Lección 8:
Markov II

Jesús Savage, Carlos Rivera

23 de abril de 2020

Índice

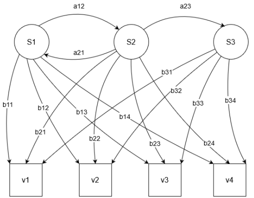
Ejercicios

Encuentre una secuencia de 10 valores de un experimento de volados, con tres monedas. Indique los parametros de la HMM: N, M, A, B, V, π y T para este experimento.

Asúmase las siguientes probabilidades, con sol=S y aguila=A:

	S_1	S_2	S_3
P(S)	0.5	0.75	0.25
P(A)	0.5	0.25	0.75

Las probabilidades de transición a cada uno de los estados es $\frac{1}{3}$. Las probabilidades iniciales son iguales a $\frac{1}{3}$



Ejercicio

El número de estados es el número de monedas: $N = 3$. Los símbolos son $V = (Sol, Aguila)$, por lo tanto $M = 2$. El modelo $\lambda = (A, B, \pi)$, es:

$$\begin{aligned}\pi &= [\pi_1 \pi_2 \pi_3] \\ &= \left[\frac{1}{3}\frac{1}{3}\frac{1}{3}\right]\end{aligned}$$

$$\begin{aligned}A &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}\end{aligned}$$

Ejercicio

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$B = \begin{bmatrix} 0.5 & 0.5 \\ 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$$

Ejercicio

1 - Usando el generador de secuencias se observa la siguiente:

$$\underline{O} = (SSSSASAAAA)$$

a) ¿Cuál es la secuencia de estados más probable?
b) ¿Cuál es la probabilidad de la secuencia observada y de la secuencia de estados más probable?

Ejercicio

Ya que todas las transiciones de estados es equiprobable, la secuencia de estados más probable es una en la cual la probabilidad de cada observación individual es máxima. Así:

$$\underline{q} = (2222323333)$$

La probabilidad de $\underline{O}$ y $\underline{q}$ (dada en el modelo) es:

$$P(\underline{O}, \underline{q}|\lambda) = (0.75)^{10} \left(\frac{1}{3}\right)^{10}$$

Ejercicio

2 - ¿Cuál es la probabilidad que la secuencia observada venga enteramente del estado 1?

$$\underline{\hat{q}} = (1111111111)$$

$$P(\underline{O}, \underline{\hat{q}}|\lambda) = (0.5)^{10} \left(\frac{1}{3}\right)^{10}$$

Formando la siguiente razón:

$$R = \frac{P(\underline{O}, \underline{q}|\lambda)}{P(\underline{O}, \underline{\hat{q}}|\lambda)} = \left(\frac{3}{2}\right)^{10} = 57.67$$

Lo que muestra que $\underline{q}$ es más probable que $\underline{\hat{q}}$

Ejercicio

3 - Considerese la siguiente secuencia de observación:

$$\underline{\hat{O}} = (SAASASSAAS)$$

¿Cómo se cambiarán las respuestas a y b?
Ya que $\underline{\hat{O}}$ tiene el mismo número de S's y de A's, las respuestas en 1 y 2 será el mismo, ya que los estados más probables ocurren en el mismo número de veces en ambos casos.

Ejercicio

4 - Si las probabilidades de transición fueran:

$$a_{11} = 0.90 \quad a_{21} = 0.45 \quad a_{31} = 0.45$$

$$a_{12} = 0.05 \quad a_{22} = 0.10 \quad a_{32} = 0.45$$

$$a_{13} = 0.05 \quad a_{23} = 0.45 \quad a_{33} = 0.10$$

¿Cómo cambiarían las probabilidades de los problemas 1 y 2?

Ejercicio

$$\underline{O} = (SSSSASAAAA)$$

$$\underline{q} = (2222323333)$$

$$P(\underline{O}, \underline{q}|\lambda) = (0.75)^{10} \left(\frac{1}{3}\right) (0.1)^{6} (0.45)^{3}$$

La nueva probabilidad de O y $\hat{q}$ que se convierte en:

$$\hat{q} = (1111111111)$$

$$P(O, \hat{q}|\lambda) = (0.50)^{10} \left(\frac{1}{3}\right) (0.9)^{9}$$

$$R = \left(\frac{2}{3}\right)^{10} \left(\frac{1}{9}\right)^{6} \left(\frac{1}{2}\right)^{3} = 1.36x10^{-5}$$

$\underline{\hat{q}}$ es más probable que $\underline{q}$

Ejercicio

Ahora, la probabilidad de $\underline{\hat{O}}$ y $\underline{q}$ no es la misma que la probabilidad de $\underline{O}$ y $\underline{q}$.

