

Ejercicio 1

Tutorial “Handling Missing Values”: Los valores faltantes pueden ser un problema para el procesamiento y análisis de los datos. Qué hacer con los valores faltantes depende tanto de la cantidad de valores como del atributo para el cual faltan. En el tutorial se vieron tres posibles formas de resolver esto:

- Filtrar atributos: para aquellos atributos con una gran cantidad de valores faltantes, y que además no aportan mucha información útil.
- Reemplazar valores: para los atributos con pocos valores faltantes, y que pueden tener utilidad en el análisis.
- Filtrar ejemplos: si quedan pocos ejemplos con valores faltantes después de haber realizado alguno de los dos pasos anteriores.

Tutorial “Normalization and Outlier Detection”: Es importante identificar ejemplos inusuales y quitarlos del data set porque en la mayoría de los casos éstos son el resultado de alguna medición incorrecta. Una forma de hacer esto que con un algoritmo de distancia entre cada uno de los valores de cada ejemplo, pero primero es necesario normalizar los datos para que los valores todos los atributos se encuentren en el mismo rango.

Ejercicio 2

El data set cuenta con 177 instancias de análisis de vinos de diferentes cultivos en Italia. El problema planteado es determinar el origen del vino dados los siguientes atributos:

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alcalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

Los 13 atributos son numéricos y no tienen valores faltantes.