explicación del que hemos sacado el ejemplo:

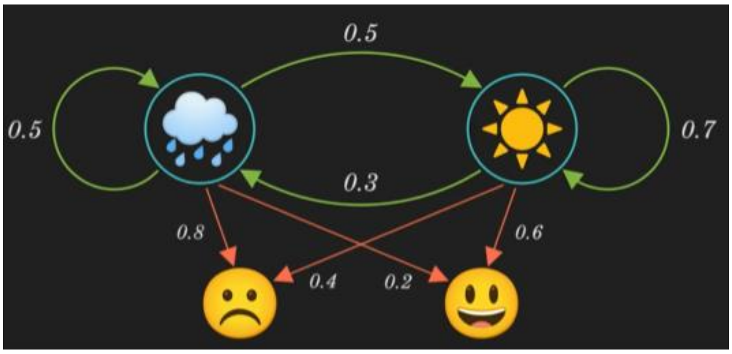
1) **ALGORITMO DE AVANCE (FORWARD)**

El algoritmo de avance nos va a permitir averiguar la probabilidad de que se dé una cierta secuencia de estados observables.

¡OJO!: lo que nos devuelve el algoritmo de avance es la probabilidad conjunta y a partir de esta podemos hallar las probabilidades de que se dé una cierta secuencia de estados.

Continuando con el ejemplo anterior, podríamos buscar la probabilidad de que el día 1 Marco esté triste y el día dos esté triste y el día tres esté feliz, es decir, P(Y = Yo, Yo, Y_1).

Simplifiquemos el ejemplo, solo va a haber dos estados no observables: que llueva (Xo) y que llueva (X1).



$$A = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

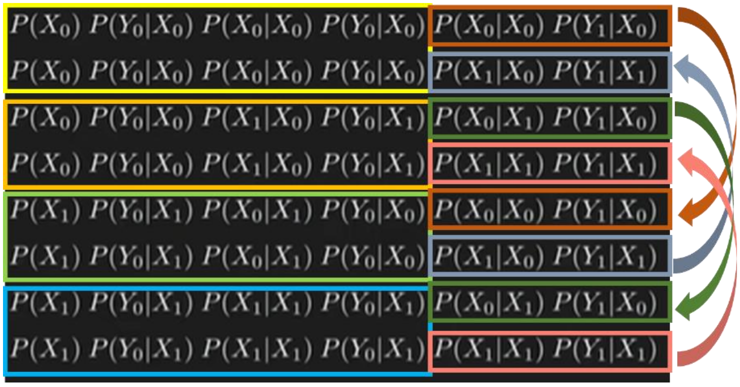
$$\pi = [0.375 \quad 0.625]$$

Sabiendo esto, podríamos intentar averiguar la probabilidad anterior haciendo:

$$P(X_0)\ P(Y_0|X_0)\ P(X_0|X_0)\ P(Y_0|X_0)\ P(X_0|X_0)\ P(Y_1|X_0)$$
$$P(X_0)\ P(Y_0|X_0)\ P(X_0|X_0)\ P(Y_0|X_0)\ P(X_1|X_0)\ P(Y_1|X_1)$$
$$P(X_0)\ P(Y_0|X_0)\ P(X_1|X_0)\ P(Y_0|X_1)\ P(X_0|X_1)\ P(Y_1|X_0)$$
$$P(X_0)\ P(Y_0|X_0)\ P(X_1|X_0)\ P(Y_0|X_1)\ P(X_1|X_1)\ P(Y_1|X_1)$$
$$P(X_1)\ P(Y_0|X_1)\ P(X_0|X_1)\ P(Y_0|X_0)\ P(X_0|X_0)\ P(Y_1|X_0)$$
$$P(X_1)\ P(Y_0|X_1)\ P(X_0|X_1)\ P(Y_0|X_0)\ P(X_1|X_0)\ P(Y_1|X_1)$$
$$P(X_1)\ P(Y_0|X_1)\ P(X_1|X_1)\ P(Y_0|X_1)\ P(X_0|X_1)\ P(Y_1|X_0)$$
$$P(X_1)\ P(Y_0|X_1)\ P(X_1|X_1)\ P(Y_0|X_1)\ P(X_1|X_1)\ P(Y_1|X_1)$$

