

Prueba técnica para Backend Jr.

Parte 2: Cuestionario Teórico

Instrucciones

Responde las siguientes preguntas de manera detallada, explicando los conceptos y dando ejemplos cuando sea necesario.

1. Machine Learning:

- ¿Qué es el overfitting y cómo se puede evitar en un modelo de machine learning?

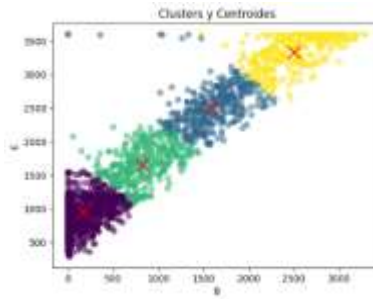
Quando entrenamos un modelo de machine learning tenemos que dividir nuestro conjunto de datos en dos partes: el conjunto de entrenamiento y el conjunto de pruebas. El conjunto de entrenamiento se utiliza para ajustar el modelo, mientras que el conjunto de pruebas se utiliza para evaluar su rendimiento. Si el modelo tiene un rendimiento más elevado en el conjunto de entrenamiento que en el conjunto de pruebas, lo más probable es que esté sufriendo de overfitting

- Explica la diferencia entre aprendizaje supervisado y no supervisado. Proporciona ejemplos de algoritmos para cada uno.

La principal diferencia esta en los datos, ya que de un lado el aprendizaje supervisado tiene un “feedback” de los datos que esta procesando. Se le dice supervisada porque el programador es el en cargado de decirle por cual camino debería de ir. Mientras que el aprendizaje no supervisado es todo lo contrario, el modelo no tiene un camino al cual lo podemos guiar, en su lugar, tiene que identificar patrones y estructuras en los datos por sí solo.

Utilice el algoritmo de Gradient Boosting, el cual es una técnica de aprendizaje supervisado, se basa en crear un modelo fuerte a partir de la combinación de varios modelos débiles.

También uno de los algoritmos de aprendizaje no supervisado más interesantes que aprendí, fue un algoritmo de agrupamiento llamado K-means. Lo que hace es que asigna cada punto de datos al cluster con la media más cercana. Adjunto imagen de como quedan los datos clasificados con este método.



- ¿Qué es la validación cruzada y por qué es importante en el entrenamiento de modelos?

Son métodos que evalúan el rendimiento de un modelo de machine learning con el objetivo de asegurarse de que el modelo no esté sobreajustando o subajustado. Es muy importante, ya que ayuda a encontrar los mejores hiperparámetros para el modelo. Seleccionar los hiperparámetros es fundamental para el rendimiento del modelo y en lo personal lo más tedioso.

2. Data Science:

- ¿Cuál es la importancia de la limpieza de datos en un proyecto de data science?

La limpieza de los datos determina que tan bueno será el modelo, porque es con lo que aprende la máquina. Si nosotros no hacemos la limpieza de datos las predicciones saldrán mal.

- Explica cómo seleccionar características relevantes para un modelo predictivo.

En primer lugar, debemos de identificar dos tipos de variables en nuestro dataset. Las características categóricas y continuas. Una vez que hayamos identificado cada una identificamos las variables que más contribuyen a la predicción y les damos más importancia que las variables irrelevantes. Existen varias técnicas para seleccionar las características relevantes. Una de ellas es encontrar la matriz de correlación, útil en la identificación de características redundantes.

- ¿Qué son las medidas de tendencia central y cómo se utilizan en el análisis de datos?

Este es un tema estadístico. Las medidas explican cuál es el centro de un conjunto de datos y son fundamentales en el análisis de datos. Porque con algo tan simple como la media o la mediana, nos da información valiosa para saber qué hacer con el conjunto de datos.

3. Backend Development:

- ¿Qué es una API RESTful y cuáles son sus principios básicos?

Una API es un conjunto de protocolos que permite a diferentes servicios comunicarse entre sí. Ahora una API RESTful es tipo de API que sigue una arquitectura REST (Representational State Transfer). El cual es una forma de diseñar servicios web que permite que los sistemas se comuniquen de manera simple y eficiente usando HTTP.

Así como cualquier API, debe de seguir un modelo de cliente-servidor. Cada solicitud del cliente al servidor debe contener toda la información necesaria para entender y procesar la solicitud.

- Explica la diferencia entre métodos HTTP GET, POST, PUT y DELETE. Da ejemplos de uso para cada uno.

Estos métodos son los que logran la comunicación entre clientes y servidores, cada uno son muy intuitivos y se utilizan en diferentes circunstancias. Por ejemplo el método GET se utiliza para solicitar datos de un servidor, POST se utiliza para enviar datos al servidor, PUT se utiliza para actualizar algún recurso del servidor y finalmente DELETE elimina un recurso del servidor.

- ¿Cómo manejarías la paginación de datos en una API RESTful?

4. MySQL:

- ¿Qué es una clave primaria y cuál es su función en una base de datos relacional?

Dependiendo del proyecto puede ser una o mas columnas en una tabla que se utiliza para identificar de manera única cada registro en la tabla. Una de las funciones mas importantes es garantizar la integridad de los datos y especialmente en bases de datos grandes ayuda a realizar queries más eficientes.

- Explica qué es una relación uno a muchos y proporciona un ejemplo de su implementación en MySQL.

Una relación uno a muchos es cuando un registro en una tabla puede estar relacionado con múltiples registros en otra tabla, pero no viceversa.

- ¿Cómo se puede optimizar una consulta SQL para mejorar su rendimiento?

Existen diferentes técnicas que podemos utilizar, aunque personalmente nunca me he enfrentado a un problema de optimización en una consulta. Utilizar la estrategia de “divide and conquer” particionando las tablas. Sin embargo, cualquier técnica que ocupemos debemos de tomar en cuenta el contexto particular de la base de datos.