



- Métodos computacionales:
Alejandro Segura
- Estadística
 - a) Incluir el código Notebook (.ipynb).
 - b) Guardar la información en una carpeta llamada **Semana15_Nombre1_Nombre2**
 - c) **Hacer una sola entrega por grupo.**

Contents

1	Estadística	3
1.1	Hidden Markov models	4
1.2	Metrópolis Hastings - Molécula de Hidrógeno (H_2 *** Challenge)	7

List of Figures

1	Probabilidad de cada secuencia oculta para el estado $\Omega_O = [V, A, R]$	5
2	Probabilidad de cada secuencia oculta para el estado $\Omega_O = [S, C, C, C, S, C, S, C]$	6
3	Nube electrónica encontrada con el algoritmo de Metropolis asociada al electrón en cada núcleo.	7
4	Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de $U_{min} \approx 2.0 \text{ eV}$ y se da a una separación nuclear de $d_{equilibrium} \approx 2.1a_0$	8
5	Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno incluyendo el efecto de Heitler y London. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de $U_{min} \approx 2.05 \text{ eV}$ y se da a una separación nuclear de $d_{equilibrium} \approx 2.3a_0$	8

1 Estadística

1.1 Hidden Markov models

1. Un proceso de Markov oculto es un proceso estocástico definido por dos variables aleatorias, la variable oculta $x(t)$ toma valores en el paso t y la variable observada $y(t)$ toma valores en el paso t . La variable $x(t)$ toma valores siguiendo la matriz de transición \mathbb{T} y la variable $y(t)$ toma valores siguiendo la matriz de emisión \mathbb{E} . Por ejemplo, definimos las siguientes matrices de transición y emisión:

$$\mathbb{T} = \begin{pmatrix} & F & T \\ F & 0.7 & 0.3 \\ T & 0.5 & 0.5 \end{pmatrix} \quad (1)$$

$$\mathbb{E} = \begin{pmatrix} & R & V & A \\ F & 0.8 & 0.1 & 0.1 \\ T & 0.2 & 0.3 & 0.5 \end{pmatrix} \quad (2)$$

Las filas de estas matrices representan las probabilidades de los estados en el tiempo $t - 1$ y las columnas las probabilidades asociadas al tiempo t . La suma de todos los valores en una fila debe ser igual a 1. Definamos los estados de transición como el estados de animo de una persona; Feliz y Triste como: $S = [F, T]$ con las siguientes probabilidades a priori $\pi = [0.4, 0.6]$. De esta forma, la probabilidad de estar feliz al día $(t - 1)$ y luego estar triste al siguiente día (t) es: $\mathbb{T}_{01} = 0.3$. Los estados de transición que siguen este proceso aleatorio no son observables por lo que se denominan ocultos. Sin embargo, esta probabilidad de transición se propaga influyendo en la manera de vestir de la persona. Definamos los estados de emisión como el color de la ropa que usa la persona al tiempo $t - 1$; Rojo, Verde, Azul como $E = [R, V, A]$. En este contexto, la probabilidad de que la persona use ropa azul dado que está feliz es $\mathbb{E}_{02} = 0.1$.

La secuencia de n eventos observados es:

$$\Omega_O = \{y(t_1), y(t_2), \dots, y(t_n)\} \quad (3)$$

Usando la propiedad de los procesos de Markov, se busca encontrar la secuencia de eventos ocultos más probable.

$$\Omega_{hidden} = \{x(t_1), x(t_2), \dots, x(t_n)\} \quad (4)$$

Suponga la siguiente secuencia de ropa usada por la persona: $\Omega_O = [V, A, R]$.

- a) Encuentre la probabilidad de cada secuencia de estados (ocultos) de ánimo.
- b) Encuentre el estado de animo más probable que tuvo la persona durante esos tres días.
- c) Calcule las probabilidades de cada estado observable (i) como la suma de las probabilidades de todos los estados ocultos. Verifique que sumando todos los estados observables la medida es 1.

$$\sum \mathbb{P}_i = 1 \quad (5)$$

Remarks:

n_h es el número de estados ocultos y n_o es el número de estados observados. El número de secuencias posibles (n_s) está dada por las variaciones con repetición de estas cantidades:

$$n_s = n_h^{n_o} = 8 \quad (6)$$

De las ocho secuencias ocultas posibles, supongamos que la secuencia oculta es $S_1 = [F, F, T]$. Entonces:

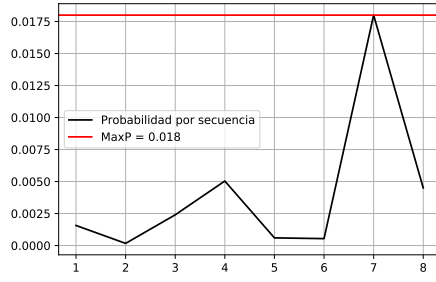


Figure 1: Probabilidad de cada secuencia oculta para el estado $\Omega_O = [V, A, R]$.

