**Maching learning**

Autor: Daniel Felipe Ladino Velásquez

Jonatan Hernández Henao

*IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia*

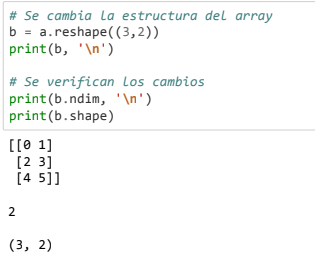
Correo-e: d.ladino@utp.edu.co

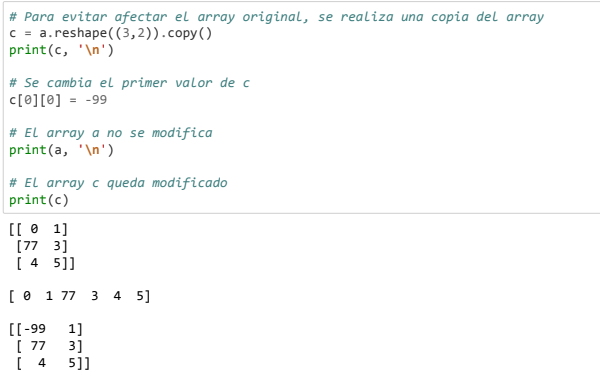
Correo-e: jonatan.hernandez@utp.edu.co

En primer se importa la librería **numpy** y se le asigna un alias **np,** esta librería contiene las diversas funciones que se necesitan para poder ejecutar el código.

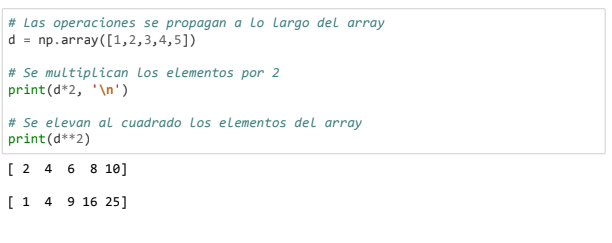
la funcion np.array lo que hace es crear un vector al cual nosotros le damos los valores, luego le decimos que nos imprima en pantalla y que salte línea print(a, '\n') y a su vez le decimos que nos imprima la dimensión del vector llamando la función (a.ndim, '\n') y luego nos imprime con a.shape el número de elementos que contiene el vector.

con a.reshape cambiamos la estructura del vector con un nuevo vector b , partiendo del vector a y vemos que los valores que se imprimen en el vector a son de una sola dimensión, en cambio el vector que imprime b su estructura cambia a una matriz de (3,2).



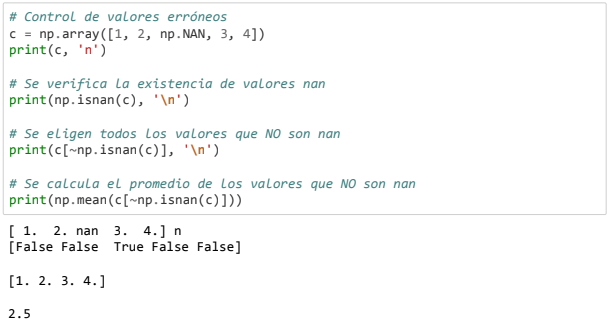


Tenemos que tener en cuenta que los cambios que hagamos en b los tendrán en el vector a, para que no nos ocurra esto, creamos una copia con la función a.reshape((3,2)).copy() e imprimimos en pantalla el resultado obteniendo una copia del vector anterior pero también se imprime el nuevo con la modificación que se le quiere hacer.

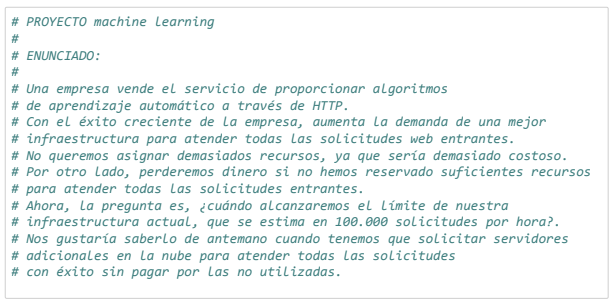


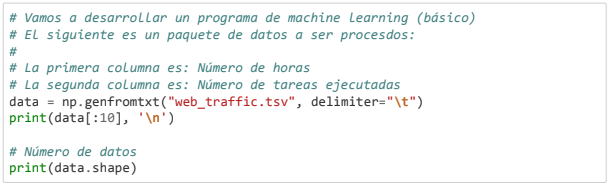
En este punto creamos un nuevo vector y lo imprimimos en pantalla, pero multiplicando todo el vector por dos y a su vez elevamos todo el vector al cuadrado y lo imprimimos.

