

Programación para *Data Science*

Unidad 7: Análisis de datos en Python - Ejercicios y preguntas

Ejercicio 0

Cargad el conjunto de datos Iris incorporado en la librería `sklearn`.

```
In [1]: # Respuesta
```

Ejercicio 1

Mostrad el nombre de los *targets* del *dataset*, el nombre de los atributos y los datos. **(0.5 puntos)**

```
In [2]: # Respuesta
```

Ejercicio 2

Representad gráficamente en un *scatter plot* la longitud frente al ancho de los pétalos. **(1.5 puntos)**

Nota: para poder incluir acentos en los textos de las etiquetas o del título del *plot*, es necesario indicar explícitamente que las cadenas de caracteres son *unicode*. Podéis hacerlo incluyendo una `u` delante de las comillas que delimitan la cadena de caracteres.

```
In [3]: # Ejemplo de cadena de caracteres unicode especificada explícitamente.
print u"pétalo"

pétalo
```

```
In [4]: # Respuesta
```

Pregunta 1

¿Que es un *box plot*? ¿Qué información podemos extraer del *box plot* mostrado en el Notebook de explicación? **(1 punto)**

Respuesta:

Pregunta 2

En el Notebook de explicación hemos utilizado un clasificador *k nearest neighbors*. Describid a grandes rasgos cómo funciona este clasificador. **(1 punto)**

Respuesta:

Pregunta 3

Ejecutad el código del clasificador *k nearest neighbors* del Notebook de explicación varias veces. ¿El resultado es siempre el mismo? ¿Por qué? **(1 punto)**

Respuesta:

Pregunta 4

Para visualizar el modelo del clasificador en el Notebook de explicación hemos usado la función `meshgrid`. Explicad qué hace esta función y para qué la hemos usado. **(1 punto)**

Respuesta:

Ejercicio 3

Visualizad gráficamente el clasificador *KNeighborsClassifier* aprendido (para cualquier partición de muestras de aprendizaje y test), mostrando tanto las muestras usadas para el aprendizaje como las muestras utilizadas para el test. Utilizad colores para mostrar la clase a la que pertenecen (`target`) las muestras de aprendizaje y las de test. **(2 puntos)**

Pista: podéis utilizar el código que hemos visto en el Notebook de explicación, añadiendo una línea que permita visualizar las muestras de test y cambiando el marcador para diferenciarlas de las muestras de aprendizaje.

```
In [5]: # Respuesta
```

Pregunta 5

Considerando la visualización generada en el ejercicio anterior, ¿qué muestras se encuentran mal clasificadas? **(1 punto)**

Respuesta:

Pregunta 6

Decimos que un método de validación cruzada es exhaustivo si entrena y aprende con todas las posibles maneras de dividir el conjunto de aprendizaje.

¿Es exhaustivo el método *KFold* utilizado en el Notebook de explicación? ¿Y el método *leave one out*? **(1 punto)**

Respuesta: