Date: 13.11.2023 / 27.11.2023

Blog Post#3

Ezgi Gümüştekin & Emir Yorgun

KNN (K-Nearest Neighbors) algorithm makes predictions based on two fundamental values:

Distance: The distance between the point to be predicted and other points is calculated. The Minkowski Distance [1] method is used for this purpose.

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Figure 1 - Minkowski Distance [1]

K (Number of Neighbors): In the KNN algorithm, the assumption is that similar things are close to each other. If K=1, there is a high probability of overfitting. If K is too large, it will result in overly general predictions. Therefore, determining the optimal K value is a crucial problem. Since the KNN algorithm makes distance-based predictions, it is highly important for the data to be scaled.

During the training phase of algorithm modeling, the model generally does not adapt itself to the data but memorizes it. This implies that the training and testing phases are almost identical, making it known as a lazy learning algorithm. It is not recommended for use with large datasets.

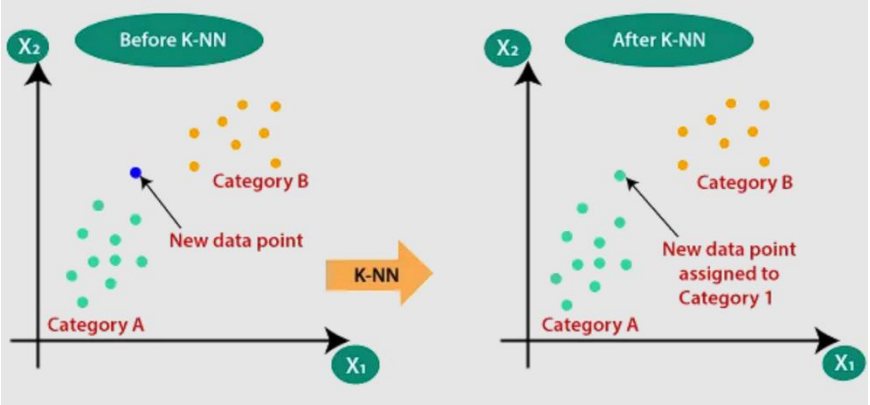

Figure 2 – KNN Before & After [4]

References

[1] Minkowski distance - Wikipedia

[2] Makine Öğrenmesi — KNN (K-Nearest Neighbors) Algoritması Nedir? | by Evren Arslan | Medium

[3] KNN (K-En Yakin Komsu). KNN algoritması, son derece basit… | by ABDULLAH ATCILI | Machine Learning Turkiye | Medium

[4] K-NN (K-Nearest Neighbors) Algoritması (Python) | by Ahmet Ataşoğlu | Medium