$$S_1 = \begin{pmatrix} F & F & T \\ V & A & R \end{pmatrix} \quad (7)$$

Recordemos la probabilidad condicional de dos sucesos:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A/B)\mathbb{P}(B) \quad (8)$$

Por tanto, la probabilidad de la secuencia es:

$$\begin{aligned} \mathbb{P}(V, A, R \cap F, F, T) &= \underbrace{\mathbb{P}(V/F)\mathbb{P}(A/F)\mathbb{P}(R/T)}_{\text{Emission Matrix}} \times \underbrace{\mathbb{P}(F)}_{\text{Prior}} \underbrace{\mathbb{P}(F/F)\mathbb{P}(T/F)}_{\text{Transition Matrix}} \\ &= 0.1 \cdot 0.1 \cdot 0.2 \times 0.4 \cdot 0.7 \cdot 0.3 = 1.68 \times 10^{-4} \end{aligned} \quad (9)$$

En general:

$$\mathbb{P}(O_i, q_i) = \prod_{i=1}^n \mathbb{P}(O_i/q_i) \prod_{i=1}^n \mathbb{P}(q_i/q_{i-1}) \quad (10)$$

2. Un casino tiene dos tipos monedas una justa y una sesgada, las probabilidades de obtener cara y sello son: $P_j = [0.5, 0.5]$ y $P_s = [0.9, 0.1]$ respectivamente. La matriz de emisión está dada por:

$$\mathbb{E} = \begin{pmatrix} & C & S \\ J & 0.5 & 0.5 \\ B & 0.9 & 0.1 \end{pmatrix} \quad (11)$$

El tipo de moneda se puede escoger siguiendo está ley de transición:

$$\mathbb{T} = \begin{pmatrix} & J & B \\ J & 0.8 & 0.2 \\ B & 0.2 & 0.8 \end{pmatrix} \quad (12)$$

Se realiza un experimento de 8 lanzamientos y se obtiene la siguiente secuencia:

$$\Omega_O = [S, C, C, C, S, C, S, C] \quad (13)$$

- Use la siguiente distribución de probabilidad a-priori $\pi = [0.2, 0.8]$ para la moneda justa y sesgada.
- Encuentre la secuencia más probable del tipo de moneda que se eligió en cada lanzamiento.
- ¿Depende el resultado de la probabilidad a-priori?
- Calcule las probabilidades de cada estado observable (i) como la suma de las probabilidades de todos los estados ocultos. Verifique que sumando todos los estados observables la medida es 1.

$$\sum \mathbb{P}_i = 1 \quad (14)$$

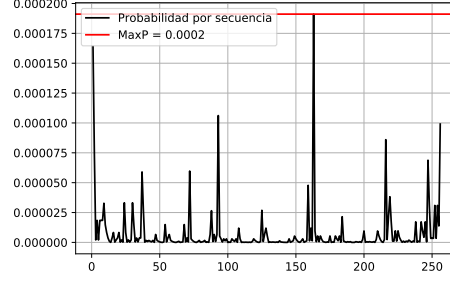


Figure 2: Probabilidad de cada secuencia oculta para el estado $\Omega_O = [S, C, C, C, S, C, S, C]$.

- Las bases nitrogenadas fundamentales que componen el ADN son: Adenina (A), Citosina (C), Guanina (G) y Timina (T). Un gen se puede representar a través de una secuencia ordenada de dichas bases. Suponga la siguiente matriz de transición entre bases:

$$\mathbb{T} = \begin{pmatrix} & A & C & G & T \\ A & 0.4 & 0.2 & 0.2 & 0.2 \\ C & 0.25 & 0.25 & 0.25 & 0.25 \\ G & 0.3 & 0.3 & 0.1 & 0.3 \\ T & 0.1 & 0.1 & 0.1 & 0.7 \end{pmatrix} \quad (15)$$

La probabilidad a-priori está dada por:

$$\pi = [0.25, 0, 0.5, 0.25] \quad (16)$$

Encuentre la probabilidad de obtener el siguiente gen:

$$g = [T, G, C, T, C, A, A, A] \quad (17)$$

$$P_g = 7.5 \times 10^{-6}$$

1.2 Metrópolis Hastings - Molécula de Hidrógeno (H_2 *** Challenge)

1. La molécula de Hidrógeno está compuesta por dos núcleos separados una distancia L . En torno a cada núcleo existe un electrón que supondremos que está en el nivel más bajo de energía, en un orbital s . La función de onda que describe a cada electrón está dada por:

$$\psi(\vec{r}, \vec{R}) = \frac{1}{\sqrt{\pi a_0^3}} e^{-|\vec{r} - \vec{R}|/a_0} \quad (18)$$

donde $a_0 = 0.529 \text{ \AA}$ es el radio de Borh y \vec{R} es la posición del núcleo respecto al sistema de referencia. Adicionalmente, el operador de energía electrostática de la molécula (en electronvoltios), que incluye la interacción entre electrones y núcleos está dado por:

$$U(\vec{r}_1, \vec{r}_2, \vec{R}_1, \vec{R}_2) = \frac{1}{|\vec{r}_1 - \vec{r}_2|} + \frac{1}{|\vec{R}_1 - \vec{R}_2|} - \frac{1}{|\vec{r}_1 - \vec{R}_1|} - \frac{1}{|\vec{r}_2 - \vec{R}_1|} - \frac{1}{|\vec{r}_1 - \vec{R}_2|} - \frac{1}{|\vec{r}_2 - \vec{R}_2|} \quad (19)$$

