

Padrón: _____ Nombre y Apellido: _____

Ejercicios V o F (Tipo Parcial)

Ejercicio 1

Determinar si las siguientes afirmaciones son verdaderas o falsas:

- La razón de la falta de datos (missing values) siempre es ajena a los datos en sí mismos. **F**
- La estandarización min-max lleva los valores de una variable al intervalo $[0,1]$. **V**
- La estandarización z-score lleva los valores de una variable al intervalo $[0,1]$. **F**
- Cuando se imputan datos faltantes por Media o Mediana la varianza estimada de la nueva variable no es válida porque está atenuada por los valores repetidos. **V**

Ejercicio 2

Determinar cuáles de las siguientes afirmaciones son Verdaderas o Falsas

- La varianza de un estimador es cuánto varía la predicción según los datos que utilicemos para testear. **F**
- Un bias muy alto significa que el modelo no se ha ajustado lo suficiente a los datos de entrenamiento. **V**
- Un modelo con varianza baja indica que pequeños cambios en los datos de entrenamiento producen cambios pequeños en la estimación. **V**
- Un bias muy alto suele ser underfitting por lo tanto se produce un error alto sólo en las muestras de test. **F**

Ejercicio 3

Determinar si las siguientes afirmaciones sobre los árboles de decisión son verdaderas o falsas:

- El atributo que se utiliza en cada nodo para repartir las instancias es aquel que más aumenta la impureza. **F**
- ID3 es un algoritmo superior a C4.5 ya que permite trabajar con campos numéricos continuos. **F**
- El criterio de impureza de Gini es numéricamente similar al criterio de la entropía. **V**
- Cuando el árbol es demasiado profundo se evita el overfitting pero puede ser muy costoso computacionalmente, por ello suelen podarse. **F**

Padrón: _____ Nombre y Apellido: _____

Ejercicio 4

Determinar si las siguientes afirmaciones son verdaderas o falsas:

- XGBoost sólo puede usarse en problemas de clasificación. **F**
- En Random Forest la votación reduce la varianza de la clasificación. **V**
- XGBoost maneja mejor el overfitting mediante regularizaciones. **V**
- Random Forest es muy útil para la selección de features. **V**

Ejercicio 5

Determinar si las siguientes afirmaciones son verdaderas o falsas:

- Las redes neuronales se optimizan mediante descenso de gradiente. **V**
- Las redes neuronales son sensibles a la escala de los datos. **V**
- Un perceptrón simple no puede modelar compuertas NAND. **F**
- Las redes neuronales muy complejas pueden sub-ajustar, esto se soluciona con métodos de regularización como Early stopping. **F**

Ejercicio 6

Una de las primeras tareas cuando se analiza un conjunto de datos es evaluar si existen valores atípicos. En caso de encontrar algunos indique cuáles de las siguientes acciones serían correctas:

- Eliminarlos siempre del dataset para no interferir en el análisis. **F**
- Tratarlos con algunas de las técnicas utilizadas en datos faltantes y corregir su valor. **F**
- No se debe accionar sobre los valores atípicos, se deben conservar tal como están. **F**
- Analizarlos para identificar la fuente de los mismos antes de decidir si se eliminan o no. **V**

Ejercicio 7

Indique cuáles de las siguientes afirmaciones son verdaderas:

- Los árboles de decisión se utilizan únicamente para problemas de clasificación. **F**
- La entropía es una medida de la impureza o aleatoriedad en los datos. En los árboles de decisión, buscamos que los nodos tengan la mayor entropía posible. **F**
- En árboles de decisión hay sobreajuste cuando el árbol es “demasiado” profundo. Una solución es detener el crecimiento del árbol antes que clasifique perfectamente a los datos. **V**
- En árboles de decisión hay sobreajuste cuando el árbol es “demasiado” profundo. Una solución es hacer crecer el árbol entero para que clasifique perfectamente a los datos y luego podar. **V**