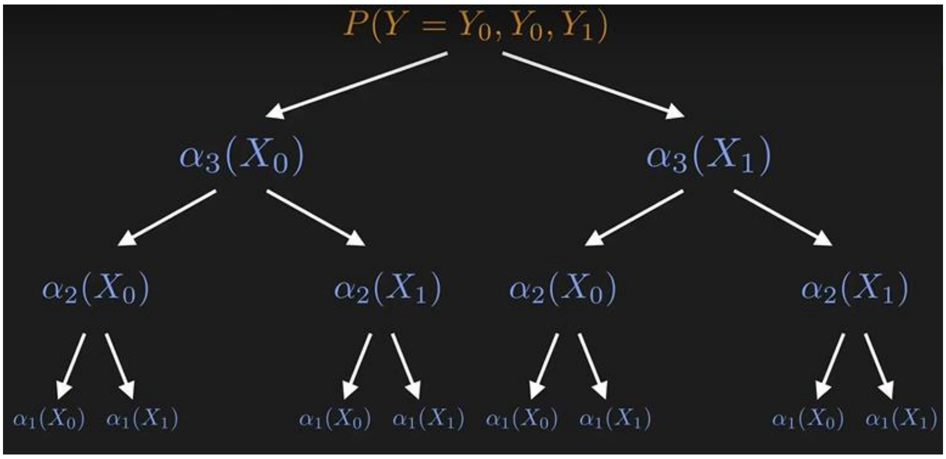
Tendríamos que hallar todas estas probabilidades y sumarlas. Esto es un proceso muy largo, sobre todo teniendo en cuenta que en este ejemplo tan solo trabajamos con dos estados ocultos y dos estados observables. El algoritmo de avance nos permite hallar esta probabilidad de forma más práctica.

Si observamos las probabilidades de arriba, vemos que hay términos que se repiten:



Lo que hacemos básicamente en el algoritmo de avance es agrupar estos términos que se repiten en α de manera que en vez de tener que hacerlo todo de nuevo, podemos multiplicar por el α que corresponda.

Para ello, vamos a ir desde un α_1 más pequeño hasta el α_{final}. Tendremos que hallar α para cada uno de los estados ocultos, es decir, $\alpha(Xo)$ y $\alpha(X_1)$:



El resultado final será la suma de los dos tipos de α_{final} que tenemos:

$$P(Y=Yo, Yo, Y_1) = \alpha_3(Xo) + \alpha_3(X_1)$$

Una vez hemos entendido cómo funciona el algoritmo de avance, solo nos queda entender cómo hallamos las α:

Para la α_1:

- $\alpha_1(Xo)$ -> El primer estado observable es Yo, es decir, queremos ver la probabilidad de que Marco esté triste si llueve:

$$\alpha_1(Xo) = P(Xo)P(Yo|Xo) = 0.375 * 0.8$$

Google Gemini: Plan Pro a 0€ durante 1 año.

Tu ventaja por ser estudiante

Entra en wlh.es/estudiacongemini

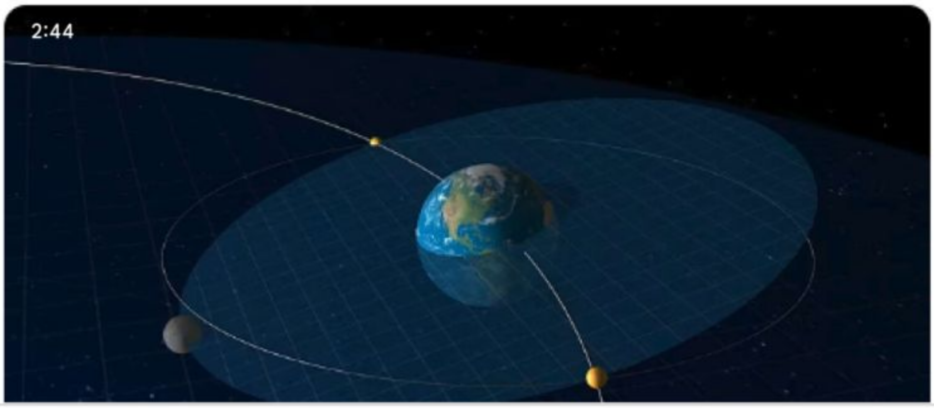
Consigue la oferta

Domina cualquier tema con el Aprendizaje Guiado de Google Gemini.

Puedes explicarme como se crea un eclipse lunar completo y sus fases?

Un momento...

¡Claro vamos a conseguir que lo entiendas a la perfección!



Introduce una petición para Gemini

Aprendizaje Guiado ×

Oferta válida hasta el 9 de diciembre de 2025

Después, 21,99€/mes. 18+. Los resultados/la compatibilidad del dispositivo varían. Comprobar la exactitud de las respuestas. Se aplican restricciones de almacenamiento y de usuario. Se requiere una cuenta de Google. Consulta los términos y condiciones.

(*) Para hacer este cálculo, hacemos uso de las matrices A, B y pi que podemos hallar a partir del dibujo.

- $\alpha_1(X1)$ -> El primer estado observable es Yo. En este caso, vamos a ver la probabilidad de que Marco esté triste si hace sol.

$$\alpha_1(X1) = P(X1)P(Yo|X1) = 0.625*0.4$$

Por tanto, a partir de los estados observables que tenemos, vamos a ver la probabilidad de que dicho estado observable se dé para cada uno de los estados ocultos que tenemos.

