==========

ALGORITHM: The k-Nearest Neighbor classifier

Let $\boldsymbol{k}$ be the number of nearest neighbors and **D** be the set of training examples

for each test example $\mathrm{z} = (\mathrm{x}', \mathrm{y}')$ **do**

Compute $\mathbf{d}(\mathbf{x}', \mathbf{x})$, the distance between $\boldsymbol{z}$ and every example, $(\mathbf{x}, \mathbf{y}) \in \mathbf{D}$

Select $\boldsymbol{D_z} \subseteq \boldsymbol{D}$, the set of k closest training examples to z.

$\boldsymbol{y' = argmax_v \sum_{(x_i + y_i) \in D_z} I(v = y_i)}$

end

ALGORITHM DESCRIPTION

Let $\boldsymbol{D}$ be a training dataset comprising n points $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{d}}$, given a test example $\boldsymbol{z} \in \mathbb{R}^{\boldsymbol{d}}$, and $\boldsymbol{k}$, the number of neighbors to consider, the algorithm computes the nearest neighbor distance (or similarity) between test instance $\mathbf{z} = (\mathbf{x}', \mathbf{y}')$ and all the training examples $(\mathbf{x}, \mathbf{y}) \in \mathbf{D}$ to determine its nearest neighbor list, $\boldsymbol{D_z}$.

NOTE: It can be computationally expensive if the training dataset is too large.

One the nearest neighbor list is obtained, the test instance is classified based on the majority class of its nearest neighbors:

$$\boldsymbol{y' = argmax_v \sum_{(x_i + y_i) \in D_z} I(v = y_i)}$$

where $\boldsymbol{v}$ is a class label, $\boldsymbol{y_i}$is the class label for the one of the nearest neighbors, and $\boldsymbol{I}(\cdot)$ is an indicator function that returns the value of **1** if its arguments is true and **0** otherwise.

NOTE: The algorithm is sensitive to the choice of $\boldsymbol{k}$ because every neighbor has equal impact on the classification.

To reduce the impact of $\boldsymbol{k}$ is to weight the influence of each nearest neighbor $\boldsymbol{x_i}$ according to its distance:

$$\boldsymbol{y' = argmax_v \sum_{(x_i + y_i) \in D_z} w_i \cdot I(v = y_i)}$$

References

Mohammed, Z. J. & Wagner, M., 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. New York: Cambridge University Press.

Tan, P.-N., Steinbach, M., Karpatne, A. & Kumar, V., 2019. *Introduction to Data Mining*. 2nd ed. New York: Pearson Education.