K-Nearest Neighbor (KNN) Algorithm

BY:

DR. VIPIN KUMAR

DEPARTMENT OF COMPUTER SCIENCE &IT

MAHATMA GANDHI CENTRAL UNIVERSITY

MOTIHARI, BIHAR

Outline…

- Defining The Instance Based Learning
- Types Of Instance Based Learning
- K-nearest Neighbors (KNN) Algorithm
- Example Of KNN In Two Dimensional Space
- Example Of Data Set To Apply The Knn Algorithm

Instance Based Learning:

- Its simply store the training example.
- Relationship of new query instance to the previously stored examples is aimed to assign a target function value for new instance.
- It includes two methods:
  - Nearest Neighbor
  - Weighted Regression
- It is assumed that instances can be represented as points in Euclidean space.
- Its some times referred as “Lazy” learning method.

K-Nearest Neighbor Learner (KNN)

- It is most basic instance-base method.
- It assumes that all instance correspond to the points in the n-dimensional Euclidean space.
- Let a instance space x can be described as feature vector

$$x = \{a_1(x), a_2(x), a_3(x), \ldots, a_n(x)\}$$

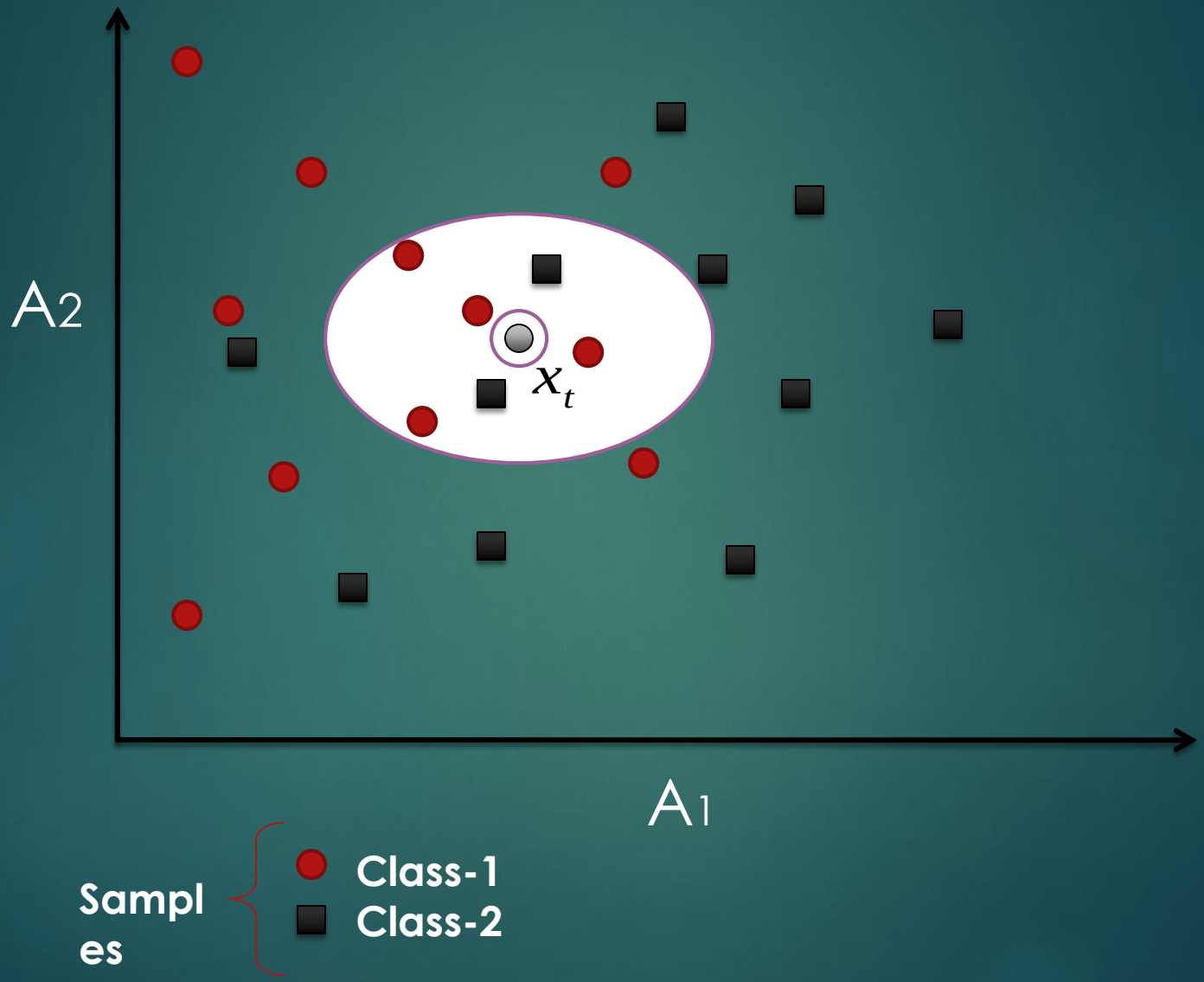
K-Nearest Neighbor Learner (KNN)

- where $a_t(x)$ denotes the value of the t^{th} attribute of instance x . Then the distance between instance x_i and x_j

can be defined:

$$d(x_i, x_j) = \sqrt{\sum_{t=1}^{n} \{a_t(x_i) - a_t(x_j)\}^2} \quad \text{.....(1)}$$

Two Dimensional Space



Distance-Weight Nearest Neighbor

- The refinement of the k-Nearest Neighbor method is to assign weight corresponding their distance to the query point.
- Then, above equation can be modified as:

$$\hat{f}(x_t) \leftarrow \arg\max_{y \in Y} \sum_{i=1}^{k} w_i \times \ell(y, f(x_i)) \text{.....(4)}$$

Distance-Weight Nearest Neighbor

- $$w_i = \frac{1}{d(x_t, x_i)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

- The real valued target functions can also be modify in a similar manner, such as:

$$\hat{f}(x_t) \leftarrow \frac{\sum_{i=1}^{k} w_i \times f(x_i)}{\sum_{i=1}^{k} w_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Example: Find out the labels of the given test sample using training set.

S.N.	A1	A2	A3	Labels
D1	3.54	5.63	3.85	2.63
D2	3.23	1.86	2.84	2.65
D3	4.76	3.87	3.75	3.39
D4	2.45	2.07	4.84	2.74
D5	3.51	4.46	2.86	3.58
Dt	3.98	8.97	3.75	?

Bibliography

- E. Alpaydin, **Introduction of Machine Learning**, Prentice Hall of India, 2006
- T.M. Mitchell, **Machine Learning**, McGraw-Hill, 1997
- C.M. Bishop, **Pattern Recognition and Machine Learning**, Springer, 2006.
- R.O. Duda, P.E. Hart and D.G. Stork, **Pattern Classification**, John Wiley, and Sons, 2001.
- Vladimir N. Vapnik, **Statistical Learning Theory**, John Wiley and Sons, 1998.
- Shawe-Taylor J. and Cristianini N., **Introduction to Support Vector Machines**, Cambridge University Press, 2000.

Thank You