A continuación, hallamos la α_2. Para ello, haremos uso de la α_1:

Para α_2, el segundo estado es Yo:

- $\alpha_2(Xo)$:

$$\begin{aligned}\alpha_2(X_0) &= \overbrace{\alpha_1(X_0)}^{\text{Llueve si Marco está triste}}\ \overbrace{P(X_0|X_0)}^{\text{Llueve el 2º día si el 1º día llovió}}\ \overbrace{P(Y_0|X_0)}^{\text{Marco está triste si llueve el 2º día}} + \overbrace{\alpha_1(X_1)}^{\text{Hace sol si Marco está triste}}\ \overbrace{P(X_0|X_1)}^{\text{Llueve el 2º día si el primer día hizo sol}}\ \overbrace{P(Y_0|X_0)}^{\text{Marco está triste si llueve el 2º día}} \\ &= 0.3 \times 0.5 \times 0.8 + 0.25 \times 0.3 \times 0.8 \\ &= 0.18\end{aligned}$$

- $\alpha_2(X_1)$: lo mismo que en el caso anterior, pero en el caso de que en el segundo día haga sol

$$\begin{aligned}\alpha_2(X_1) &= \alpha_1(X_0)\ P(X_1|X_0)\ P(Y_0|X_1) + \alpha_1(X_1)\ P(X_1|X_1)\ P(Y_0|X_1) \\ &= 0.3 \times 0.5 \times 0.4 + 0.25 \times 0.7 \times 0.4 \\ &= 0.13\end{aligned}$$

Haremos lo mismo para las α_3, solo que para averiguar estas últimas, usaremos las α_2:

$$\begin{aligned}\alpha_3(X_0) &= \alpha_2(X_0)\ P(X_0|X_0)\ P(Y_1|X_0) + \alpha_2(X_1)\ P(X_0|X_1)\ P(Y_1|X_0) \\ &= 0.18 \times 0.5 \times 0.2 + 0.13 \times 0.3 \times 0.2 \\ &= 0.0258\end{aligned}$$

$$\begin{aligned}\alpha_3(X_1) &= \alpha_2(X_0)\ P(X_1|X_0)\ P(Y_1|X_1) + \alpha_2(X_1)\ P(X_1|X_1)\ P(Y_1|X_1) \\ &= 0.18 \times 0.5 \times 0.6 + 0.13 \times 0.7 \times 0.6 \\ &= 0.1086\end{aligned}$$

Por tanto, la probabilidad de que se dé la secuencia es:

$$P(Y = Yo, Yo, Y_1) = 0.13 + 0.1086 = 0.1344$$

AVANCE NORMALIZADO

Al igual que nos ocurría cuando veíamos el tema de alineaciones, cuando trabajamos con una gran cantidad de datos, al calcular las probabilidades, estas nos puedan salir muy pequeñas. Esto hace que la precisión con la que calculemos las probabilidades dependa de la precisión del ordenador con el que hagamos los cálculos.

Nos pueden salir probabilidades del tipo 0.00000234, si nuestro ordenador solo llega a 5 decimales, esta probabilidad se quedaría en 0.00000. Los logaritmos se usan para que podamos calcular probabilidades sin depender de la precisión del ordenador.

Para solucionar esto, en alineaciones hacíamos la log-probabilidad. Sin embargo, en avance estamos haciendo sumas por lo que no podemos hacer el logaritmo, tendremos que normalizar antes.

Para hacer avance normalizado tenemos que dividir cada valor de entrada por la suma de todos los valores de entrada anteriores:

$$\hat{\alpha}_1(e_j) = \frac{\alpha'_1(e_j)}{\sum_{e \in \mathcal{E}} \alpha'_1(e)} \quad \text{para todo } 1 \leq j \leq m$$

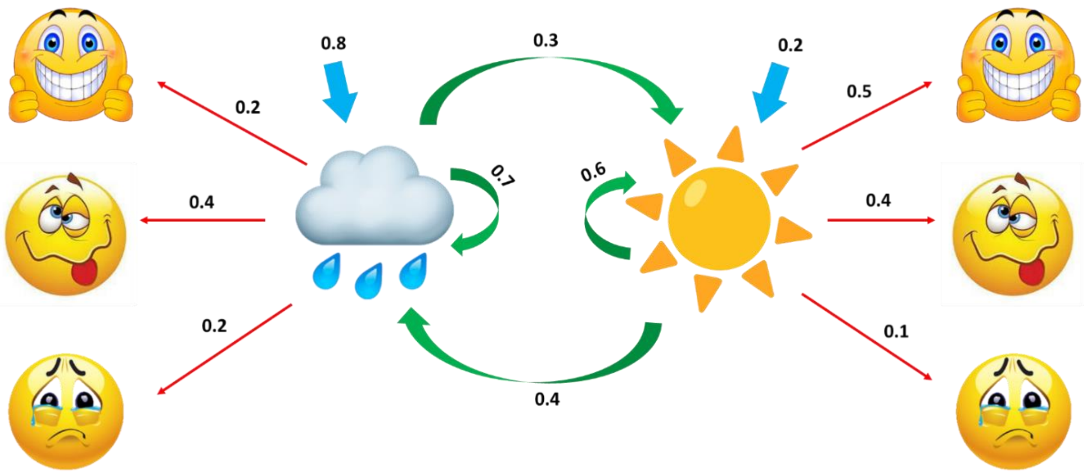
(*) De la misma forma, el alfa 2 normalizada sería alfa 2 / suma de las alfas normalizadas. -> Así con todas las alfas que tengamos.

