



BOURBAKI

COLEGIO DE MATEMÁTICAS

B O U R B A K I

COLEGIO DE MATEMÁTICAS

Índice

01. Introducción _____ pág. 02

02. Examen final _____ pág. 04

01 Introducción

Bienvenidos a nuestro curso de Matemáticas Avanzadas para la Ciencia de Datos, nuestro curso tiene tres módulos dedicados a estudiar las ideas matemáticas más útiles para comprender los algoritmos y modelos matemáticos más comunes en Ciencia de Datos. Los tres módulos son los siguientes

- Probabilidad y Estadística
- Álgebra Lineal
- Optimización y cálculo diferencial

Todos los módulos tienen una duración de 8 semanas. El curso está acompañado de ejercicios y tareas en Python para practicar y reforzar los conocimientos aprendidos así como las implementaciones en bases de datos de los algoritmos estudiados. Pueden consultar el repositorio de esta semana en [este link](#).

La estructura de cada una de las semanas es la siguiente:

1. Treinta minutos dedicados a estudiar un artículo de referencia que motivará los conceptos matemáticos de esta semana.
2. Dos horas dedicadas a estudiar el tema de la semana y algunos ejercicios.

3. Una hora y media dedicada a practicar lo aprendido utilizando Python.

El primer módulo de probabilidad consta de los siguientes temas:

1. Kolmogorov, independencia y condicionamiento
2. Variables aleatorias discretas y sus momentos
3. Ley de los grandes números y máxima verosimilitud
4. El teorema límite central y los intervalos de confianza
5. Tests estadísticos
6. Método de monte carlo para cadenas de markov
7. Inferencia bayesiana
8. Simulación y redes bayesianas

02 Examen final

Para aprobar el examen es necesario resolver 8 ejercicios de la siguiente lista.

1. Supongamos que tomamos 10 post-its y en cada uno escribimos alguna de las letras de la palabra MATEMÁTICO. Si doblamos cada uno de los post-its y los metemos en una urna, ¿cuál es el espacio de probabilidad asociado a esta experiencia? ¿qué probabilidad hay de obtener la palabra matemático al sacar los diez post-its de manera aleatoria?
2. Generar la ley de probabilidad de las siguientes variables aleatorias:
 - a) El número que puede salir en la ruleta francesa (sólo tienen un cero).
 - b) El color que puede salir en la ruleta francesa.
 - c) Graficarlas
3. Utilizando las probabilidades del ejercicio anterior evaluar si los siguientes eventos son dependientes o independientes
 - a) Obtener un numero impar,
 - b) Que sea de color negro.

4. ■ Calcule la probabilidad condicional de obtener un número impar dado que salió el color negro $\mathbb{P}(A|B)$.

■ Calcule la probabilidad condicional de que el número sea negro, dado que salio un impar $\mathbb{P}(A|B)$.

■ Graficar $\mathbb{P}(-|B)$ en ambos casos.

5. Supongamos que deseamos segmentar a nuestros clientes en cinco clases distintas de acuerdo a tres rangos de compras: 10, 20, 30, 40 y 50. Si tenemos una base de datos de tamaño 1000, definir las variables aleatorias correspondientes utilizando dos distribuciones de nuestros clientes distintas. Calcular la esperanza y la varianza correspondiente.

6. Enunciar en castellano (sin fórmulas) la ley de los grandes números y mencionar dos aplicaciones a la ciencia de datos.

7. Elegir un fenómeno aleatorio, para el que exista conocimiento sobre su distribución de probabilidad basados en estudios previos.

■ Especifique cuál es la variable aleatoria y grafique su espacio de probabilidad.

■ Genere 5 muestra aleatoria con esa distribución, cada una de ellas con tamaño $n=10, n=100, n=1000, n=5000, n=10000$

■ Realice el paso anterior varias veces y describa sus observaciones

8.
 - En el ejercicio anterior genere 500 muestras de tamaño 200
 - Grafique la distribución de al menos 3 muestras
 - Calcule la media de cada una de las 500 muestras y grafique su distribución (distribución de los promedios de la muestra)
 - Describa sus observaciones
9. Con los datos anteriores:
 - Calcule el intervalo de confianza de la décima muestra para un nivel de confianza de 95 porciento
 - Genera una muestra más pero ahora de tamaño 400
 - Calcule el intervalo de confianza para un nivel de confianza de 95 porciento
 - Explique sus observaciones
10. Calcular el test de Kolmogorov-Smirnov para la adecuación con una distribución gaussiana y una laplace de los retornos de la serie del tiempo del oro.
11. ¿En las hipótesis de cuál algoritmo de machine learning no se incluye la hipótesis de independencia? ¿Se incluye la de indécticamente distribuido?
12. Explicar qué significa este prior.



BOURBAKI
COLEGIO DE MATEMÁTICAS

escuela-bourbaki.com