

- En el mapeo de valores categóricos, ¿pueden los modelos multiplicar cadenas directamente por los pesos aprendidos?

1 / 1 punto

- ☒ No
☐ Sí

✓ **Correcto**

¡Correcto! La ingeniería de características debe aplicarse con anticipación para convertir cadenas en valores numéricos.

2. You are working on a taxi tip prediction model and your raw dataset has columns for the latitude and longitude of both pickup and dropoff locations. These do not assume a Gaussian distribution. Among the choices below, which two approaches are more likely to make your model learn better from these features? Assume that you are just starting the feature engineering process. (Select two answers)

1 / 1 punto

- ☐ Because the data does not assume a Gaussian distribution, you should normalize these location features following the formula:

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$$

This puts the values into the range [0, 1] so it can help the training converge faster.

- ☒ Bucketize the locations into discrete bins then do a feature cross of the latitude and longitude.

✓ **Correcto**

¡Correcto! Es posible que dividir en cubos y luego cruzar las características mejore la calidad predictiva de los datos. Por ejemplo, un cruce de *latitud x longitud* de las ubicaciones de recolección puede permitir que el modelo aprenda qué combinación de estas características produce una propina más grande.

- ☐ No considere estas características porque notó que los datos sin procesar tienen una columna para la distancia del viaje en kilómetros. Tu intuición dice que esto tiene un mayor impacto en el monto de la propina en comparación con las ubicaciones.

- ☒ Organice las ubicaciones en contenedores discretos.

✓ **Correcto**

¡Correcto! La agrupación en cubos de las coordenadas agrupará las proximidades cercanas. Con esta técnica, el modelo puede aprender qué áreas es probable que den una propina mayor.

3. ¿Cuál de los siguientes debe implementar al servir su modelo para garantizar su rendimiento?

1 / 1 punto

- ☐ Aumente la calidad predictiva de sus datos a través de técnicas de ingeniería de características como escalado, normalización, estandarización, segmentación, etc.
- ☐ Asegúrese de que cada solicitud se procese en tiempo real.
- ☒ Asegúrese de que todas las transformaciones de datos sean iguales en cualquier escenario
- ☐ Use la menor cantidad de funciones posible para ahorrar recursos informáticos



Correcto

¡Buen trabajo! Las mismas transformaciones que hizo durante la ingeniería de funciones se deben aplicar a la entrada del usuario al servir su modelo. **Esto es para evitar sesgos de servicio de entrenamiento.** Por ejemplo, si una función se normaliza durante el entrenamiento, la entrada de servicio también debe normalizarse. De lo contrario, el modelo podría comportarse de forma inesperada.