Cuando lleguemos a la última alfa, podemos calcular la probabilidad condicionada despejando:

Probabilidad condicionada = $\frac{\text{última } \alpha}{\text{suma de todas las } \alpha'}$ -> Esto pasa a ser una multiplicación, por lo que sí podemos hacer el logaritmo.

2) ALGORITMO DE RETROCESO (BACKWARD)

Para entender este algoritmo, vamos a cambiar ligeramente el ejemplo. Tendremos los estados ocultos: llueve o hace sol. Y tres estados observables: estar feliz, estar triste y estar confuso porque no entiendes nada de biotecnología:



MATRIZ Pi = [0.8   0.2]

MATRIZ A

	0.7	0.3
	0.4	0.6

MATRIZ B

	0.2	0.4	0.4
	0.5	0.4	0.1

El algoritmo de avance, nos permitía hallar la probabilidad de que se diera una cierta secuencia de estados observables. Por ejemplo, si Marco está confuso el día 1, triste el día 2, confuso el día 3 y feliz el día 4:

$\alpha_1(\ )$	$\alpha_2(\ )$	$\alpha_3(\ )$	$\alpha_4(\ )$
$\alpha_1(\ )$	$\alpha_2(\ )$	$\alpha_3(\ )$	$\alpha_4(\ )$
$\beta_1(\ )$	$\beta_2(\ )$	$\beta_3(\ )$	$\beta_4(\ )$
$\beta_1(\ )$	$\beta_2(\ )$	$\beta_3(\ )$	$\beta_4(\ )$

α_2 nos permite conocer la probabilidad de que el día 2 esté triste y de que el día 1 esté feliz, es decir, las α (algoritmo de avance) nos permitirá hallar las probabilidades de los estados observables de la secuencia que están ocurriendo o que han ocurrido: "***El algoritmo Forward*** *se utiliza para calcular la probabilidad de observar una secuencia de* *símbolos a lo largo de un modelo oculto de Markov, desde el estado inicial hasta un estado final."*

Para conocer la probabilidad de que el día 3 esté feliz, necesitaremos la β_2 (algoritmo de retroceso): *"El **algoritmo Backward**, por otro lado, se utiliza para calcular la probabilidad de observar los símbolos restantes después de estar en un estado determinado."*

El algoritmo de retroceso se realiza de forma similar al de avance. La diferencia principal se encuentra en que vamos a ir desde β_{final} hasta la $\beta_{inicial}$.

Si tenemos la secuencia:

S_k = {Y1: triste, Y2: feliz, Y3: confuso}

Z_k = {Xo: lluvia, X1: sol}

k indica la posición en la que nos encontramos

Este algoritmo, tal y como su propio nombre indica, comienza desde el final:

- **β_5** (día 5 está triste):
  - $\beta_5(Xo) = 1$
  - $\beta_5(X1) = 1$

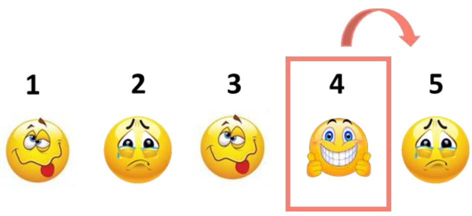
Las probabilidades en los estados ocultos lluvia y sol son iguales a 1 en la última posición porque se está asumiendo que se ha observado la secuencia completa "confuso, triste, confuso, feliz, triste". Al estar en el último símbolo, no hay más símbolos observados después de ese punto, por lo que la probabilidad de observar los símbolos restantes (en este caso, ninguno) es 1.

