

Tarea: Fundamentos de probabilidad.

Estimados estudiantes,

Por favor, resuelvan los ejercicios utilizando procesos matemáticos correctos.

Agregar soluciones en formato adjunto y cargar en la tarea correspondiente.

## 1 Clásicos

Ejercicio : Determinar si  $p_1, \dots, p_4$  definen una distribución de probabilidad finita, siendo  $p_1 = a + 0.001, p_2 = a - 0.01, p_3 = a^2 + 0.001, p_4 = 1 + a^2$

Ejercicio : Determinar si  $p_1, \dots, p_4$  definen una distribución de probabilidad finita, siendo  $p_1 = a - 0.001, p_2 = a - 0.01, p_3 = 3a^2 + 0.001, p_4 = 1 - 2a^2$

Ejercicio: Un experimento consiste en lanzar un dado y luego una moneda y luego otro dado. Describa el espacio de todos los posibles resultados, asigne probabilidades uniformemente y calcule (a)  $P(\text{observar cara})$  (b)  $P(\text{observar cara y un lado superior impar del dado})$

Ejercicio: Un experimento consiste en leer dos cartas consecutivamente (de una baraja de 52 cartas agrupadas en 4 bloques de 13 cartas) Describa el espacio de todos los posibles resultados, asigne probabilidades uniformemente y calcule (a)  $P(\text{observar un as})$  (b)  $P(\text{observar un lado superior par del dado y una reina})$

Ejercicio: Un experimento consiste en lanzar un dado y luego leer una carta (de una baraja de 52 cartas agrupadas en 4 bloques de 13 cartas) Describa el espacio de todos los posibles resultados, asigne probabilidades uniformemente y calcule (a)  $P(\text{observar un as})$  (b)  $P(\text{observar un lado superior par del dado y una reina})$

## 2 Probabilidades en modelos de lenguajes

Ejercicio :

Un equipo de investigación está desarrollando un sistema de resumen automático basado en un LLM. Saben que el modelo tiene una probabilidad del 70% de generar un resumen adecuado cuando el texto de entrada tiene más de 5000 palabras, y una probabilidad del 50% cuando el texto tiene menos de 5000 palabras. Si se selecciona aleatoriamente un documento de 10,000 palabras y otro de 3,000 palabras, si se asumen eventos independientes, calcule:

a) La probabilidad de que ambos documentos generen un resumen adecuado.

b) La probabilidad de que al menos uno de los documentos genere un resumen adecuado.

Ejercicio: Una empresa de desarrollo de modelos de lenguaje se consultará si sus empleados están de acuerdo con hacer horas extras con pago doble o no. Los agrupará según sean solteros, estén casados y tengan menos de 2 hijos, y los demás. Describa el espacio de todos los posibles resultados, asigne probabilidades uniformemente y calcule (a)  $P(\text{tener dos hijos o más})$  (b)  $P(\text{estar de acuerdo con las horas extras})$

Ejercicio:

Un equipo de lingüistas está estudiando el uso de palabras raras en los LLM. Saben que la probabilidad de que un LLM genere una palabra rara es del 20% cuando se entrena en un conjunto de datos general, y del 30% cuando se entrena en un conjunto de datos especializado. Si se selecciona aleatoriamente un LLM entrenado en un conjunto de datos general y otro entrenado en un conjunto de datos especializado, si se asumen eventos independientes, calcule:

- a) La probabilidad de que ambos LLM generen una palabra rara.
- b) La probabilidad de que al menos uno de los LLM genere una palabra rara.

Ejercicio:

Un equipo de lingüistas está estudiando el uso de palabras fuera del lenguaje de consulta en los LLM. Saben que la probabilidad de que un LLM genere una palabra fuera del lenguaje es del 25% cuando se entrena en un conjunto de datos general, y del 40% cuando se entrena en un conjunto de datos especializado en matemáticas. Si se selecciona aleatoriamente un LLM entrenado en un conjunto de datos general y otro entrenado en un conjunto de datos especializado, si se asumen eventos independientes, calcule:

- a) La probabilidad de que ambos LLM generen una palabra que esté dentro del lenguaje de consulta.
- b) La probabilidad de que al menos uno de los LLM genere una palabra fuera del lenguaje.