

## Examen para SAA05

### Intento 1

#### Pregunta 1

¿Cómo se calcula el “recall” de un modelo entrenado?

- a. dividiendo el número de casos de una clase identificados correctamente por todos los casos de esa clase identificados por el modelo.
- b. Ninguna de las anteriores.
- c. Dividiendo el número de casos correctamente clasificados de todas las clases por el número total de casos.
- d. **dividiendo el número de casos de una clase identificados correctamente por todos los casos reales de dicha clase.**

#### Pregunta 2

La técnica de submuestreo para casos de clasificación no balanceada consiste en

- a. Ninguna de las anteriores.
- b. añadir errores a propósito en la clase mayoritaria para confundir al modelo.
- c. añadir casos a la clase minoritaria para igualar el número de casos entre clases.
- d. **eliminar casos de la clase mayoritaria para igualar el número de casos entre clases.**

#### Pregunta 3

En la predicción de un modelo de aprendizaje automático, ¿qué es un falso positivo?

- a. Es un caso que se predice como negativo pero en realidad es positivo.
- b. Es un caso que se predice como positivo y realmente es positivo.
- c. **Es un caso que se predice como positivo pero en realidad es negativo**
- d. Es un caso que se predice como negativo y realmente es negativo.

#### Pregunta 4

¿Cómo se calcula el “accuracy” de un modelo entrenado?

- a. dividiendo el número de casos de una clase identificados correctamente por todos los casos de esa clase identificados por el modelo.
- b. Ninguna de las anteriores.
- c. **Dividiendo el número de casos correctamente clasificados de todas las clases por el número total de casos.**
- d. dividiendo el número de casos de una clase identificados correctamente por todos los casos reales de dicha clase.

#### Pregunta 5

La optimización de hiper-parámetros se realiza normalmente mediante la utilización de un proceso de búsqueda cuyo objetivo consiste en encontrar la mejor selección de valores para un conjunto finito de hiper-parámetros con el objetivo de generar el mejor modelo posible.

Seleccione una:

**Verdadero**

Falso

#### Pregunta 6

**En la predicción de un modelo de aprendizaje automático, ¿qué es un falso negativo?**

- a. **Es un caso que se predice como negativo pero en realidad es positivo.**
- b. Es un caso que se predice como positivo y realmente es positivo.
- c. Es un caso que se predice como positivo pero en realidad es negativo
- d. Es un caso que se predice como negativo y realmente es negativo.

#### Pregunta 7

**A la hora de buscar los hiperparámetros óptimos para tener un buen modelo...**

- a. se busca que el proceso sea muy rápido.
- b. Ninguna de las anteriores.
- c. **se suele utilizar la métrica de accuracy para guiar el proceso.**
- d. Se suele hacer a ciegas.

#### Pregunta 8

**¿Qué es la validación cruzada a la hora de evaluar un modelo?**

- a. Consiste en calcular el error de los datos de entrenamiento en vez de utilizar los de test.
- b. Consiste en cruzar los datos que tenemos con los de otro compañero y calcular cada uno el error del otro.
- c. **Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones del conjunto de datos.**
- d. Consiste en calcular la métrica de evaluación y buscar su valor simétrico respecto del eje horizontal.

#### Pregunta 9

**¿Qué métrica de evaluación de modelos de regresión penaliza los errores grandes?**

- a. **El error cuadrático medio.**
- b. El error medio absoluto.
- c. El accuracy
- d. El recall.

#### Pregunta 10

**En la técnica Random Search para buscar los hiperparámetros óptimos**

- a. **se modifican aleatoriamente las soluciones previamente generadas en el espacio de búsqueda.**
- b. se utilizan los hiperparámetros recomendados por el diseñador del modelo.
- c. se utilizan valores de hiperparámetros distribuidos en el espacio de búsqueda siguiendo una distribución con forma de red.
- d. Ninguna de las anteriores.

## Intento 2

### Pregunta 1

Quando tenemos un dataset con desequilibrio, suele ocurrir que obtenemos un alto valor de precisión en la clase mayoritaria y un bajo recall en la clase minoritaria.

Seleccione una:

**Verdadero**

Falso

### Pregunta 2

¿Cómo se calcula el valor de la métrica “precision” de un modelo entrenado?

- a. Dividiendo el número de casos correctamente clasificados de todas las clases por el número total de casos.
- b. dividiendo el número de casos de una clase identificados correctamente por todos los casos de esa clase identificados por el modelo.**
- c. Ninguna de las anteriores.
- d. dividiendo el número de casos de una clase identificados correctamente por todos los casos reales de dicha clase.

### Pregunta 3

En los modelos de aprendizaje automático para casos de regresión, hay que considerar los datos anómalos y si queremos penalizar errores grandes o no.

Seleccione una:

**Verdadero**

Falso

### Pregunta 4

En la detección de anomalías en conjuntos de datos, tenemos que

- a. Las anomalías siempre van a ser errores que se han cometido al recoger los datos mal.
- b. Las anomalías pueden representar casos o eventos sospechosos que requieren atención.**
- c. No se puede saber qué es un caso normal y qué es una anomalía.
- d. Lo mejor es eliminar las anomalías del dataset.

### Pregunta 5

La técnica de sobremuestreo para casos de clasificación no balanceada consiste en

- a. añadir errores a propósito en la clase mayoritaria para confundir al modelo.
- b. eliminar casos de la clase mayoritaria para igualar el número de casos entre clases.
- c. Ninguna de las anteriores.
- d. añadir casos a la clase minoritaria para igualar el número de casos entre clases.**

### Pregunta 6

¿Qué algoritmo de detección de anomalías se basa en que éstas son fáciles de separar?

- a. Redes neuronales profundas.
- b. Isolation Forest**
- c. Regresión logística.
- d. KNN

### Pregunta 7

En la validación cruzada aleatoria, el bloque de validación se escoge:

- a. buscando el caso con el valor más alto de entropía.
- b. seleccionando el último caso entrenado.
- c. **de forma aleatoria.**
- d. Ninguna de las anteriores.

### Pregunta 8

En un problema de clasificación ¿Para qué sirve la matriz de confusión?

- a. Para saber qué operaciones matemáticas aplicar para corregir el modelo.
- b. Ninguna de las anteriores.
- c. **Para saber si los errores se dan más prediciendo una clase o la otra.**
- d. Para confundir al modelo y que salga más resiliente.

### Pregunta 9

**Aplicar técnicas de tipo “ensemble” ayuda a mejorar la precisión de un modelo, pero debe ser un modelo con los datos muy bien balanceados.**

Seleccione una:

Verdadero

**Falso**

### Pregunta 10

**La principal desventaja de la validación cruzada dejando uno fuera es**

- a. que hay riesgo de perder información.
- b. que es muy probable que haya overfitting.
- c. que se van perdiendo datos a lo largo del proceso.
- d. **que es un proceso muy costoso computacionalmente.**