Por tanto, para hallar β_4 habría que multiplicar por β_5, esto es multiplicar por 1 por lo que no lo vamos a escribir. El algoritmo verdaderamente comienza a continuación:

- **β_4**

Al igual que hacíamos con las alfas, tendremos que hallar las betas para todos los estados ocultos posibles:

- $\beta_4(Xo) = P(Z_5=Xo/Z_4=Xo)*P(S_5=Y3/Z_5=Xo) + P(Z_5=X1/Z_4=Xo)*P(S_5=Y3/Z_5=X1) = 0.7*0.4 + 0.3*0.1 = 0.31$
  - Si la secuencia es de tamaño 5, la primera beta que hallamos es la 4, de manera que estamos en la zona 4 para hallar la zona 5:



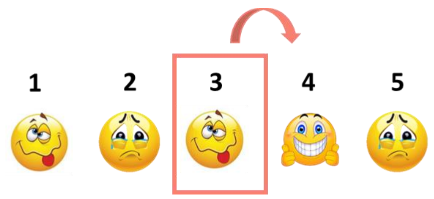
- Este algoritmo pretende averiguar los símbolos siguientes. En este caso, si estamos en el estado 4: feliz, queremos saber qué estado es más probable que esté en el día 5
- Si estamos viendo β_4 (Xo), estamos asumiendo que en el día 4 ha llovido, por eso Z_4 que hemos puesto son Xo. Esto cambia en el caso de β_4 (X1) tal y como veremos a continuación.
- Si traducimos lo que estamos haciendo:

$$\beta4\ (Xo) = \overbrace{P(Z5=Xo/Z4=Xo)}^{\text{Llueve el día 5, sabemos que ha llovido el día 4}} * \overbrace{P(S5=Y1/Z5=Xo)}^{\text{Está triste el día 5 si sabemos que llueve el día 5}} + \overbrace{P(Z5=X1/Z4=Xo)}^{\text{Hace sol el día 5, sabemos que ha llovido el día 4}} * \overbrace{P(S5=Y1/Z5=X1)}^{\text{Está triste el día 5 si sabemos que hace sol el día 5}}$$

  - β_4 (X1) = $P(Z_5=Xo/Z_4=X1)*P(S_5=Y3/Z_5=Xo) + P(Z_5=X1/Z_4=X1)*P(S_5=Y3/Z_5=X1) = 0.22$

- **β_3**:

Para hallar el resto de betas, necesitaremos usar la beta posterior. En este caso, vamos a mirar para que el siguiente símbolo de la secuencia, es decir, que el día 4 esté feliz:



  - β_3 (Xo) = β_4 (Xo) * $P(Z_4=Xo/Z_3=Xo)$* $P(S_4=Y2/Z_4=Xo)$ + β_4 (X1) *$P(Z_4=X1/Z_3=Xo)$* $P(S_4=Y2/Z_4=X1)$

  - β_3 (X1) = β_4 (Xo) * $P(Z_4=Xo/Z_3=X1)$* $P(S_4=Y2/Z_4=Xo)$ + β_4 (X1) *$P(Z_4=X1/Z_3=X1)$ * $P(S_4=Y2/Z_4=X1)$

- **β_2**: para que el día 3 esté confuso:
  - β_2 (Xo) = β_3 (Xo) * $P(Z_3=Xo/Z_2=Xo)$ * $P(S_3=Y3/Z_3=Xo)$ + β_3 (X1) *$P(Z_3=X1/Z_2=Xo)$* $P(S_3=Y3/Z_4=X1)$

  - β_2 (X1) = β_3 (Xo) * $P(Z_3=Xo/Z_2=X1)$ * $P(S_3=Y3/Z_3=Xo)$ + β_3 (X1) *$P(Z_3=X1/Z_2=X1)$ * $P(S_3=Y3/Z_4=X1)$

- **β_1**: para que el día 2 esté triste:

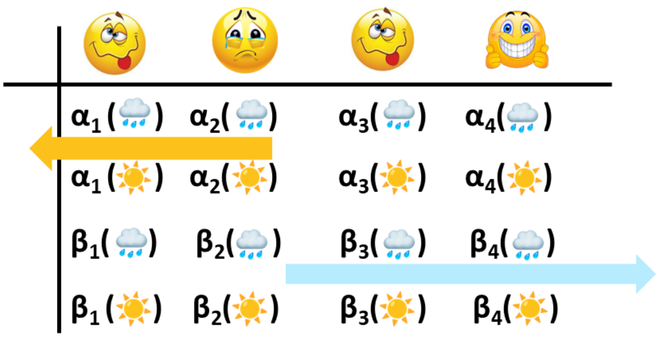
- β_1 (Xo) = β_2 (Xo) * P(Z_2=Xo/Z_1=Xo) * P(S_2=Y1/Z_2=Xo) + β_2 (X1) *P(Z_2=X1/Z_1=Xo) * P(S_2=Y1/Z_2=X1)

- β_1 (X1) = β_2 (Xo) * P(Z_2=Xo/Z_1=X1) * P(S_2=Y1/Z_2=Xo) + β_2 (X1) *P(Z_2=X1/Z_1=X1) * P(S_2=Y1/Z_2=X1)

El resultado final del algoritmo Backward es una tabla de valores beta que contiene la probabilidad de observar los símbolos restantes después de estar en un estado determinado, para cada posición en la secuencia de observación. En nuestro ejemplo, el resultado final sería una tabla de valores beta que muestra la probabilidad de estar en el estado lluvioso o soleado en cada posición de la secuencia de observación.

