K-Nearest Neighbor Learning

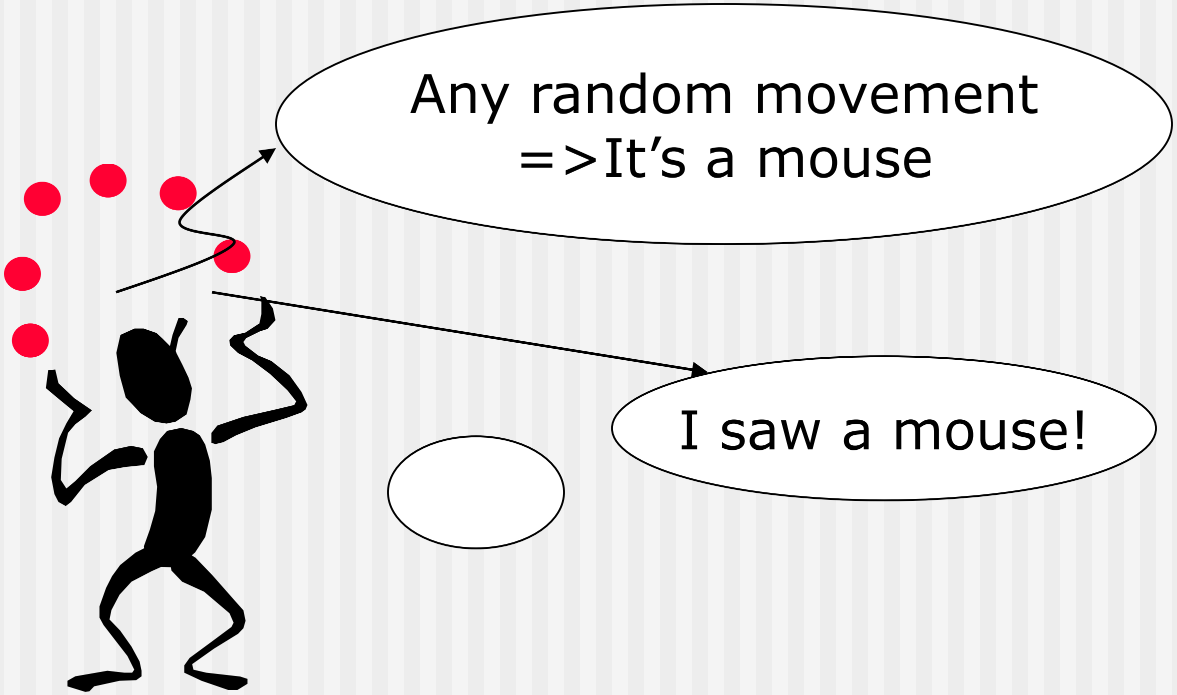
Dipanjan Chakraborty

Different Learning Methods

- Eager Learning
  - Explicit description of target function on the whole training set
- Instance-based Learning
  - Learning=storing all training instances
  - Classification=assigning target function to a new instance
  - Referred to as “Lazy” learning

Different Learning Methods

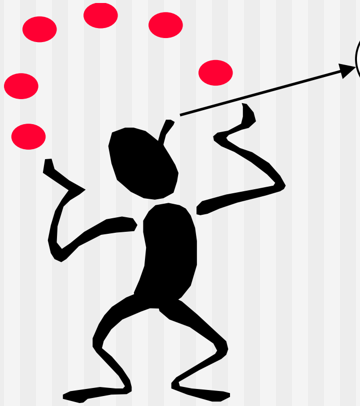
- Eager Learning



Any random movement =>It’s a mouse

I saw a mouse!

Instance-based Learning



Its very similar to a Desktop!!

