

Minería de Datos

Lab. #4

Ivan Saavedra, Ph.D.

saavedrai@uninorte.edu.co

Universidad del Norte
División de Ingenierías
Dpto. Ingeniería de Sistemas



202030

1 Monte Carlo simulations

Asuma un juego en la cual un jugador realiza lanzamientos de un dado personalizado para obtener un resultado de 1 a 100. Si el jugador obtiene un resultado entre 1-51, la casa gana, pero si el resultado se encuentra entre 52-100, el jugador gana. Se necesita encontrar la distribución del monto resultante después de que el jugador realice un número x de apuestas. ¿Será que el jugador podrá perderlo todo?

Metodología

- Crear simulador del lanzamiento del dado
 - Devuelve verdadero si el jugador gana
- Crear función que simula la apuesta
 - Dinero Total: El jugador comienza con \$10,000
 - Monto de la Apuesta: el monto a apostar en cada apuesta es de \$100
 - 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 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

3. MCMC

Asuma que necesita realizar muestras de una distribución gamma definida como $f(x) = \left(\frac{x^{c-1}}{\Gamma(c)b^c}\right)^{-x}$ en el intervalo de 0 a 20.

Metodología

- Crear función para encontrar $f(x)$
- Crear función de MCMC
 - Utilizar la distribución uniforme para generar los valores de los estados
 - Visualice distribución real vs simulada para una corrida del 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

Análisis de Distribuciones de Probabilidad

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 Viernes 25 de Septiembre del 2020 vía catalogo web enlace de laboratorios.