le decimos al vector que nos imprima en pantalla (a>4, '\n’) los valores dentro del vector mayores a cuatro (4), después le decimos que a los valores dentro del vector que estén por encima del número 4 los modifique por uno (a[a>4] = 1), y a su vez imprimimos en pantalla el vector, luego reemplazamos todos los valores iguales a uno dentro del vector por (a[a==1] = 777) y lo imprimidos en pantalla.

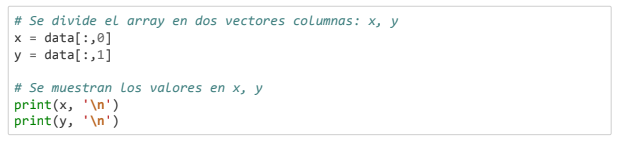


En esta función se utiliza el np.NAN para hallar los valores del vector que son erróneos, en la creación del vector se le adiciona un np.NAN y se imprime en pantalla el vector y veremos como sale el error en el vector mostrado en pantalla, luego llamamos la función np.isnan para preguntar dentro del vector que valores son NAN y donde lo encuentre imprime un (true) dando aviso de su existencia en esa posición. Lo que hace esta función es entrar al vector y a partir de la función anterior retirar cuales son NAN y elige los valores que no son para luego decirle que los imprima en pantalla (c[~np.isnan(c)], '\n'). esta lee el comando de adentro hacia afuera para poder hallar el promedio en los valores del vector np.mean(c[~np.isnan(c)], '\n').



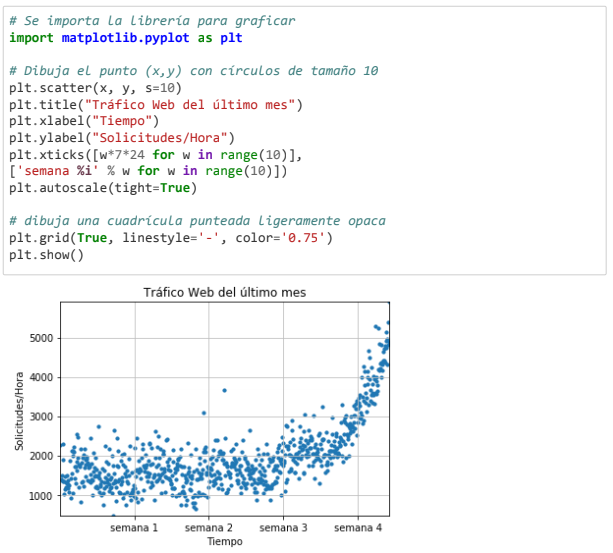


En esta función le pedimos que lea los datos informativos que se tienen en el archivo de texto, data = np.genfromtxt("web\_traffic.tsv", con esta función lo que se hace pedirle una tabulación a los datos delimiter="\t"), y con este print nos dará los datos que le solicitemos (data[:10], '\n').



Con esta instrucción lo que se hace es separar los datos completos de las dos columnas anteriores en dos vectores diferentes y los imprimimos en pantalla.

Le buscamos con (x.ndim, '\n') y (y.ndim, '\n') la dimensión de cada vector y a su vez con (x.shape,'\n') y (y.shape) nos dará la cantidad de elementos contenidos en el vector.



Se importa la librería (matplotlib.pyplot) se le asigna un alias (as plt), esto para poder graficar los datos que tenemos en el archivo de texto.

plt.scatter(x, y, s=10), esta es una función de la librería que nos permite modificar el tamaño de los puntos en el grafico que representan cada dato.

plt.title("Tráfico Web del último mes"), con esta instrucción colocamos el título principal del gráfico.

plt.xlabel("Tiempo"), con esta instrucción colocamos el título en el eje x del gráfico.

plt.ylabel("Solicitudes/Hora"), con esta instrucción colocamos el título en el eje y del gráfico.

plt.xticks([w\*7\*24 for w in range(10)]

['semana %i' % w for w in range(10)]), lo que hace esta instrucción es coger la cantidad de semanas, se multiplica por 7 que es el número de días de una semana, y a su vez se multiplica por las horas del día, con esto se agrupan los datos de las semanas y se divide en el grafico lógicamente, en un rango de 10 datos.