

K- nearest neighbours

K nearest neighbours is a supervised classification algorithm. It can be used for both classification and regression. KNN favours time and all parameters of consideration. It makes predictions using the training dataset. For a new instance, predictions are made by searching through the training dataset for the k nearest instances. Nearness is calculated through Euclidean distance. Euclidean distance is calculated to determine the k neighbours which are most similar to the new instance.

Euclidean distance is calculated as the square root of the sum of the squared differences between a new instance (n) and an existing point (i) across all attributes.

$$\text{EuclideanDistance}(x, x_i) = \sqrt{\sum (x_j - x_{ij})^2}$$

Selecting the value of k is the most critical section as it diversify the output emotion. The best approach to select k is from the formula $k = \text{number}^{(1/2)}$ where number is the total points.

There are many distance measures other than Euclidean distance which can be used with KNN such as Hamming Distance (distance between binary vectors), Manhattan Distance (distance between vectors using the sum of their absolute difference), Minkowski Distance (generalised form of Euclidean and manhattan distance). The best distance metric can be chosen by the data properties.

KNN can be used both for regression and classification.

For Regression: Prediction is based on mean or median of k-similar instances

For Classification: Output is based on the mode of k-most similar instances

It is advisable to best prepare data before using KNN. KNN succeeds more if all the data is scaled and normalized. Any missing data should be excluded from the dataset so that all the distances can be calculated. KNN works well with lower dimensional data (small number of input variables).

KNN algorithm is one of the simplest classification algorithm. It can give highly competitive results with simplicity.

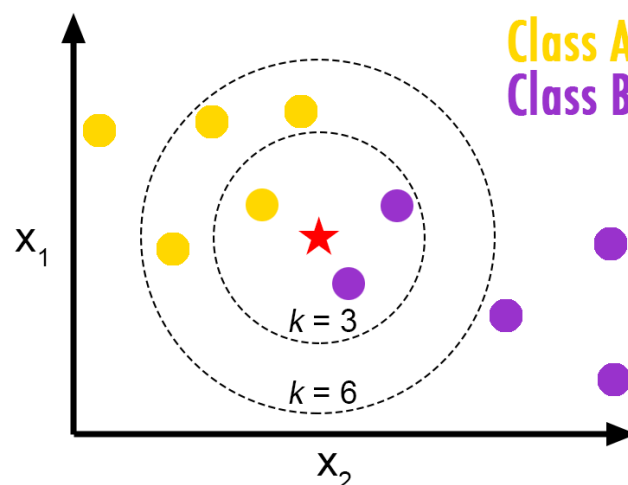


Diagram representing KNN for two classes of data

Tweets		Emotions						
	Enjoyment	Anticipation	Anger	Disgust	Sad	Surprise	Fear	Trust
I feel happy	64.2	25.5	.2	1	.3	1	.8	2
I have been in a foul mood these past few days	0	0	46.5	41	10.5	0	2	0
Feeling hateful due to the intermission between the movie	0	8.5	36	21	24	0	10.5	0
He has shown faith in me	4	25.5	0	0	0	5	0	65.5