3) ALGORITMO FORWARD-BACKWARD O SMOOTHING

Tal y como comentábamos cuando explicamos el algoritmo Backward, en una cierta secuencia, las alfas (algoritmo forward) nos permiten averiguar qué estados observables han ocurrido antes mientras que, las betas (algoritmo backward) nos permiten averiguar qué estados van a ocurrir después:



Si combinamos ambos algoritmos, tenemos el algoritmo **Forward-Backward**. Este nos devuelve la probabilidad de estar en cada estado oculto en cada instante de tiempo, dado un conjunto de observaciones. Es decir, proporciona una estimación de la probabilidad de que el modelo haya estado en un estado oculto en un momento dado, dado el conocimiento de las observaciones realizadas.

Si hemos hallado que:

α_3 (Xo) = 0.00535

α_3 (X1) = 0.01373

β_3 (Xo) = 0.49552

β_3 (X1) = 0.48976

-> Si llueve: α_3 (Xo)* β_3 (Xo) = 0.00265

-> Si hace sol: α_3 (X1)* β_3 (X1) = 0.0072

P(secuencia) = 0.00265 + 0.0072 = 0.00938

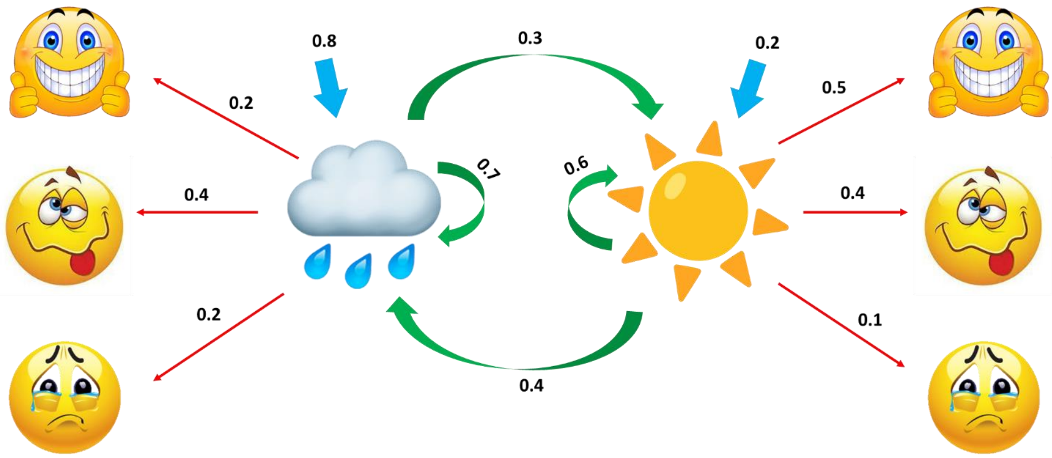
Vídeo explicación que he usado para entenderlo:

4) ALGORITMO VITERBI

En el algoritmo de avance hallábamos la probabilidad de que se diera cierta secuencia de estados. Sin embargo, Viterbi nos permite para una cierta secuencia de estados ocultos, en qué región (estado observable) es más probable qué se dé dicha secuencia.

Es decir, si nos dan P(Y = Yo, Y1, Yo), ¿a qué secuencia de estados ocultos (tiempo) es más probable que pertenezca?, es decir, ¿es más probable que haga sol, llueva, haga sol o que llueva, haga sol y haga sol, o que llueva los tres días…?

Continuemos con el ejemplo de antes:



Las matrices A, B y pi serán:

MATRIZ Pi = [0.8   0.2]

MATRIZ A

	0.7	0.3
	0.4	0.6

MATRIZ B

	0.2	0.4	0.4
	0.5	0.4	0.1

Nuestro objetivo va a ser averiguar qué tiempo es más probable que haya hecho cada uno de los tres días, si sabemos que Marco ha estado triste-feliz-triste.

Sabiendo está secuencia, vamos a ir viendo cada día si es más probable que llueva o que haga sol:

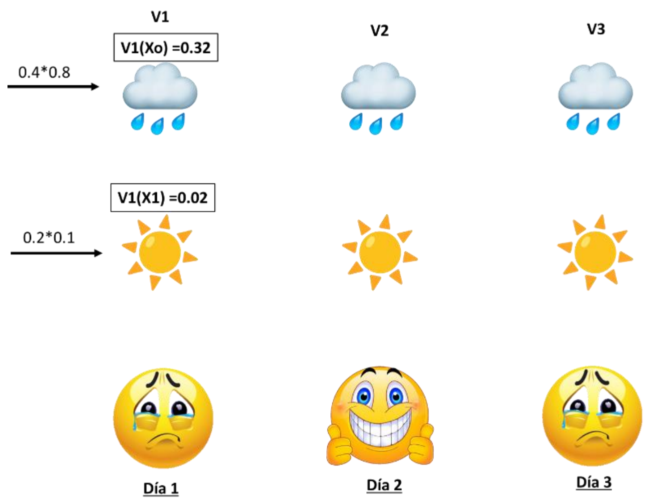
- **V1**. Primero vamos a averiguar si, sabiendo que en el día 1 Marco está triste, es más probable que llueva o que haga sol.