$$\underline{\hat{O}} = (SAASASSAAS)$$

$$P(\underline{\hat{O}}, \underline{q}|\lambda') = \frac{1}{3}(0.1)^6(0.45)^5(0.25)^4(0.75)^6$$

$$P(\underline{\hat{O}}, \underline{\hat{q}}|\lambda') = (0.50)^{10}\left(\frac{1}{3}\right)(0.9)^9$$

Con

$$R = \left(\frac{1}{9}\right)^6\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^4\left(\frac{2}{3}\right)^6 = 1.67x10^{-7}$$

Claramente porque $a_{11} = 0.9$, $\hat{q}$ es más probable

Problemas básicos de las HMM's

Problema 1:
Dada la secuencia de observaciones $\underline{O} = (O_1, O_2, ..., O_T)$ y el modelo $\lambda = (A, B, \pi)$, ¿Cómo se calcula eficientemente la probabilidad de la secuencia observada, dado el modelo λ?
Este problema se llama problema de evaluación.

Problemas básicos de las HMM's

Problema 2:
Dada la secuencia de observación $\underline{O} = (O_1, O_2, ..., O_T)$ y el modelo λ, ¿Cómo se escoje la secuencia óptima $\underline{q}_i = (q_1, q_2, ..., q_T)$ que mejor explique las observaciones dadas?
Se trata de encontrar las variables escondidas en el modelo.

Problemas básicos de las HMM's

Problema 3:
¿Cómo se calculan los parámetros del modelo $\lambda = (A, B, \pi)$ que maximizan la $p(\underline{O}|\lambda)$?
El procedimiento que máximiza la $p(\underline{O}|\lambda)$ se utiliza para entrenar los HMM's.

Problemas básicos de las HMM's

1- Solución de Problema 1.
Evaluación de la probabilidad.
Calcular $P(\underline{O}|\lambda)$ dado $\underline{O} = (O_1, O_2, ..., O_T)$ y $\lambda(A, B, \pi)$.
La forma más directa de hacer esto es a través de enumerar cada secuencia posible de tamaño T (el número de observaciones). Hay N^T posibles secuencias. Considérese una secuencia de estados fijo:

$$\underline{q} = (q_1, q_2, ..., q_T)$$

Entonces la probabilidad de tener la secuencia $\underline{O} = (O_1, O_2, ..., O_T)$ dada la secuencia $\underline{q} = (q_1, q_2, ..., q_T)$ y el modelo $\lambda(A, B, \pi)$:

$$P(\underline{O}|\underline{q}, \lambda) = \frac{P(\underline{O}, \underline{q}, \lambda)}{P(\underline{q}, \lambda)}$$

$$= b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdot ... \cdot b_{q_T}(O_T)$$

Donde q_1 es el estado inicial.

Problemas básicos de las HMM's

La probabilidad de tener esta secuencia de estados

$$P(\underline{q}|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} ... a_{q_{T-1} q_T} = \frac{P(q,\lambda)}{P(\lambda)}$$

La probabilidad de que ocurran $\underline{O}$ y $\underline{q}$ simultáneamente es simplemente el producto de los dos términos anteriores:

$$= P(\underline{O}|\underline{q},\lambda)P(\underline{q}|\lambda) = \frac{P(\underline{O},\underline{q},\lambda)}{P(\underline{q},\lambda)} \cdot \frac{P(\underline{q},\lambda)}{P(\lambda)}$$

$$= \frac{P(\underline{O},\underline{q},\lambda)}{P(\lambda)} = P(\underline{O},\underline{q}|\lambda)$$

Problemas básicos de las HMM's

Recordar que para una función de densidad conjunta si se hace la suma sobre una de las variables se encuentra la función de densidad de la otra variable:

$$\sum_{i=1}^{N} f(x_i, y_j) = f(y_j)$$

Por lo tanto la probabilidad de $\underline{O}$ (dado el modelo) es obtenida sumando la probabilidad conjunta sobre todas las posibles secuencias de estados $\underline{q}$

$$P(\underline{O}|\lambda) = \sum_{all\ \underline{q}} P(\underline{O}, \underline{q}|\lambda)$$

$$= \sum_{all\ \underline{q}} P(\underline{O}|\underline{q}, \lambda) P(\underline{q}|\lambda)$$

$$= \sum_{q_1, q_2 \ldots q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \ldots a_{q_{T-1} q_T} b_{q_T}(O_T)$$

Problemas básicos de las HMM's

Por lo tanto para calcular la probabilidad de las observaciones $\underline{O}$ dado el modelo λ, $P(\underline{O}|\lambda)$, de acuerdo a la fórmula necesita $2T \cdot N^T$ cálculos, ya que para $t = 1, 2, ..., T$ hay N posibles estados que pueden ser alcanzados (es decir, hay N^T posibles secuencias de estados), por cada secuencia de estados se necesitan $2T$ operaciones para cada término de la sumatoria. Para ser precisos, se necesitan $(2T - 1)N^T$ multiplicaciones y $N^T - 1$ sumas.

