Logotipo, nombre de la empresa

Descripción generada automáticamente

Universidad de Montevideo – Introducción a la Ciencia de Datos

Examen 2022 – S2

22 de noviembre, 2022

1. Suponga que usted invierte en acciones de Apple. Para saber cuándo es el momento óptimo de vender la acción usted extrae datos históricos de como la acción de Apple ha cambiado en el tiempo (precio en el correr del tiempo) y decide entrenar un modelo. Cómo se debe realizar la validación del modelo?
2. Cuáles son los problemas asociados a los modelos de Bag of Words?
3. En marco de la regresión lineal:
   * Qué significado tiene el parámetro R2?
   * Entre que valores está definido ?
   * Cómo se calcula y a partir de qué componentes?
4. En cuanto a la correlación de Pearson
   * Cómo se calcula?
   * Qué significado tiene su coeficiente?
   * Cómo se determina la significancia del valor del coeficiente?
   * En qué se diferencia la correlación de Pearson de la de Spearman?
5. Cuáles son los principales componentes de una serie de tiempo? Describa brevemente cada una de ellas.
6. En la siguiente imagen se muestran las predicciones de 2 algoritmos de clustering para 6 datasets distintos. Cada fila representa las predicciones de un algoritmo. Estos algoritmos son DBSCAN y K-Means. Qué fila corresponde a qué algoritmo? Justifique su respuesta.

Chart, scatter chart

Description automatically generated

1. Usted posee un dataset con 100.000 registros, de los cuales 90.000 pertenecen a la clase negativa y=0 y 10.000 a la clase positiva y=1. Cómo debería usted partir los datos para hacer una correcta validación?
2. Explique 4 aplicaciones de Procesamiento de Lenguaje Natural no vistas en clase.
3. Explique qué representan y cómo se calculan los siguientes indicadores. A qué tipos de variables (nominales, ordinales, intervalo o ratio) se puede aplicar?

* Media
* Moda
* Media Ponderada
* Mediana

1. A la hora de utilizar una regresión lineal:
   * Que diferencias hay entre extrapolar e interpolar? Cuál de estas genera valores con mayor incertidumbre?
   * Qué consideraciones deberíamos de tener a la hora de utilizar una regresión lineal para predecir un valor futuro?