

# Tarea 1

Patrones y aprendizaje automatizado - 2023II

Fecha de entrega:

*Las personas no iniciadas en la teoría General de la improbabilidad preguntan hasta hoy día por qué, de hecho, Trurl probabilizó al dragón y no al elfo, o al gnomo. Lo hacen por ignorancia, ya que no saben que el dragón es, sencillamente, más probable.* Los dragones de la probabilidad. Stanislaw Lem

## 1. Instrucciones

Responde las siguientes preguntas y haz las implementaciones en código como se indica a continuación:

- Indica los integrantes de tu equipo en un archivo de texto llamado 00Integantes.txt. Todos deberán presentar la misma tarea dentro de la actividad en el classroom.
- Haz las implementaciones en un script python que se llame `tarea.n.py`.
- Sube todos los archivos a la entrada de la tarea correspondiente en el classroom (preferentemente sin comprimir a menos de que sea inevitable debido a su tamaño o complejidad).

## 2. Probabilidad y estadística

1. Considera la variable aleatoria  $X \sim B(n, p)$  con una población de  $n = 100$  y con  $p = 0,08$  (y por lo tanto  $q = 0,92$ ) ¿Cuál es la probabilidad asociada para todos los valores entre 5.5 y 8?

Tip: Recuerda que la forma de la media  $\mu$  y  $\sigma$  para la binomial tienen formas analíticas ya derivadas que puedes verificar en cualquier texto de probabilidad o Wikipedia

2. Considera la v.a. continua  $X$  con función de densidad como se muestra a continuación:

$$f(x) = \begin{cases} 2x & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

¿Cuál es la esperanza de  $X$ ?

3. Implementa una función que tome un arreglo de vectores **X** y que regrese dos conjuntos ajenos denominados **X\_train** y **X\_test**. Los parámetros que debe tomar esta función es un iterable con los vectores a separar así como un escalar **p** que indique la proporción de datos que quedará en el primer conjunto.

Debe regresar dos iterables donde el primer conjunto tiene la proporción **p** de **X** y el segundo **(1-p)** de **X**.

A continuación se presenta un ejemplo de la invocación de la función.

```
def repartidor(X, p=0.5):
    ...
    return X_train, X_test

X = [0, 2, 3, 4, 1, 8]
m = len(X)
X_tr, X_ts = repartidor(X, 0.5)
#X_tr puede ser [0, 8, 1]
#X_ts entonces quedaria [2, 4, 3]
assert( len(E) ) == m//2
assert( set(E).intersection(set(T)) == set() )
```

4. Usando el conjunto de datos de IRIS, haz lo siguiente usando el modelo de predicción que desees (e.g. modelo lineal, logístico, soporte vectorial, etc.):
- Determina y explica todas las métricas que pueden ser calculadas por medio de la matriz de confusión.
  - Grafica la matriz de confusión y explica los valores graficados.

La forma de importar el conjunto de datos de IRIS se muestra a continuación, así como un hint sobre el método de predicción para realizar:

```
import pandas as pd
import numpy as np
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline
from sklearn.svm import SVC
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
#
# Load the data set
#
iris = datasets.load_iris()
X = iris.data
y = iris.target
#
# Create the training and test split
#
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.25, random_state=1, stratify=y)
```

### 3. Regresión lineal

Usando el conjunto de datos de IRIS, haz lo siguiente usando el modelo de regresión lineal multivariado:

- Determina el comportamiento de cada variable de forma independiente y menciona qué tipo de función siguen (ver clase de ayudantía de regresión lineal).
- Usando el método OLS determina los coeficientes  $\beta$  del ajuste lineal, el valor  $R^2$ . ¿Qué puedes decir del estadístico F y de la significancia/importancia de cada variable en el ajuste del modelo?
- Investiga qué significan las medidas de Skew, Kurtosis, Omnibus y Durbin-Watson y menciona cómo se interpretan en este modelo en particular.
- En el caso de que alguna variable pueda prescindir del modelo, determina el nuevo ajuste, eliminando dichas variables y compara los resultados.

Si desean leer **Los dragones de la probabilidad** lo pueden hacer en esta liga <https://lacanciondelasirena.wordpress.com/2017/08/14/los-dragones-de-la-probabilidad-stanislaw-lem/>.