

Padrón: \_\_\_\_\_ Nombre y Apellido: \_\_\_\_\_ Hojas: \_\_\_\_\_

---

## Recuperatorio Parcial 23-11-2023

### Ejercicio 1

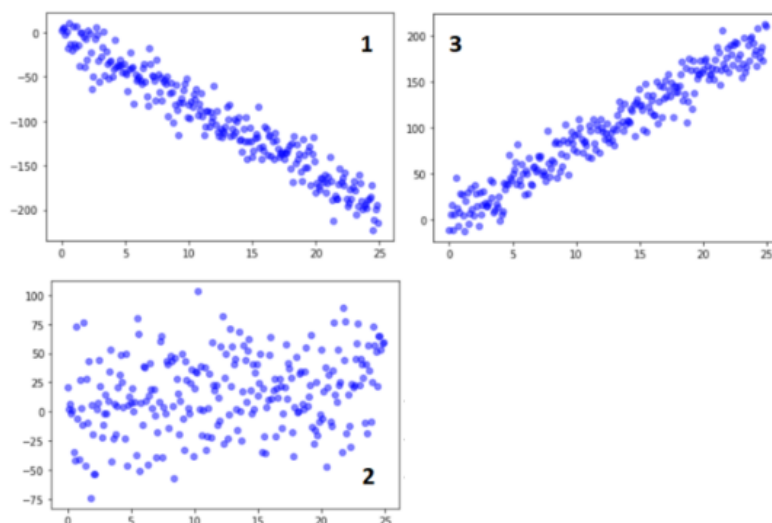
Seleccione una **técnica de preprocesamiento** aplicada en el **Checkpoint 1 del TP1** y describa detalladamente cómo se implementó dando ejemplos concretos del dataset utilizado.

#### EJEMPLO CONCRETO DE APLICACIÓN AL TP1

Detección y corrección de valores faltantes, imputación, encoding, generación de features, etc

### Ejercicio 2

Dados los siguientes **coeficientes de correlación de pearson**, para dos variables X e Y, realizar el gráfico de dispersión corresponde cada uno: a)  $r = 0.80$  b)  $r = -0.86$  c)  $r = 0.35$   
¿Qué puede mencionar de la relación de las variables en cada caso?



- 1) Corresponde a :  $r = -0.86$  se observa una asociación lineal negativa fuerte.
- 2) Corresponde a :  $r = 0.35$  dada la dispersión de los puntos se observa ausencia de correlación/ asociación lineal entre las variables.
- 3) Corresponde a :  $r = 0.80$  se observa una asociación lineal positiva fuerte.

### Ejercicio 3

Explique brevemente en qué consiste y para qué se utiliza el método **Early Stopping** en Redes Neuronales.

Es una técnica de regularización para Redes Neuronales que consiste en detener el entrenamiento de la red cuándo el error sobre el set de validación comienza a aumentar.

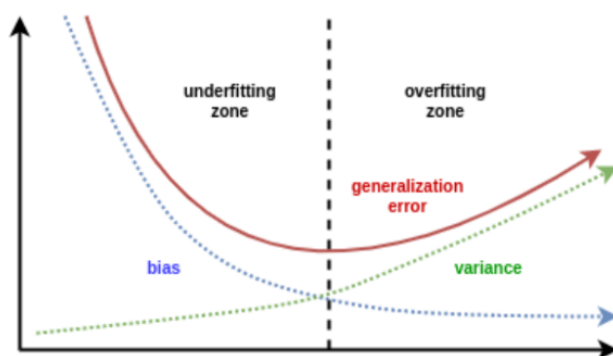
Padrón: \_\_\_\_\_ Nombre y Apellido: \_\_\_\_\_ Hojas: \_\_\_\_\_

Este método busca quedarse con los pesos en la instancia óptima. Se utiliza para evitar el **overfitting** (sobreajuste) ya que ayudan a una mejor generalización, es decir, que el modelo funcione adecuadamente en datos que nunca vió.

#### Ejercicio 4

Explique los siguientes conceptos y cómo se relacionan con la **complejidad** de un modelo de aprendizaje automático: a) Error de Sesgo (**Bias**) b) Overfitting (**sobreajuste**)

- a) **Bias: modelos menos complejos tienen mayor sesgo.**
- b) **Overfitting: modelos más complejos aumentan la varianza**



#### Ejercicio 5

Se entrenaron dos **modelos de clasificación** para detectar personas sanas (clase 0) y personas enfermas (clase 1). Luego se evaluaron los modelos en los datos de test y se obtuvieron las siguientes **matrices de confusión**:

		Modelo A		Modelo B	
True	0	13	12	15	10
	1	16	22	11	27
		Predicted		Predicted	
		0	1	0	1

Padrón: \_\_\_\_\_ Nombre y Apellido: \_\_\_\_\_ Hojas: \_\_\_\_\_

---

Calcular para la Clase 1:

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

a) F1 Score Modelo A

**Recall Clase 1:**  $22 / 22 + 16 = 11/19$   
**Precisión Clase 1:**  $22 / 22 + 12 = 11/17$

b) F1 Score Modelo B

**Recall Clase 1:**  $27 / 27 + 11 = 27 / 38$   
**Precisión Clase 1:**  $27 / 27 + 10 = 27 / 37$

c) ¿Qué modelo elegiría en términos de Accuracy? ¿Por qué?

**Accuracy A :**  $TN + TP / TN + TP + FP + FN$   
**Accuracy A:**  $13 + 22 / 13 + 22 + 16 + 12$   
**Accuracy A :**  $35 / 63$

**Accuracy B :**  $TN + TP / TN + TP + FP + FN$   
**Accuracy B :**  $15 + 27 / 15 + 27 + 11 + 10$   
**Accuracy B :**  $42 / 63$

**El modelo B tiene mayor accuracy**

**Nota:** Los resultados pueden expresarse como fracción