

# Minería de Datos

## Lab. #4

Ivan Saavedra, Ph.D.

[saavedrai@uninorte.edu.co](mailto:saavedrai@uninorte.edu.co)

Universidad del Norte  
División de Ingenierías  
Dpto. Ingeniería de Sistemas



202110

# 1 Monte Carlo simulations

Asuma un juego en la cual un jugador selecciona una ficha de una bolsa llena de fichas. Cada ficha contiene un número del 1 al 100. El jugador apuesta si la ficha a sacar es par o impar. En este juego, las fichas con números pares entre 45 y 55 van a ser contados como impares y los impares entre 45 y 55 como pares. Si se apuesta en que el número a salir es par y sale 50, el jugador pierde. Se necesita encontrar la distribución del monto resultante después de que el jugador realice un número  $n$  de apuestas. Será que el jugador podrá perderlo todo?

## Metodología

- Crear simulador de la selección de la ficha
  - Devuelve verdadero si el jugador gana
- Crear función que simula la apuesta
  - Dinero Total: El jugador comienza con \$5,000
  - Monto de la Apuesta: el monto a apostar en cada apuesta es de \$50
  - Juegos Totales: El número de veces que el jugador juega.
- Crear el ciclo que llame a la función que simule las apuestas
- Visualizar el resultado de los montos resultantes al finalizar un número de apuestas (número de juegos vs monto restante)

## 2. Markov Chain

Usted se encuentra realizando un churn análisis de un producto de video streaming. Usted sabe que el producto en evaluación pierde y gana clientes frecuentemente. Asuma que las probabilidades de transición son constantes para 4 productos considerados incluyendo al que se encuentra bajo estudio. Netflix mantiene una probabilidad de 0.25 de que sus usuarios cancelen el servicio y se marchen a otro. Los usuarios de Disney+ continúan en este servicio con una probabilidad 0.4, algunos usuarios se devuelven a Netflix con una probabilidad de 0.3 y otros pasan al servicio de HBO+ con una probabilidad de 0.1. HBO+ tiene la probabilidad mas alta de perder usuarios para Netflix de 0.5 y de 0.2 para los otros servicios. Los usuarios de Hulu se subscriben a Netflix con una probabilidad de 0.3 y otros pasan a Disney con una probabilidad de 0.15 y de los restantes unos se quedan en este servicio con una probabilidad de 0.25.

### Metodología

- Construir la matriz de probabilidad de transición
- Validar que la matriz se encuentre bien definida
- Encuentre la distribución estacionaria

### 3. Markov Chain

Usted debe realizar un generador de texto y autocompletado para generar la frase final de despedida de correos electrónicos. Usted tiene la posibilidad de trabajar con un conjunto de frases ejemplo para cumplir esta actividad. Al finalizar usted necesita saber cual es la probabilidad de que la frase generada como cierre de correo sea “Thank you”.

#### Metodología

- Utilizar el conjunto de datos de ejemplo para construir la matriz de probabilidad de transición
  - “Thank you for your help”
  - “Thank you”
  - “Thank you for all your help”
  - “Thank you for your assistance”
  - “Thank you for everything”
- Validar que la matriz se encuentre bien definida
- Crear la función de generación de frases usando el modelo de Markov
- Crear el ciclo que llama a la función anterior un numero
- Identifique las frases que se generaron con “Thank you” y calcule su probabilidad

## 4. MCMC

Asuma que necesita realizar muestras de una distribución gamma definida como  $f(x) = \left(\frac{1}{1+e^{\beta t+a}}\right)$ .

### Metodología

- Crear función para encontrar  $f(x)$
- Crear función de MCMC
- Crear ciclo que llame a la función MCMC un numero de x veces para promediar el valor máximo de la función en cada corrida del MCMC
- Visualice la distribución real vs simulada
- Prueba con un valor de  $\beta=-2$  y  $\alpha=-5$  y  $\beta=-2$  y  $\alpha=5$

# Lab-4 MCMC

## Entregable:

- Un archivo de Jupyter Notebook con el desarrollo del análisis
- Las conclusiones y respuestas al objetivo del análisis deben ser contestadas en el mismo notebook.
- Se sugiere que comente las secciones de manera adecuada para una mejor interpretación de su análisis.
- La fecha de entrega es el Sábado 20 de Marzo del 2021 vía catalogo web enlace de laboratorios.