

¿Qué fórmula representa un *cambio de conjunto de datos*?

1 / 1 punto

- ☒ $PAGS_{tren_}(y, x) = PAGS_{servir_}(y, x)$
- ☐ $P_{entrenar}(y|x) \neq P_{servir}(y|x)$ y $P_{entrenar}(x) = P_{servir}(x)$
- ☐ $P_{entrenar}(y|x) = P_{servir}(y|x)$ y $P_{entrenar}(x) \neq P_{servir}(x)$



Correcto

¡Bien hecho! El caso más genérico de sesgos de distribución es cuando la distribución conjunta de insumos y productos difiere entre capacitación y servicio.

2. ¿Qué medida se utiliza normalmente para determinar el grado de *desviación de los datos* ?

1 / 1 punto

- ☒ Distancia de Chebyshev (L-infinito)
- ☐ Distancia euclidiana (L2)
- ☐ Distancia Manhattan (L1)
- ☐ distancia de hamming



Correcto

¡Eso es todo! La distancia de Chebyshev se define como $\max \lim_{i \rightarrow \infty} (|x_i - y_i|)$

3. **Distribution skew** occurs when the distribution of the training dataset is significantly different from the distribution of the serving dataset, and is typically caused by: (check all that apply).

1 / 1 punto

- ☒ Different data sources for training and serving data.



Correcto

Way to go! Data sources between training and serving often change and so this is another case of distribution skew.

- ☐ A data source that provides some feature values is modified between training and serving time.
- ☒ Faulty sampling method that selects a sample for training which is not representative of serving data distribution.

✓ **Correcto**

Spot on! A faulty sampling mechanism that chooses a non-representative subsample is an example of distribution skew.

- ☒ Trend, seasonality, changes in data over time.

✓ **Correcto**

Keep it up! Data distributions between training and serving often change and so this is another case of distribution skew.

- ☐ There is different logic for generating features between training and serving. For example, if you apply some transformation only in one of the two code paths.
- ☐ Occurs when serving and training data don't conform to the same schema. For example, int32 != float.

4. TensorFlow Data Validation (TFDV) helps TFX users maintain the health of their ML pipelines. TFDV can analyze training and serves data to:

1 / 1 punto

- ☐ Perform feature selection.
- ☒ Detect data anomalies.

✓ **Correcto**

That's the way! TFDV can check your data for error in the aggregate across an entire dataset or by checking for errors on a per-example basis.

- ☐ Perform feature engineering.
- ☒ Compute descriptive statistics.

✓ **Correcto**

¡Perfecto! TFDV va más allá de calcular estadísticas relevantes, también tiene buenas herramientas de visualización basadas en navegador.

- ☐ Implemente la canalización en una aplicación móvil.
- ☒ Inferir un esquema.



Correcto

¡Bien hecho! En resumen, los esquemas describen las expectativas de datos "correctos" y, por lo tanto, pueden usarse para detectar errores en los datos.