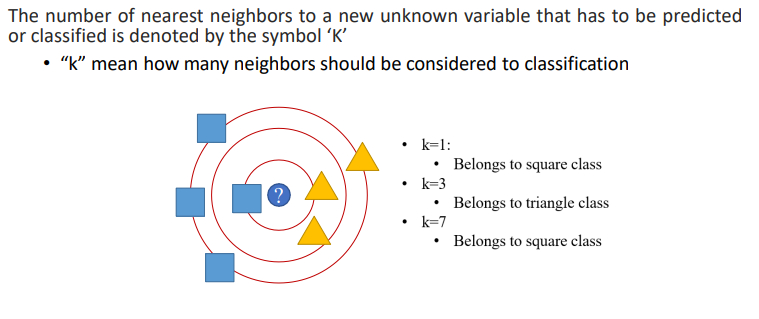
# K Nearest Neighbors

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

* The abbreviation KNN stands for “K-Nearest neighbours ”
* It is a supervised machine learning algorithm
* The KNN algorithm aims to locate all of the closest neighbours around a new unknown data point in order to figure out what class it belongs to
* It’s a distance-based approach



Imagine you have a bunch of friends, each with different interests. Now, you meet a new person and want to figure out what they might like based on who they're similar to.

KNN, or k-Nearest Neighbors, works similarly. It's like asking your friends who are most similar to the new person and then guessing what they might like based on what those similar friends enjoy.

Here's how it works:

1. Finding Neighbors: You look at the data points (your friends) closest to the new point (the new person) in the feature space (interests).
2. Majority Rule: Then, you see what the majority of those closest points are (what your similar friends like), and that's what you guess the new person might like.

Uses:

* KNN is used for classification and regression tasks. In classification, it predicts the class label of a new data point based on its neighbors. In regression, it predicts the value of a continuous target variable based on the average of its neighbors.
* It's used in recommendation systems, handwriting recognition, and detecting spam emails.

Advantages:

* Simple and easy to understand.
* No training phase (it doesn't really "learn" from the data; it just memorizes it), so it's quick to implement.

Disadvantages:

* Computationally expensive during testing because it needs to compare the new data point with all existing data points.
* Sensitive to irrelevant features or noisy data.
* Requires a meaningful distance metric (like Euclidean distance), which might be challenging to define in some cases.

Why "Lazy" Algorithm?

* It's called a "lazy" algorithm because it doesn't do any real work until you give it a new data point to classify. It just stores the data and waits until you ask for a prediction.

KNN is called a lazy learner because when we supply training data to this algorithm, the algorithm does not train itself at all.

KNN does not learn any discriminative function from the training data. But it memorizes the entire training dataset instead.

*There is no training time in KNN.*

But, this skipping of training time comes with a cost.

*Each time a new data point comes in and we want to make a prediction, the KNN algorithm will search for the nearest neighbors in the entire training set*.

Hence the prediction step becomes more time-consuming and computationally expensive.