What is Nearest Neighbour Classification

Let us say a new person is expected to join your organization or institution. You want to know whether he is a fan of CSK, or SRH or MI. Your only means of settling this question is a bunch of photographs, where he is in formals and his friends are wearing IPL uniforms.

Intuitively, we can say, "Look at the photographs. See which uniforms/jerseys are more and there is your answer!"

Nearest neighbour algorithms do just that. We need to carefully define what is a neighbour and what is nearest thats all.

More formally, nearest neighbor classifiers classify unlabeled examples by analysing the classes of *similar* labeled examples.

Despite the simplicity of this idea, nearest neighbor methods are extremely powerful. They have been used successfully for:

- Computer vision applications, including optical character recognition and facial recognition in both still images and video.

- Predicting whether a person will enjoy a movie or music recommendation.

- Identifying patterns in genetic data, perhaps to use them in detecting specific proteins or diseases.

The k-NN Algorithm

This is perhaps the simplest machine learning algorithm. As noted it is still used widely due to its effectiveness.

The k-NN algorithm gets its name from the fact that it uses information about an example's k-nearest neighbors to classify it. We identify the k records in the training data that are the "nearest" in similarity. The unlabeled test instance is assigned to the class of the majority of the k nearest neighbors.

The ε Neighbourhood Algorithm

Here instead of looking at the nearest k-neighbours and counting the majority, we look the neighbours within a specified 'radius' (namely ε) and classify the example as belonging to the majority in that radius.