Usar un sistema de unidades donde $a_0 = 1$ y poner los núcleos en las posiciones: $\vec{R}_1 = [0, 0, L/2]$ y $\vec{R}_2 = [0, 0, -L/2]$. La idea es calcular el valor esperado de la energía potencial como función de la separación de los núcleos de la molécula. Este promedio está dado (sin considerar el espín electrónico) por:

$$\langle U \rangle = \int d^3r_1 d^3r_2 |\psi(\vec{r}_1, \vec{R}_1)|^2 |\psi(\vec{r}_2, \vec{R}_2)|^2 U(\vec{r}_1, \vec{r}_2, \vec{R}_1, \vec{R}_2) \quad (20)$$

- a) Note que esta integral se puede ver como el valor promedio de la energía potencial evaluada en el muestreo de la densidad de probabilidad de cada electrón. Usando el algoritmo de Metropolis en coordenadas cartesianas, hacer el muestreo del cuadrado de la función de onda fijando $L = 2$ y 10^5 pasos en la cadena. La distribución de puntos de la nube electrónica asociada a cada electrón está dada por [3]:

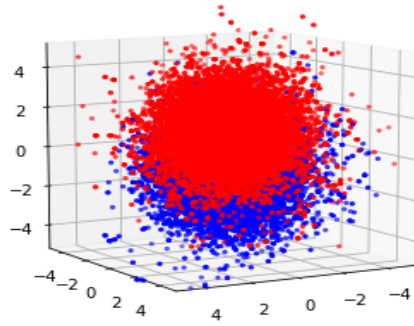


Figure 3: Nube electrónica encontrada con el algoritmo de Metropolis asociada al electrón en cada núcleo.

- b) Calcule el valor esperado de la energía potencial para el siguientes separaciones nucleares.
`1 = np.linspace(1.0,3.0,10)`. Debería obtener algo similar a [5]:

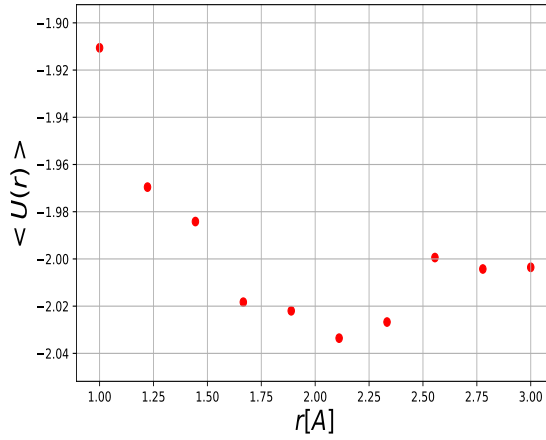


Figure 4: Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de $U_{min} \approx 2.0 \text{ eV}$ y se da a una separación nuclear de $d_{equilibrium} \approx 2.1a_0$.

c) Incluya el efecto de Heitler y London (buscar el sentido físico):

$$\langle U \rangle = \int d^3r_1 d^3r_2 |\psi(\vec{r}_1, \vec{R}_1)\psi(\vec{r}_2, \vec{R}_2) + \psi(\vec{r}_2, \vec{R}_1)\psi(\vec{r}_1, \vec{R}_2)|^2 U(\vec{r}_1, \vec{r}_2, \vec{R}_1, \vec{R}_2) \quad (21)$$

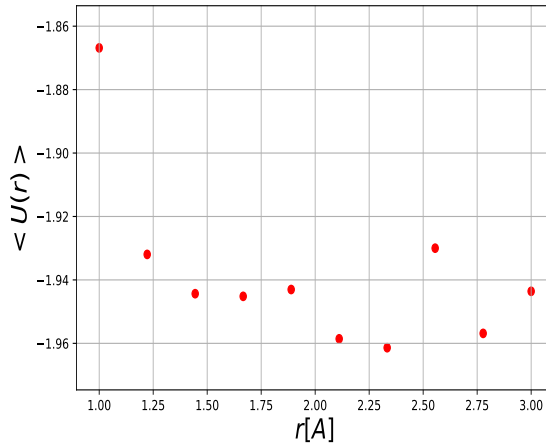


Figure 5: Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno incluyendo el efecto de Heitler y London. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de $U_{min} \approx 2.05 \text{ eV}$ y se da a una separación nuclear de $d_{equilibrium} \approx 2.3a_0$.

d) Note que los valores están de acuerdo a los observados en la naturaleza: La distancia de separación de los protones es $d_{equilibrium} = 2.4a_0$ y la energía de ligadura es 2.8 eV . Nice isn't it!