

Universidad Panamericana
Facultad de Ingeniería

Gestión y análisis del conocimiento (BDP COM145)
Prof. Mtra. Sarahí Aguilar González

Primer Examen Parcial
Febrero 2022

Nombre completo:

ID:

Calificación:

"Certifico que el trabajo realizado en este examen es estrictamente personal y reconozco que no cumplir las reglas o COMETER ACTOS DESHONESTOS en los exámenes puede resultar en la baja definitiva de la Universidad."

Firma:

Lea cuidadosamente y responda acordemente.

1. Describe las etapas del ciclo de vida de un proyecto de ciencia de datos.
Valor: 2 puntos
2. Describe la importancia de los ambientes de conda.
Valor: 1 punto
3. Describe las ventajas de utilizar estructuras de datos de NumPy sobre listas estándar de Python.
Valor: 1 punto
4. Utilizando NumPy, desarrolle y resuelva lo siguiente en un Jupyter notebook de Google Colaboratory.
Valor: 3 puntos

En estadística, el error cuadrático medio (ECM) de un estimador (un procedimiento para estimar un valor) mide el promedio de los errores al cuadrado, es decir, la diferencia cuadrática promedio entre los valores estimados y los valores reales.

El ECM puede ser expresado como:

$$\text{ECM} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2.$$

Donde:

- n es el total de valores reales
- \hat{Y} son los valores estimados
- Y son los valores reales

Ejemplo del cálculo de \hat{Y} y del ECM considerando un estimador lineal ($y = a \cdot x + b$):

$a = 2$

$b = 1$

$X = [1, 2, 3]$

$Y = [3, 5, 10]$

$n = \text{total de valores reales} = 3$

Calculando \hat{Y} ...

$$\hat{y}_1 = a \cdot x_1 + b = 2 \cdot 1 + 1 = 3$$

$$\hat{y}_2 = a \cdot x_2 + b = 2 \cdot 2 + 1 = 5$$

$$\hat{y}_3 = a \cdot x_3 + b = 2 \cdot 3 + 1 = 7$$

Calculando el ECM...

$$\text{ECM} = ((\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2 + (\hat{y}_3 - y_3)^2) / n = ((3 - 3)^2 + (5 - 5)^2 + (7 - 10)^2) / 3 = 3$$

Considerando un estimador lineal ($y = a \cdot x + b$), escriba una función que:

- reciba como entrada:
 - el valor a
 - el valor b
 - un arreglo X (compuesto por X_1, X_2, \dots, X_n)
 - un arreglo Y (compuesto por Y_1, Y_2, \dots, Y_n)
- calcule \hat{Y}
- calcule el ECM
- devuelva como salida el ECM

5. Utilizando NumPy y/o Pandas, desarrolle y resuelva lo siguiente en un Jupyter notebook de Google Colaboratory.

Valor: 4 puntos

Responda las siguientes tres preguntas que la tienda departamental Bristol tiene sobre sus ventas hasta la fecha:

1. ¿Cuál ha sido la cantidad máxima de unidades vendidas por un producto?
2. ¿Qué porcentaje de los productos disponibles a la venta han vendido más de \$1,000,000?
3. ¿Cuál es la cantidad de unidades vendidas por producto mínima, promedio y máxima de cada categoría?

Para responder las preguntas anteriores, le han compartido dos conjuntos de datos y el modelo entidad-relación correspondiente:

1. ventas: ventas por producto hasta la fecha de la tienda departamental Bristol
Disponible para su descarga en:
<https://raw.githubusercontent.com/sarahaguilar/fundamentos-cdd/main/datos/ventas.csv>
2. catalogo: productos disponibles a la venta en la tienda departamental Bristol
Disponible para su descarga en:
<https://raw.githubusercontent.com/sarahaguilar/fundamentos-cdd/main/datos/catalogo.csv>