Para calcular las V1 haremos:

- **V1(Xo)** = $P(Yo/Xo)P(Xo) = 0.4*0.8 =$ **0.32**
- **V1(X1)** = $P(Yo/X1)P(X1) = 0.1*0.2 =$ **0.02**

Recuerda:	
Lluvia = Xo	Triste = Yo
Sol = X1	Feliz = Y1

Hagamos un esquema de cómo va avanzando el algoritmo para verlo todo de forma más visual:



- **V2**: en el día 2 Marco está feliz. Tenemos cuatro posibilidades. Calculamos cada una de estas probabilidades:
  - Que hoy llueva si ayer llovió:

    $$P(Y1,Xo) = P(Y1/Xo)P(Xo/Xo) = 0.2*0.7 = 0.14$$

  - Que hoy llueva si ayer hizo sol:

    $$P(Y1,Xo)' = P(Y1/Xo)P(Xo/X1) = 0.5*0.3 = 0.15$$

  - Que hoy haga sol si ayer llovió:

    $$P(Y1,X1) = P(Y1/X1)P(X1/Xo) = 0.2*0.4 = 0.08$$

  - Que hoy haga sol si ayer hizo sol:

    $$P(Y1,X1)' = P(Y1/X1)P(X1/X1) = 0.5*0.6 = 0.3$$

  Queremos averiguar cuál es la V2(lluvia) y la V2(sol), por lo que vamos a multiplicar cada una de estas posibilidades por la V1 que le corresponde:

  - Día 1 llovió, por lo que multiplico por la V1 (día 1) en el que llovió:

    $$P(Y1,Xo) = 0.14 * V1(\text{lluvia}) = 0.0448$$

P(Y1,Xo)´= 0.15 * V1(lluvia) = 0.048

- Día 1 hizo sol, por lo que multiplico por la V1 (día 1) en la que hizo sol:

P(Y1,X1) =0.08 *V1(sol) = 0.016

P(Y1,X1)´= 0.3*V1(sol) = 0.06

Sin embargo, estamos buscando qué tiempo es el más probable, por lo que vamos a coger los máximos en el caso de que en el día 2 haga sol o en el día 2 llueva:

- **Día 2 llueve:**

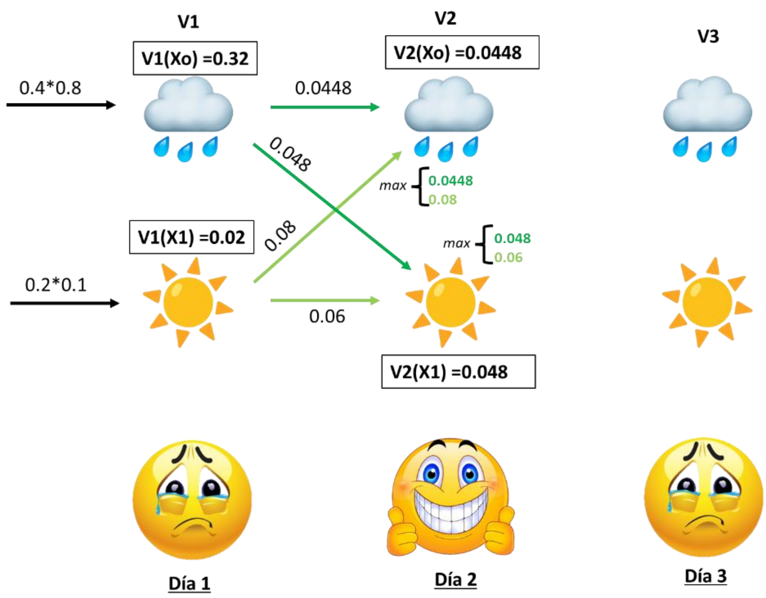
  P(Y1,Xo) = 0.14 *V1(lluvia) = 0.0448
  P(Y1,X1) =0.08 *V1(sol) = 0.016

