

# Trabajo práctico 1

## Análisis multivariado UNaB 2025


### Primer cuatrimestre

Fecha máxima de entrega:  
Domingo 25 de mayo de 2025 23:59 hs

#### Integrantes del grupo:

- Quintana, Oscar
  - Gómez, Dante
  - Lozano, Juan
  - Ruiz Dias, Laura
  - Garrido, Pablo
- 

#### 1. Regresiones **LISTO PARA SUBIR**

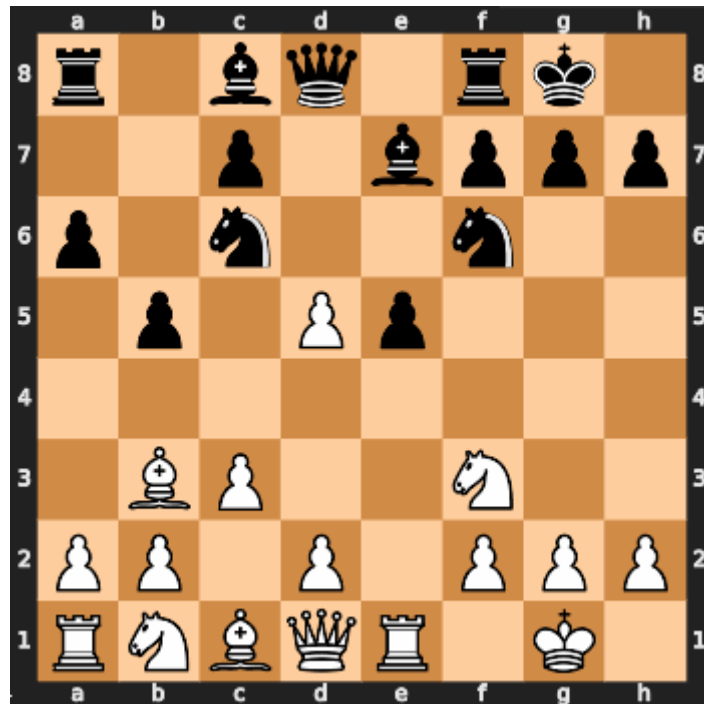
 ejercicio1\_grupo7.ipynb

Con los datos presentes en la carpeta Datos/Forecasting de la clase 2 construyan un modelo de regresión lineal que les permita realizar una predicción futura de la variable **Sales quantity**. Experimenten con distintas transformaciones de los datos, regresiones polinómicas u otras técnicas que las utilizadas en clase para obtener el mejor modelo posible en testing.

#### 2. Regresión logística **LISTO PARA SUBIR**

 ejercicio2\_grupo7.ipynb

Construyan un modelo de regresión logística en el lugar señalado en el archivo **Trabajo\_practico.ipynb** (presente en los archivos de la clase 4) e implementen su propio modelo de ajedrez que juegue contra sí mismo la posición denominada **Ataque Marshall**.




### 3. Regresiones locales **LISTO PARA SUBIR**

 ejercicio3\_grupo7.ipynb

Construyan una regresión logística pesada y una regresión lineal pesada. Utilizarlas para hacer predicciones sobre los datasets en la carpeta Datos de la clase 4 y muestren mediante gráficos de ejemplo como el parámetro  $\tau$  afecta los modelos. Además, demuestren que un modelo con un  $\tau$  muy pequeño tendrá buena performance en train mas no en test en relación con otros  $\tau$  (es decir, muestre que un  $\tau$  muy pequeño lleva al overfit).

### 4. Bayes ingenuo **LISTO PARA SUBIR**

 ejercicio4\_grupo7

Programen su propia implementación de naive Bayes. Utilicen las técnicas vistas para procesamiento del lenguaje natural para procesar el dataset presente en la carpeta Datos de la clase 5 y hagan predicciones con su modelo sobre si el texto es relevante o no (columna relevance). Su predicción podría mejorar mucho si balancean su dataset antes de pasarlo por su modelo.

### 5. LDA **LISTO PARA SUBIR**

ejercicio5\_grupo7.ipynb

Programen su propia implementación de LDA y utilicenla para predecir el género de los audios procesados en el archivo dataset.csv de la carpeta Datos de la clase 6. Pueden intentar mejorar el feature engineering de los audios crudos que se encuentran en la subcarpeta Raw si así lo consideran.

## 6. GDA **LISTO PARA REVISAR**

ejercicio6\_grupo7.ipynb

Implementen su propia versión de un modelo GDA sobre el mismo dataset que el punto anterior y compare sus resultados. ¿Que modelo dió mejor accuracy? ¿Que similitudes o diferencias encuentran en los parámetros obtenidos? ¿Es posible sacar alguna conclusión de estos resultados?

### • Revisar el manejo del determinante en GDA:

En la línea:

```
if det_sigma >= 0:  
    det_sigma = 1e-10 # Evita la division por 0
```

Esto fuerza el determinante a un valor pequeño si es mayor o igual a cero, pero debería ser si es menor o igual a cero.

Corregir a:

```
if det_sigma <= 0:  
    det_sigma = 1e-10
```

**EL PUNTO 5 Y CONCLUSIÓN, LO PUSE MÁS ACORDE A UN ESTUDIANTE QUE A UN SUPER PROFESIONAL...**