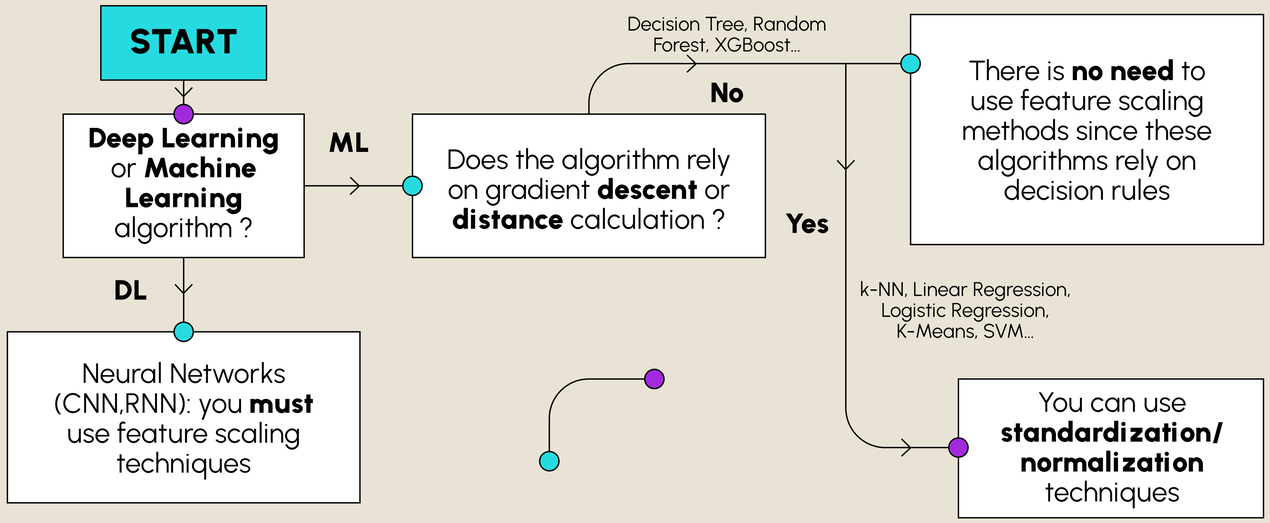
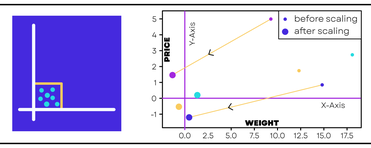
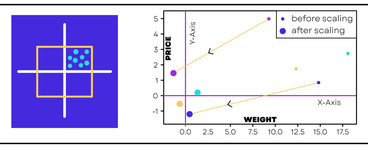
Cheat Sheet **Feature Scaling**



Normalization or standardization ? The main feature scaling techniques

Feature Scaling	Normalization	Standardization
Definition	Normalization is a technique which consists in shifting and scaling the values of a feature to redefine their interval of definition (between 0 and 1, or -1 and 1).	Standardization is a normalization technique which consists into centering the values around the mean (0) and redefining a unit standard deviation.
Use Cases	In order to gather the values of different numerical columns on a common scale and to avoid distortion.	In order to compare various measures with different units so that the features contribute to the same extent to the analysis without creating any bias.
Formula	$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$	$X' = \frac{X - \mu}{\delta}$
Explanatory pictures		
What columns values are used for the scaling process ?	Minimal and maximal values of the column	Mean and standard deviation of the column
Values Range	Between [0,1] or [-1,1]	No limit
Sensitivity to outliers	Sensitive to outliers	Way less sensitive to outliers

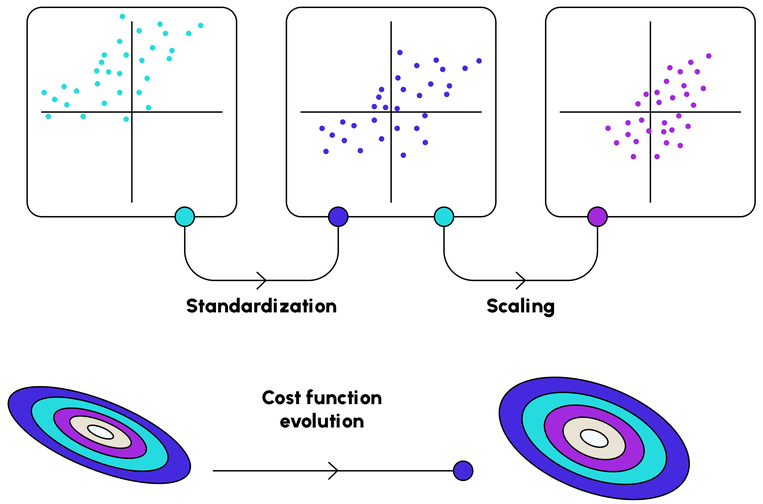
ScikitLearn Transformers	MinMaxScaler():
	`from sklearn.preprocessing import MinMaxScaler`
	`scaler = MinMaxScaler ()`
	`scaled_data = scaler.fit_transform(data)`
		MinMaxScaler():
		`from sklearn.preprocessing import StandardScaler`
		`scaler = StandardScaler ()`
		`scaled_data = scaler.fit_transform(data)`

Feature distribution	No specific distribution but not Gaussian/Normal	(Quasi) Gaussian/normal distribution amongst
Associated algorithms	k-NN, Neural Networks	Linear regression, Logistic regression, LDA
Also known under the name of...	Scaling Normalization	Z-Score Normalization

Other scaling methods

Technique	MaxAbsScaler	RobustScaler	QuantileTransformer	PowerTransformer
Definition	Transformation of each variable so that the maximal value is 1 (scaling by the absolute value of the maximal value of the column, similar to the MinMaxScaler technique for positive values).	Remove the median and transform the variable through inter quartile interval.	Also known as Rank Scaler, this technique transforms the columns such that the distribution is normal or uniform.	Transforms the columns such that the distribution is Gaussian.
Use Cases	In order to gather different numerical columns on a common scale with only positive values.	When the outliers have a negative impact on the StandardScaler technique.	For a relatively large dataset and non linear.	For data that are unequally distributed.
Specifications	Does not distort the original distribution.	The range is wider than with MinMaxScaler or StandardScaler.	Spread the most occurring values in a same column.	The variance becomes more stable and the Specifications asymmetry is minimized.
Sensitivity to outliers	Sensitive to outliers (cannot guarantee the stability of the feature scalings)	Insensitive to outliers.	Insensitive to outliers.	Insensitive to outliers.

Impact on the cost function and performances improvement :



Normalizing and Standardizing operations allow to redefine the outlines of the cost function such that they become more circular and less elliptical.
This implies that the optimization becomes simpler since it becomes closer to a first order method (gradient descent). Thus, we will benefit from a faster convergence towards the global minimum.