**1**

**Dado un modelo que tiene alta performance en el dataset de entrenamiento y baja performance en el dataset de test ¿qué podemos decir?**

El modelo tiene bajo sesgo y baja varianza

El modelo tiene alto sesgo y alta varianza

El modelo tiene overfitting

El modelo tiene overfitting en algunos atributos

**2**

**¿Qué representan los residuos?**

La diferencia entre los valores reales de Y y la media de Y

La diferencia entre los valores de Y reales y los valores de Y predichos

La raíz cuadrada de la pendiente

El valor predicho de Y para el valor medio de X

**3**

**¿Cómo se puede medir la fuerza de la relación lineal entre dos variables numéricas?**

Mediante un diagrama de dispersión

Mediante el coeficiente de correlación

Determinando la pendiente

Determinando la intersección con Y

**4**

**Suponiendo una relación lineal entre X e Y, ¿cuál de las siguientes afirmaciones es verdadera si el coeficiente de correlación es igual a -0.30?**

No existe correlación

La pendiente ?1 es negativa

La variable X es más grande que la variable Y

La varianza de X es negativa

**5**

**Se determina que un intervalo de confianza del 95% para ß1 (beta 1) es [15, 30]. Interprete el significado de este intervalo.**

Podemos tener un 95% de confianza en que el valor medio de Y estará entre 15 y 30 unidades.

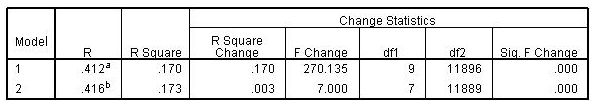
Podemos tener un 95% de confianza en que el valor X aumentará entre 15 y 30 unidades por cada aumento de una unidad en Y.

Podemos tener un 95% de confianza en que el valor promedio de Y aumentará entre 15 y 30 unidades por cada aumento de una unidad en X

Al nivel de significancia del 5%, no hay evidencia de una relación lineal entre Y y X.

**6**

**Se completó un análisis de regresión en dos partes, la segunda de los cuales agregó términos de interacción a las variables explicativas incluidas en el primer modelo. ¿La adición de los términos de interacción mejoró la cantidad de varianza explicada en un grado estadísticamente significativo?**



No podemos decidir a partir de estos datos

Verdadero

Falso

**7**

**Los datos de un registro de ocupación de padres se divide en 9 categorías nominales. Parece una buena variable explicativa para un análisis de regresión, por lo que se decide dividirla en variables ficticias. En este caso ¿cuántas variables dummies se necesitará crear?**

Nueve

Diez

Ocho

Ninguna de las anteriores

**8**

**Se realiza una regresión lineal múltiple con dos variables explicativas, que explica el 70% de la varianza total en el resultado. La variable A representa de forma única el 30% de la varianza total y la variable B el 25% de la varianza total. ¿Qué explica el 15% restante de la varianza que se ha explicado?**

Una tercera variable predictora, la Variable C

La varianza compartida entre la Variable A y la Variable B

El error en la medición de la Variable A y la Variable B

**9**

**Para la regresión de ridge, si el parámetro de regularización = 0, ¿qué significa?**

Los coeficientes grandes no se penalizan

No se tienen en cuenta los problemas de sobreajuste

La función de pérdida es la misma que la función de pérdida de mínimos cuadrados ordinaria

Todas las anteriores

**10**

**Para la regresión de lasso, si el parámetro de regularización es muy alto, ¿qué opciones son verdaderas? (Seleccione dos opciones)**

Puede usarse para seleccionar features importantes de un conjunto de datos

Reduce los coeficientes de las features menos importantes a exactamente 0

La función de pérdida es la misma que la función de pérdida de mínimos cuadrados ordinaria

La función de pérdida es la misma que la función de pérdida de regresión de ridge

**11**

**¿Cuál de las siguientes afirmaciones es verdadera?**

La regresión de Ridge y Lasso son técnicas para reducir la complejidad del modelo y evitar el sobreajuste que puede resultar de una regresión lineal simple

La regresión de ridge reduce los coeficientes y ayuda a reducir la complejidad del modelo y la multicolinealidad.

La regresión de lasso no solo ayuda a reducir el sobreajuste, sino que también puede ayudarnos en la selección de features

Todas las anteriores

**12**

**Supongamos que hemos ajustado un modelo de regresión complejo en un conjunto de datos. Ahora, estamos utilizando la regresión de Ridge con el parámetro de ajuste lambda para reducir su complejidad. Elija las opciones a continuación que describen la relación de sesgo y varianza con lambda.**

En caso de lambda muy pequeña; el sesgo es bajo, la varianza es baja

En caso de lambda muy pequeña; el sesgo es bajo, la varianza es alta

En caso de lambda muy pequeña; el sesgo es alto, la varianza es baja

En caso de lambda muy pequeña; el sesgo es alto, la varianza es alta