  El máximo entre estos dos números es 0.0448 -> **V2(Xo) = 0.0448**

- **Día 2 hace sol:**

  P(Y1,X1)´= 0.3*V1(sol) = 0.06
  P(Y1,Xo)´= 0.15 * V1(lluvia) = 0.048

  El máximo entre estos dos números es 0.048 -> **V2(X1) = 0.048**



Para hallar la V3, repetiremos el mismo proceso que para hallar la V2. Observamos todos los casos, los multiplicamos por la V2 que les corresponda y nos quedamos con el máximo en cada caso:

- Día 2 llovió:

  P(Xo,Yo) = P(Xo/Yo)*P(Xo/Xo)*V2(Xo) = 0.4 *0.7*0.0448 = 0.012544
  P(X1,Yo) = P(X1/Yo)*P(X1/Xo)*V2(Xo) = 0.1*0.3*0.0448 = 0.0001344

- Día 2 hizo sol:
  P(Xo,Yo)´= P(Xo/Yo)*P(Xo/X1)*V2(X1) = 0.4*0.4*0.048 = 0.000768
  P(X1,Yo)´= P(X1/Yo)*P(X1/X1)*V2(X1) = 0.1*0.6*0.048 = 0.000288

A continuación, hacemos los máximos correspondientes:

- El día 3 llueve:

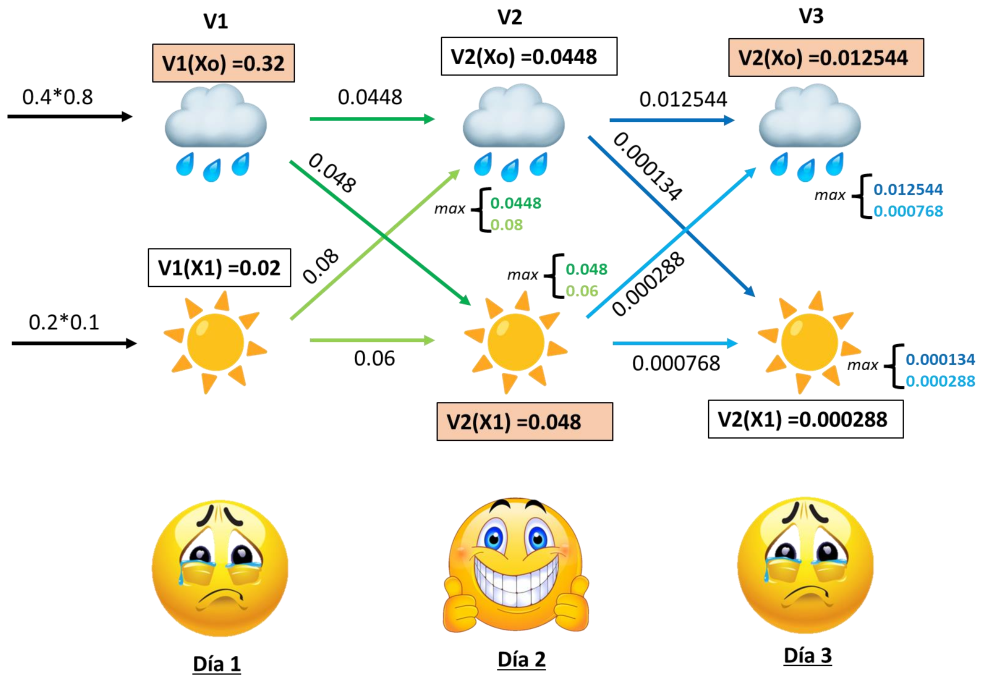
  P(Xo,Yo) = **0.012544**
  P(Xo,Yo)´= 0.000768
  V3(Xo) = 0.012544

- El día 3 hace sol:

  P(X1,Yo) = 0.0001344
  P(X1,Yo)´= **0.000288**
  V3(X1) = 0.000288

No olvidemos que todo esto lo estamos haciendo para sacar la secuencia de estados ocultos más probable a partir de una secuencia de estados observables. Lo último que nos queda es comparar: V1(Xo) y V1 (X1), el que sea mayor, será el estado correspondiente a la secuencia final. De la misma forma, comparamos el resto de V.

Veámoslo gráficamente:



Por tanto, la secuencia que hemos hallado con el algoritmo de Viterbi es: lluvia-sol-lluvia.

Hagamos un resumen:

ALGORITMOS	QUÉ DEVUELVE	QUÉ PUEDO CALCULAR CON ESTE ALGORITMO
Avance	Probabilidades conjuntas	-Probabilidad de una secuencia dada -La región más probable para la última posición de una secuencia
Retroceso	Un conjunto de probabilidades que indican la probabilidad de observar los símbolos restantes después de estar en un estado determinado.	-Región más probable para la primera posición -Probabilidad de parte de la secuencia
Avance-retroceso = Smoothing	Probabilidades de las posiciones intermedias	-Región más probable en zonas intermedias -Probabilidad de la secuencia -Probabilidad condicionada
Viterbi	Una sucesión de regiones	La sucesión de regiones