

1. Los métodos de interpretación se pueden agrupar según (marque todo lo que corresponda):

1 / 1 punto

☐ Ya sean complejos o básicos.

☒ Ya sean locales o globales.

✓ **Correcto**

Tu lo tienes. También se pueden agrupar según generen interpretaciones locales o globales.

☒ Ya sean intrínsecos o post-hoc.

✓ **Correcto**

¡Buen trabajo! Una forma de agrupar los métodos de interpretabilidad es si el modelo es intrínsecamente interpretable o si se trata como una caja negra y se utilizan herramientas externas para analizarlo.

☒ Ya sean específicos del modelo o independientes del modelo.

✓ **Correcto**

¡Avanza! Los métodos específicos del modelo están limitados a ciertos tipos de modelos, mientras que los métodos agnósticos del modelo se aplican a cualquier modelo después de entrenarlo.

2. Un aspecto clave que ayuda a mejorar la interpretabilidad es la presencia de características monótonas.

1 / 1 punto

☐ FALSO

☒ Verdadero

✓ **Correcto**

¡Así es! Esto coincide con nuestro conocimiento de dominio para muchas características en muchos tipos de problemas. Cuando tratamos de comprender el resultado de un modelo, si las características son monótonas, coincide con nuestra intuición sobre la realidad del mundo que estamos tratando de modelar.

3. Muchos modelos clásicos son modelos intrínsecamente interpretables, como las redes neuronales transparentes, intuitivas y relativamente fáciles de entender.

1 / 1 punto

☒ FALSO

☐ Verdadero

✓ **Correcto**

¡Absolutamente correcto! La compleja arquitectura de las NN las convierte en "cajas negras" cuando intentas interpretarlas.