

# PROCESAMIENTO DE MATRICES CON NUMPY

Septiembre del 2019

Semana: 3-4

## Objetivos de aprendizaje y competencias

1. Objetivos de aprendizaje:
  - a. Importar un .csv con Pandas y transformar datos a un arreglo de numpy
  - b. Hacer cálculos simples en python
  - c. Usar funciones matemáticas en numpy
  - d. Hacer gráficos simples de funciones matemáticas con matplotlib
  - e. Hacer sumas, productos, mínimos, máximos y valores absolutos.
  - f. Resolver ecuaciones lineales
  - g. Crear y usar array
  - h. Indexar partes de un array, usando tajadas, índices, ejes
  - i. Obtener y cambiar la forma de un array
  - j. Usar funciones para crear array típicos
  - k. Hacer operaciones aritméticas con array
  - l. Hacer copias y vistas, eliminar filas y columnas
  - m. Hacer operaciones matriciales
  - n. Calcular estadísticas de filas y columnas
2. Competencias que se desarrollan:
  - a. Analizar un problema para formular una solución
  - b. Idear un algoritmo para resolver un problema
  - c. Trabajar en equipo resolviendo un problema
  - d. Presentar los resultados de un proyecto

## Descripción

El objetivo de este proyecto es procesar un conjunto de datos haciendo transformaciones, cálculos, selección, limpieza y estandarización.

Las tareas a realizar son:

1. Leer el archivo datos.csv y almacenar los datos en un array de numpy. Los datos corresponden a observaciones de los valores de mediciones del Centro Meteorológico de Chile. Los valores de las variables son numéricos y se han registrado con dos decimales. Las variables son:

- 1) Días (t)
- 2) Temperatura máxima (T1)
- 3) Temperatura mínima (T2)
- 4) Presión máxima (P1)
- 5) Presión mínima (P2)
- 6) Velocidad del viento (r)
- 7) Precipitaciones (s)

Deben unir los datos en un array de la siguiente forma:

`['t', 'T1', 'T2', 'P1', 'P2', 'r', 's']`

los nombres de las variables se usarán para rotular los gráficos y para preparar los resúmenes de datos.

2. Desarrollar una función que permita calcular los siguiente para cada una de las variables T1, T2, P1, P2, r y s:
  - a. Calcular el valor máximo, mínimo, media aritmética, desviación estándar y desviación estándar relativa. Investigue usando Google el significado de la media y la desviación estándar. ¿Cómo se calculan estos valores usando Numpy?
  - b. Graficar cada una de las variables respecto a los días (t), en donde t será el eje x, y, las variables serán el eje y. Escriba un comentario discutiendo la forma de cada gráfico. ¿Nota alguna tendencia en estos gráficos?.
  - c. Sabiendo que algunos valores tienen errores de medición porque salen del rango normal del proceso detecte estos valores cuando sean superiores a la media en **más** de tres veces la desviación estándar, o, a la media en **menos** de tres veces la desviación estándar. Esos valores (llamados **outliers**) se reemplazan por la moda (investigue el significado de la moda en google y aprenda a usar este valor en numpy). Use un gráfico **boxplot** para buscar outliers visualmente.
  - d. Realizar nuevamente los gráficos del punto 2.b y comparar ambos gráficos analizando y comentando las diferencias.
3. Cree dos nuevas columna  $rs = r * s$  y  $rsacum = \sum rs$ , es decir, la suma acumulada de los valores de  $rs$ , esto es:  $rsacum = \sum rs_{[0]} \text{ hasta } rs_{[i-1]}$ .
4. Crear una nueva columna con el porcentaje que T2 es de T1, buscar en google como se calcula porcentaje.
5. Construya un nuevo array de forma que  $rs[i,j] = r[i] * s[j]$ . Calcule lo anterior para los diez valores mayores de r. Haga un **heatmap** de este array.
6. Calcule y grafique la función:

$$f(t) = \frac{r*s}{t}$$

7. Agregue al array, una nueva variable:

$$R = e^{\frac{r^*s}{t}}$$

grafique esta variable y calcule su máximo, mínimo, media y desviación estándar. Compare con el gráfico (6).

8. Agregue al array, una nueva variable con la siguiente fórmula:

$$W = \log\left(\frac{r^*s}{t}\right)$$

grafique esta variable y calcule su máximo, mínimo, media y desviación estándar. Compare con el gráfico (6).

9. A continuación calcule las distancias entre cada par de puntos  $(T_1, P_1)$  y  $(T_2, P_2)$  usando la siguiente fórmula:

$$D = \sqrt{(T_1 - T_2)^2 + (P_1 - P_2)^2}$$

y grafique D en función de t.

10. Para el análisis final se deben mantener los registros que tengan un valor de  $R \geq 1$ , pero aquellos con  $R > 2$  deben eliminarse.

## Antecedentes

Para realizar este proyecto deberán considerar el estudio del [Tutorial de Numpy](#) (abrir en google chrome y traducir de inglés a español).

[Manual de referencia de Numpy](#)

## Instrucciones

La entrega de los trabajos se debe hacer en un cuaderno Jupyter, documentando las partes del código y los resultados como sea necesario.