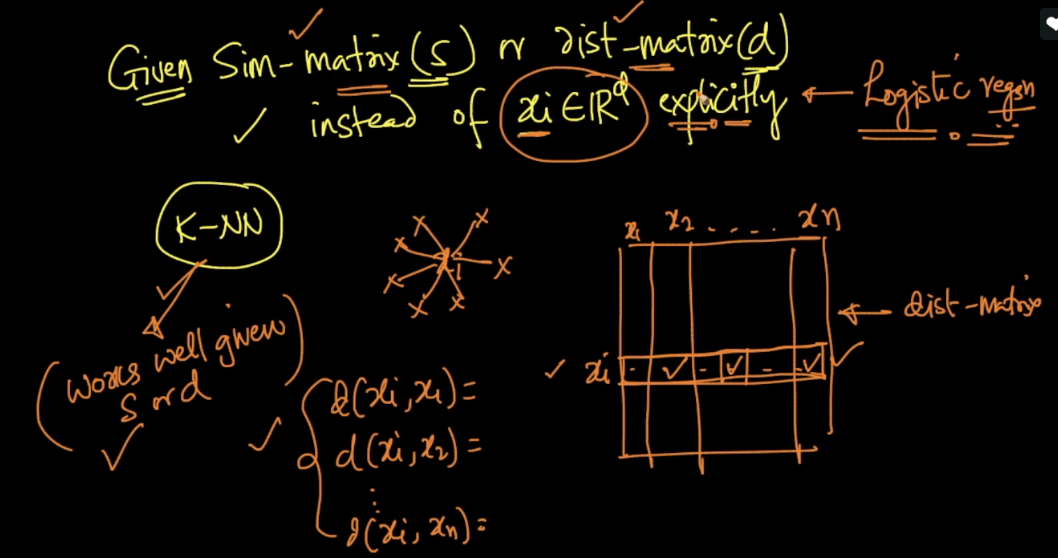


for some kind of problems(like pharma), the individual feature values will not be given. Also it is not possible to convert into vector form(like chemical compound or formula can’t be converted to numerical vector). Hence in that case, similarity values will be given.   
For example, we are given 2 chemical compounds 'A' and 'B'. Here the Domain expert knows only the composition of ingredients used in these compounds. He only knows at the max, how to compute the similarity scores of two compounds. So for each chemical compound they give similarity score with all the other compound. In such cases, apart from similarity scores we do not have anything else to move forward, so we will go with similarity matrix between compounds. When we have only similarity matrix, then the only algorithm left over for us is K-NN.



Since we have similarity matrix we can convert them to distance simply by 1 / similarity. Since we have distance of each point, that’s what k-NN needed. All classifiers can’t be implemented using similarity matrix or distance matrix. There is also one classifier called SVM kernel can implement such type of problems.

**Comments:**

* How do we find the distance between unseen Xq(vector) and given distances(scalar value) are the values between seen data?

our query point will also present in our similarity matrix right so you can get the distances for query points directly