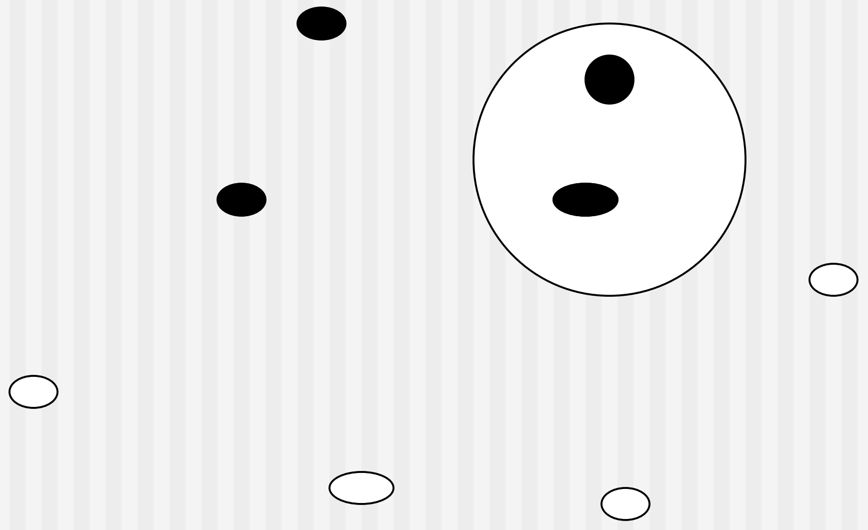
Instance-based Learning

- K-Nearest Neighbor Algorithm
- Weighted Regression
- Case-based reasoning

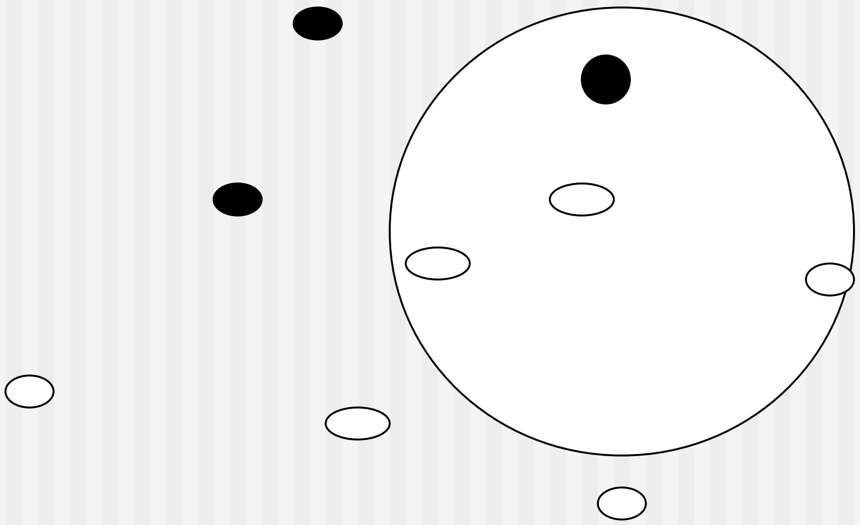
K-Nearest Neighbor

- Features
  - All instances correspond to points in an n-dimensional Euclidean space
  - Classification is delayed till a new instance arrives
  - Classification done by comparing feature vectors of the different points
  - Target function may be discrete or real-valued

