

UNIVERSIDAD POLITÉCNICA DE VICTORIA
MINERÍA DE DATOS APLICADA
EXAMEN UNIDAD 1

IMPORTANTE: REvisa al final las instrucciones de entrega

ALUMNO: _____

PARTE 1: CÁLCULO, ÁLGEBRA LINEAL Y PROBABILIDAD Y ESTADÍSTICA

INSTRUCCIONES: CONTESTA LO QUE SE TE PIDE

1. **(CÁLCULO)** A PARTIR DE LAS SIGUIENTES FUNCIONES:

- a) $f(x, y) = 3x + \text{sen}(xy^2)$
- b) $g(x) = \text{sen}(a^2) + \text{sen}(x)\cos(x)$
- c) $h(x, y, z) = \text{sen}(xyz^2) + \tan(x)\cos(xz)$
- d) $i(a, b) = 2^a + \cos(3b) + \text{sen}(a^b)$

OBTEN LAS SIGUIENTES DERIVADAS

- 1. $\frac{\partial f(x, y)}{\partial x}$
- 2. $\frac{dg(x)}{dx}$
- 3. $\frac{\partial h(x, y, z)}{\partial z}$
- 4. $\frac{\partial i(a, b)}{\partial b}$
- 5. $\nabla h(x, y, z)$ (Vector gradiente de la función h)

2. **(ÁLGEBRA LINEAL)** REALIZA LAS OPERACIONES QUE SE TE PIDEN CONSIDERANDO LAS SIGUIENTES MATRICES:

$$A = \begin{pmatrix} 2 & 3 \\ 5 & -4 \\ 8 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & -1 & -2 \\ 5 & 4 & 3 \\ 3 & 2 & 10 \end{pmatrix} \quad C = \begin{pmatrix} 3 & -1 & 9 \\ 4 & 1 & 2 \end{pmatrix}$$
$$\mathbf{v} = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix} \quad \mathbf{u} = \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$$

- 1. $B * A$ (Multiplicación)
- 2. $C^T * C$ (Multiplicación)
- 3. $\mathbf{v} + 3\mathbf{u}$ (Suma de vectores)
- 4. $\mathbf{w} \times \mathbf{u}$ (Producto Cruz)
- 5. $(\mathbf{w} \times \mathbf{v}) \cdot \mathbf{v}$ (Producto Cruz y después Producto Punto)

2. (ESTADÍSTICA) A PARTIR DE LOS SIGUIENTES DATOS:

13 20 52 4 37 49 12 57 26 53 31 25 39 14 48 64

CALCULA LO SIGUIENTE:

1. MEDIA ARITMÉTICA
2. MEDIANA
3. MODA
4. VARIANZA
5. DESVIACIÓN ESTÁNDAR

PARTE 2: SCRIPT EN PYTHON – PROGRAMAR ALGORITMO KMEANS BÁSICO

INSTRUCCIONES: A PARTIR DEL SCRIPT REALIZADO EN EL PROBLEMARIO, ES DECIR, A PARTIR DE LOS 17 PASOS MENCIONADOS, IMPLEMENTADOS EN PYTHON COMPLETAMENTE FUNCIONALES, SE DEBE DE REALIZAR LO SIGUIENTE:

PASOS SIGUIENTES:

18. UNA VEZ QUE SE TIENEN GRAFICADOS LOS 3 CENTROIDES CON SUS RESPECTIVOS PATRONES (FILAS, INSTANCIAS DE LA MATRIZ Xtrain) MÁS CERCANOS, SE DEBEN DE RECALCULAR LOS VALORES DE LOS CENTROIDES, ESTO SE HACE DE LA SIGUIENTE MANERA:

- SE CALCULA LA MEDIA ARITMÉTICA DE TODOS LOS PATRONES ASOCIADOS AL K-ÉSIMO CENTROIDE, ES DECIR, LA MEDIA (MEAN) DE LOS PATRONES MÁS CERCANOS AL K-ÉSIMO CENTROIDE.
- EL VALOR OBTENIDO DE LA MEDIA ARITMÉTICA SERÁ UN VECTOR EN \mathbb{R}^2 , ESTO YA QUE CADA PATRON ASOCIADO ES UN VECTOR EN \mathbb{R}^2
- ESTE VALOR REPRESENTA EL NUEVO CENTROIDE RECALCULADO

19. REPETIR EL PASO ANTERIOR HASTA QUE LA DIFERENCIA ENTRE LOS CENTROIDES DE UNA ITERACIÓN ANTERIOR (OLD-CENTROIDS, ES DECIR, CENTROIDES VIEJOS) Y LOS CENTROIDES RECALCULADOS EN EL PASO ACTUAL SEA MUY PEQUEÑA, PUEDE SER DIGAMOS DE 0.01. NOTA: LA DIFERENCIA LA OBTIENEN DE CADA DIMENSION, ES DECIR, DE CADA ELEMENTO DEL VECTOR.

