K Nearest Neighbors - Classification

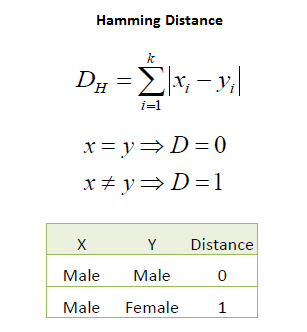
K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970’s as a non-parametric technique.

Algorithm

A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor.



For continuous data , for categorical data use hamming distance



|  |  |  |
| --- | --- | --- |
| We can now use the training set to classify an unknown case (Age=48 and Loan=$142,000)  using Euclidean distance. If K=1 then the nearest neighbor is the last case in the training set with Default=Y. |  |  |
|  |  |  |
| D = Sqrt[(48-33)^2 + (142000-150000)^2] = 8000.01  >> Default=Y |  |  |