Por ejemplo, para $N = 5$ estados, $T = 100$ observaciones se necesitan $2 \cdot 100 \cdot 5^{100} \approx 10^{72}$. Se necesita un método eficiente para resolver este problema.

El procedimiento hacia adelante

Considérese la variable hacia adelante $\alpha_t(i)$ definida como:

$$\alpha_t(i) = P(O_1, 0_2, ..., O_t, q_t = i|\lambda)$$

es decir, la probabilidad de observar la secuencia de $O_1, O_2, ..., O_t$ (hasta el tiempo t) y el estado i en el tiempo t, dado el modelo λ. Se puede resolver $\alpha_t(i)$ inductivamente, como se indica a continuación.
1- Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

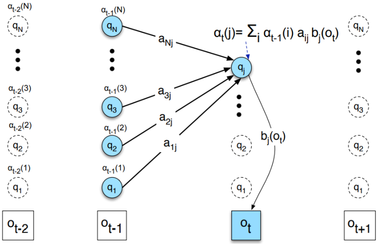
2- Inducción:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] \cdot b_j(O_{t+1});$$

$1 < t < T - 1, \qquad 1 < j < N.$

El procedimiento hacia adelante

Como se puede ver por la siguiente figura el estado es S_j es alcanzado de venir de cualquier de los estados $S_1, S_2, ..., S_N$.



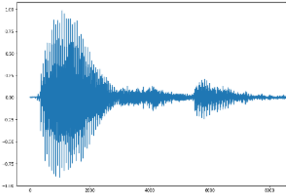
3) Termino:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{1}$$

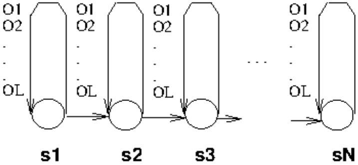
Para calcular $\alpha_t(j), (1 \leq t \leq T, 1 \leq j \leq N)$, requiere $N^2 T$ calculos, en lugar de $2TN^T$

Reconocimiento de palabras aisladas usando HMMs

Cada palabra se representa con un modelo de cadena de Markov oculto, λ_i. Para señales de voz, como la que se muestra a continuación:



El tipo de HMM para representar este tipo de señales de voz es uno de izquierda a derecha, como se muestra en la siguiente figura.

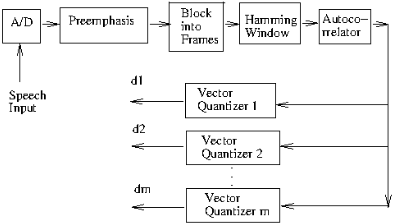


Reconocimiento de palabras aisladas usando HMMs

Entonces para crear modelo de cadena de Markov oculto de la palabra i a ser reconocida, λ_i, lo primero que hay que hacer es un cuantizador vectorial genérico que abarque los vectores de LPC (correlaciones) de todas las palabras a ser reconocidas. Es decir, se cuentan las muestras de entrenamiento de todas las palabras, se encuentran sus correlaciones y después los vectores de LPC de los centroides del cuantizador.
Los simbolos del modelo λ_i representan los indices de los centroides a los cuales los vectores de LPC se parecen más.
Los estados ocultos representan la parte temporal de la señal de voz.
Una vez encontrado el modelo, punto que se vera más adelante, para reconocer la palabra se encontraria los simbolos utilizando el cuantizador vectorial genérico y a partir de ahí la probabilidad $P(O|\lambda_i)$

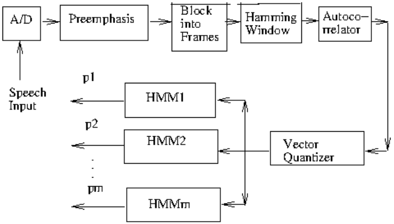
Reconocimiento de palabras aisladas usando HMMs

Como recordatorio se pone la figura de reconocimiento de palabras de voz aisladas, en este caso cada palabra se representa por un cuantizador.



Reconocimiento de palabras aisladas usando HMMs

Se muestra a continuación el diagrama de bloques para un reconocedor de palabras aisladas usan HMMs.



Dado el vector de observaciones $\underline{O} = (O_1, O_2, ..., O_t, ..., O_T)$, obtenido después de cuantizar la señal de voz, en donde cada observación O_t corresponde al índice del centroide que más se parece al vector LPC de uno de los bloques de la señal de voz, se obtiene la probabilidad $p_j = P(O|\lambda_j)$. Se reconoce la palabra i si:

$$i = argmax[P(O|\lambda_j)], 1 \leq j \leq M$$