20. AL FINAL SE DEBEN VER DOS GRÁFICAS: UNA CON LOS CENTROIDES ORIGINALES OBTENIDOS DE MANERA ALEATORIA Y LOS PATRONES ASOCIADOS A CADA UNO DE ELLOS, Y LA OTRA GRÁFICA CON LOS ÚLTIMOS CENTROIDES OBTENIDOS Y LOS PATRONES ASOCIADOS. RECUERDA QUE DADO QUE SON 3 CENTROIDES, DEBE DE HABER TRES GRUPOS DE PUNTOS (PATRONES) EN DONDE CADA GRUPO TENGA UN COLOR DIFERENTE.

A CONTINUACIÓN SE MUESTRA EL PSEUDOCÓDIGO DEL ALGORITMO K-MEANS, EL CUAL SE REALIZA EN PYTHON SIGUIENDO LAS INSTRUCCIONES (PASOS) QUE LES INDIQUÉ:

1. Initialize cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$ randomly.

2. Repeat until convergence: {

For every i , set

$$c^{(i)} := \arg \min_j \|x^{(i)} - \mu_j\|^2.$$

For each j , set

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}.$$

}

IMPORTANTE:

INSTRUCCIONES DE ENTREGA PARA LA PARTE 1 (CÁLCULO, ÁLGEBRA LINEAL Y ESTADÍSTICA):

- ENTREGAR LA PARTE DE CÁLCULO, ÁLGEBRA LINEAL Y ESTADÍSTICA EL MARTES 19 DE JUNIO, LA HORA LÍMITE ES 4:40 PM (TIENEN LA HORA DE CLASE DEL LUNES 18 PARA CONTINUAR POR SI AÚN NO TERMINAN)
- EN HOJAS BLANCAS TAMAÑO CARTA (HOJA DE MÁQUINA, UTILIZAR AMBOS LADOS, SI ES RECICLADA PUES SOLO USEN EL LADO QUE ESTÁ EN LIMPIO), PUEDE SER A MANO, PERO DEBE SER ENTENDIBLE, SINO NO LO RECIBO. SI DESEAN TRANSCRIBIRLO EN COMPUTADORA ESTÁ BIEN.
- EL DESARROLLO DEBE DE CONSISTIR CONSIDERANDO LO SIGUIENTE: CÁLCULO: ÍNDICA QUÉ REGLAS Y FÓRMULAS DE DERIVACIÓN UTILIZASTE, ÁLGEBRA LINEAL: SÓLO DESARROLLA Y REALIZA LAS OPERACIONES, ESTADÍSTICA: LAS FÓRMULAS QUE UTILIZASTE, LA SUSTITUCIÓN DE LOS RESULTADOS Y EL RESULTADO FINAL.
- OBLIGATORIO PORTADA CON LA SIGUIENTE INFORMACIÓN Y EN EL SIGUIENTE FORMATO:

UNIVERSIDAD POLITÉCNICA DE VICTORIA

ITI

MINERÍA DE DATOS APLICADA

EXAMEN UNIDAD 1 – PARTE 1

ALUMNO: NOMBRE COMPLETO

- ENTREGAR HOJAS GRAPADAS, NO CLIPS, NO CARPETAS, NO ENGARGOLADO, NO NADA, SÓLO GRAPADAS, ÚNICAMENTE GRAPADAS, **GRAPADAS**. (SINO NO SE RECIBE)

INSTRUCCIONES DE ENTREGA PARA LA PARTE 2 (SCRIPT DE PYTHON CON EL ALGORITMO DE K-MEANS):

- EL SCRIPT SE DEBE DE ENVIAR COMPLETAMENTE FUNCIONAL ANTES DE LAS 11:59 PM DEL MARTES 18 DE JUNIO DEL 2018
- **ENVIAR AL CORREO ELECTRÓNICO:** stoviasa@upv.edu.mx
- **ASUNTO: MDD-EXAMEN-SCRIPT-U1-ApellidoPaterno-ApellidoMaterno**
- **NOTA-IMPORTANTÍSIMA:** SOLO ENVIAR EL ARCHIVO QUE CONTIENE EL CÓDIGO. SI EL ASUNTO NO VIENE ASÍ, TAL CUAL, NO LO RECIBO. DEBEN DE RESPETAR EL FORMATO DEL ASUNTO. SOLO RECIBO UNA VERSIÓN (LA PRIMERA QUE ENVÍEN) ASÍ QUE NO SE EQUIVOQUEN AL ENVIAR LA PRIMERA VEZ.

ES TODO, ÉXITO.