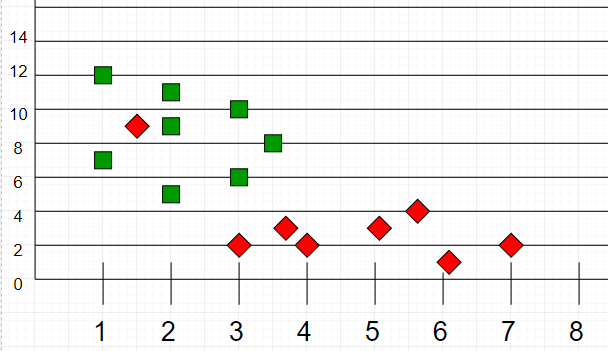
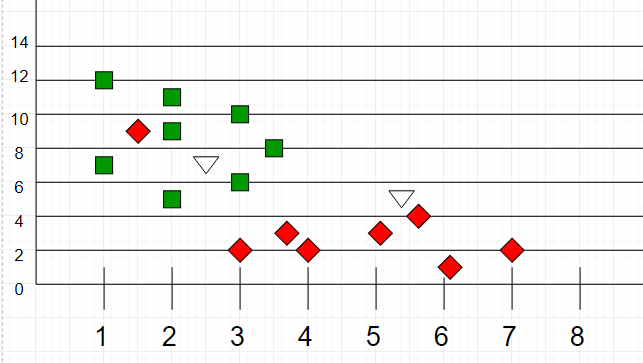
**KNN Algorithm Intuition**

K-Nearest Neighbours is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection.  
It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as [GMM](https://en.wikipedia.org/wiki/Mixture_model), which assume a Gaussian distribution of the given data).  
We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.  
As an example, consider the following table of data points containing two features: 



Now, given another set of data points (also called testing data), allocate these points a group by analyzing the training set. Note that the unclassified points are marked as ‘White’.



K in KNN is a parameter that refers to the number of the nearest neighbours to include in the majority voting process.

**How do we choose K?**

Sqrt(n), where n is a total number of data points(if in case n is even we have to make the value  odd by adding 1 or subtracting 1 that helps in select better)

**When to use KNN?**

We can use KNN when Dataset is labelled and noise-free and it’s must be small because KNN is a *“Lazy learner”*. Let’s understand KNN algorithm with the help of an example

|  |  |  |  |
| --- | --- | --- | --- |
| NAME | AGE | GENDER | CLASS OF SPORTS |
| Ajay | 32 | 0 | Football |
| Mark | 40 | 0 | Neither |
| Sara | 16 | 1 | Cricket |
| Zaira | 34 | 1 | Cricket |
| Sachin | 55 | 0 | Neither |
| Rahul | 40 | 0 | Cricket |
| Pooja | 20 | 1 | Neither |
| Smith | 15 | 0 | Cricket |
| Laxmi | 55 | 1 | Football |
| Michael | 15 | 0 | Football |

Here male is denoted with numeric value 0 and female with 1. Let’s find in which class of people Angelina will lie whose k factor is 3 and age is 5. So we have to find out the distance using

*d=√((x2-x1)²+(y2-y1)²) to find the distance between any two points.*

So let’s find out the distance between Ajay and Angelina using formula

*d=√((age2-age1)²+(gender2-gender1)²)*

*d=√((5-32)²+(1-0)²)*

*d=√729+1*

*d=27.02*

Similarly, we find out all distance one by one.

|  |  |
| --- | --- |
| Distance between Angelina and | Distance |
| Ajay | 27.02 |
| Mark | 35.01 |
| Sara | 11.00 |
| Zaira | 9.00 |
| Sachin | 50.01 |
| Rahul | 35.01 |
| Pooja | 15.00 |
| Smith | 10.00 |
| Laxmi | 50.00 |
| Michael | 10.05 |

So the value of ***k***factor is 3 for Angelina. And the closest to 3 is 9,10,10.5 that is closest to Angelina are Zaira, Smith and Michael.

                                       Zaira         9           cricket

                                      Michael      10         cricket

                                      smith          10.5      football

so according to KNN algorithm, Angelina will be in the class of people who like cricket. So this is how KNN algorithm works.