

3.2. Correlación entre variables.

- La correlación se refiere a la relación entre dos variables y cómo pueden o no cambiar juntas.
- El método más común para calcular la correlación es el coeficiente de correlación Pearson, que asume una distribución normal de las variables.
- Su valor está en $[-1,1]$, siendo un valor de -1 una correlación perfectamente negativa, y un valor de $+1$ una correlación perfectamente positiva.
- Un valor de 0 indica que no existe correlación.
- Algunos algoritmos de aprendizaje automático pueden ver penalizado su rendimiento si hay variables altamente correlacionadas entre sus entradas.

Ejercicio 88_01. Correlaciones en Python 1

```
import pandas as pd
nomFichero = 'winequality-red.csv'
# No es necesario añadir nombres de columnas, ya que el csv los trae.
# Se usa ; como separador, y no la coma por defecto.
dfVinos = pd.read_csv(nomFichero, sep=";")
dfDatos = dfVinos.iloc[:,0:4];
# Mostramos las 3 primeras filas
print(dfDatos.head(3))
```

	fixed Acidity	volatile Acidity	citric Acidity	residual Sugar
0	7.4	0.70	0	1.9
1	7.8	0.88	0	2.6
2	7.8	0.76	0.04	2.3

Ejercicio 88_02. Correlaciones en MATLAB 1

```
% Importar datos desde un archivo CSV en MATLAB
nomFichero = 'winequality-red.csv';
dfVinos = readtable(nomFichero, 'Delimiter', ',');
% Seleccionar las primeras cuatro columnas
dfDatos = dfVinos(:, [1:4]);
% Mostrar las primeras tres filas
disp('Las 3 primeras filas de la tabla:');
disp(dfDatos(1:3, :));
```

Las 3 primeras filas de la tabla:

Var1	fixedAcidity	volatileAcidity	citricAcidity	residualSugar
0	7.4	0.70	0	1.9
1	7.8	0.88	0	2.6
2	7.8	0.76	0.04	2.3

Ejercicio 89_01. Correlaciones en Python 2

```
# Calculamos ahora Las correlaciones entre todas las parejas de variables:  
print("Correlaciones entre las variables:\n")  
correlaciones = dfDatos.corr(method='pearson')  
print(correlaciones)
```

Correlaciones entre las variables:

	fixed acidity	volatile acidity	citric acidity	residual sugar
fixed acidity	1.000000	-0.67696	0.70478	-0.08590
volatile acidity	-0.67696	1.000000	-0.83131	-0.28904
citric acidity	0.70478	-0.83131	1.000000	0.59803
residual sugar	-0.085909	-0.28904	0.59803	1.000000

Ejercicio 89_02. Correlaciones en MATLAB 2

```
% Importar datos desde un archivo CSV en MATLAB  
nomFichero = 'winequality-red.csv';  
dfVinos = readtable(nomFichero, 'Delimiter', ',');  
% Seleccionar las primeras cuatro columnas  
dfDatos = dfVinos(:, [2:5])  
% dfDatos = dfVinos  
% Mostrar las 3 primeras filas  
disp('Las 3 primeras filas de la tabla (quitando la primera): ');  
disp(dfDatos([2:4], :));  
% Mostrar las 3 primeras filas (quitando la primera)  
% Calcular correlaciones entre las variables  
correlaciones = corrcoef(table2array(dfDatos), 'rows', 'pairwise');  
% Mostrar las correlaciones en formato de tabla  
nombres_variables = dfDatos.Properties.VariableNames;  
correlaciones_tabla = array2table(correlaciones, 'RowNames', nombres_variables,  
'VariableNames', nombres_variables);  
% Mostrar las correlaciones en formato de tabla  
disp('Correlaciones entre las variables:');  
disp(correlaciones_tabla);
```

Correlaciones entre las variables:

	fixedAcidity	volatileAcidity	citricAcidity	residualSugar
fixedAcidity	1	-0.67696	0.70478	-0.08590
volatileAcidity	-0.67696	1	-0.83131	-0.28904
citricAcidity	0.70478	-0.83131	1	0.59803
residualSugar	-0.085909	-0.28904	0.59803	1

Correlaciones entre las variables:

	fixedAcidity	volatileAcidity	citricAcid	residualSugar
fixedAcidity	1.000000	-0.256131	0.671703	0.114777
volatileAcidity	-0.256131	1.000000	-0.552496	0.001918
citricAcid	0.671703	-0.552496	1.000000	0.143577
residualSugar	0.114777	0.001918	0.143577	1.000000