- 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

## List of Figures

# 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:

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

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

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 $Ps = [0.4, 0.6]$. 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:

$$\Omega_O = \{y(t_1), y(t_2), ..., y(t_n)\} \tag{3}$$

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

$$\Omega_{hidden} = \{x(t_1), x(t_2), ..., x(t_n)\} \tag{4}$$

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

a) Encuentre la probabilidad de cada secuencia de estados de ánimo (ocultos).

b) Encuentre el estado de animo más probable que tuvo la persona durante esos tres días.

$n_h$ 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:

$$n_s = n_h^{n_o} = 8 \tag{5}$$

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

$$S_1 = \begin{pmatrix} F & F & T \\ V & A & R \end{pmatrix} \tag{6}$$

Recordemos la probabilidad condicional de dos sucesos:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A/B)\mathbb{P}(B) \tag{7}$$

Por tanto, la probabilidad de la secuencia es:

$$\begin{aligned} \mathbb{P}(V, A, R \cap F, F, T) &= \underbrace{\mathbb{P}(V/F)\mathbb{P}(A/F)\mathbb{P}(R/T)}_{\text{Emission Matrix}} \times \underbrace{\mathbb{P}(F)}_{\text{Prior}} \underbrace{\mathbb{P}(F/F)\mathbb{P}(T/F)}_{\text{Transition Matrix}} \\ &= 0.1 \cdot 0.1 \cdot 0.2 \times 0.4 \cdot 0.7 \cdot 0.3 = 1.68 \times 10^{-4} \end{aligned} \tag{8}$$

En general:

$$\mathbb{P}(O_i, q_i) = \prod_{i=1}^{n} \mathbb{P}(O_i/q_i) \prod_{i=1}^{n} \mathbb{P}(q_i/q_{i-1}) \tag{9}$$

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:

$$\mathbb{E} = \begin{pmatrix} & C & S \\ J & 0.5 & 0.5 \\ B & 0.9 & 0.1 \end{pmatrix} \tag{10}$$

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

$$\mathbb{T} = \begin{pmatrix} & J & B \\ J & 0.8 & 0.2 \\ B & 0.2 & 0.8 \end{pmatrix} \tag{11}$$

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

$$\Omega_O = [S, C, C, C, S, C, S, C] \tag{12}$$

a) Defina alguna distribución de probabilidad a-priori ($P_0$).

b) Encuentre la secuencia más probable del tipo de moneda que se eligió en cada lanzamiento.

c) ¿Depende el resultado de la probabilidad a-priori?

3. 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:

$$\mathbb{T} = \left(\begin{array}{c|cccc} & A & C & G & T \\ \hline A & 0.4 & 0.2 & 0.2 & 0.2 \\ C & 0.25 & 0.25 & 0.25 & 0.25 \\ G & 0.3 & 0.3 & 0.1 & 0.3 \\ T & 0.1 & 0.1 & 0.1 & 0.7 \end{array}\right) \tag{13}$$

La probabilidad a-priori está dada por:

$$P_0 = [0.25, 0, 0.5, 0.25] \tag{14}$$

Encuentre la probabilidad de obtener el siguiente gen:

$$g = [T, G, C, T, C, A, A, A] \tag{15}$$

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