1-Nearest Neighbor



3-Nearest Neighbor



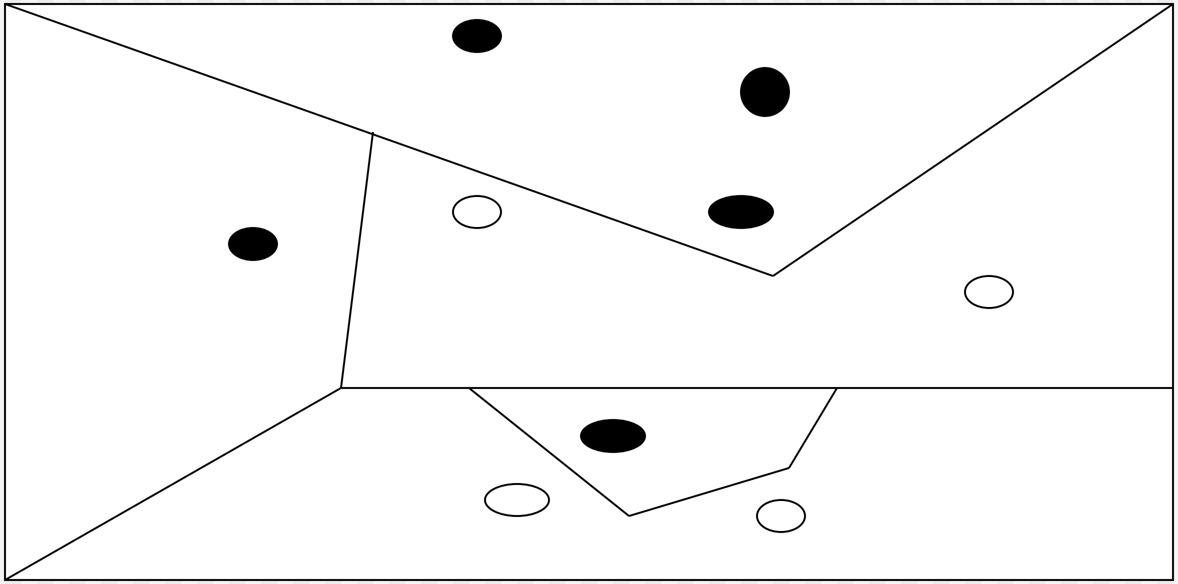
K-Nearest Neighbor

- An arbitrary instance is represented by $(a_1(x), a_2(x), a_3(x),.., a_n(x))$
  - $a_i(x)$ denotes features
- Euclidean distance between two instances

  $d(x_i, x_j)$=sqrt (sum for r=1 to n $(a_r(x_i) - a_r(x_j))^2$)
- Continuous valued target function
  - mean value of the k nearest training examples

Voronoi Diagram

- Decision surface formed by the training examples



Distance-Weighted Nearest Neighbor Algorithm

- Assign weights to the neighbors based on their ‘distance’ from the query point
  - Weight ‘may’ be inverse square of the distances

→ All training points may influence a particular instance

  - Shepard’s method

Remarks

+Highly effective inductive inference method for noisy training data and complex target functions

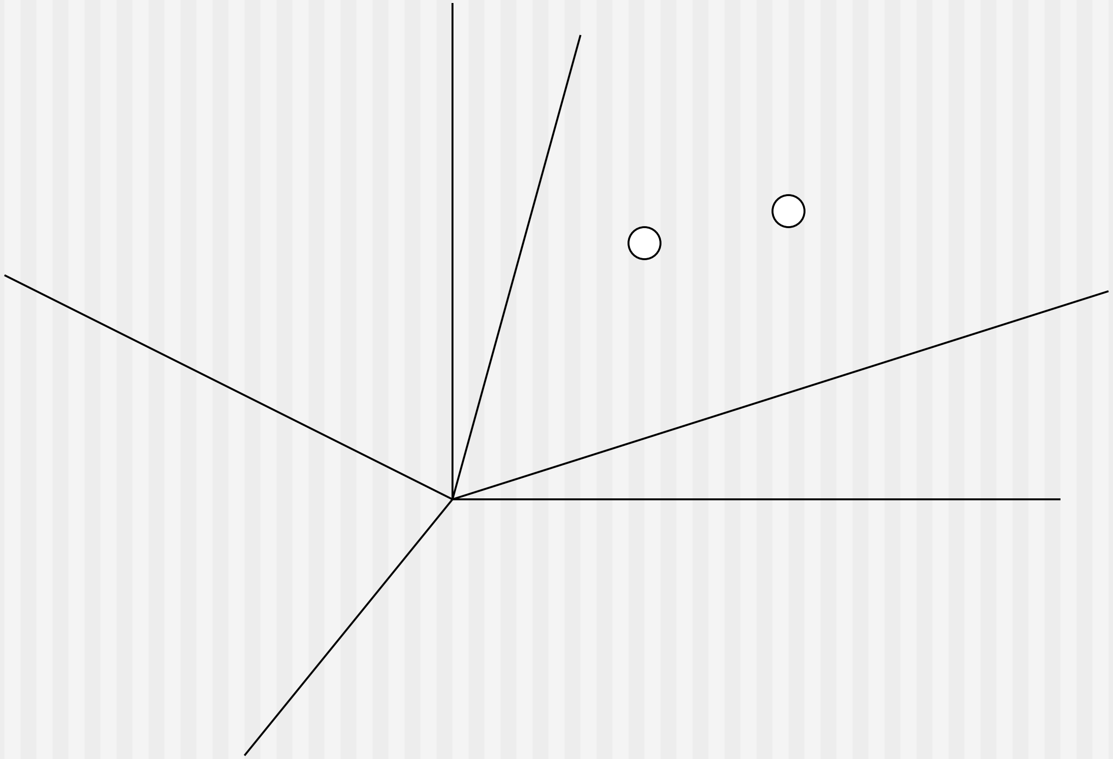
+Target function for a whole space may be described as a combination of less complex local approximations

+Learning is very simple

- Classification is time consuming

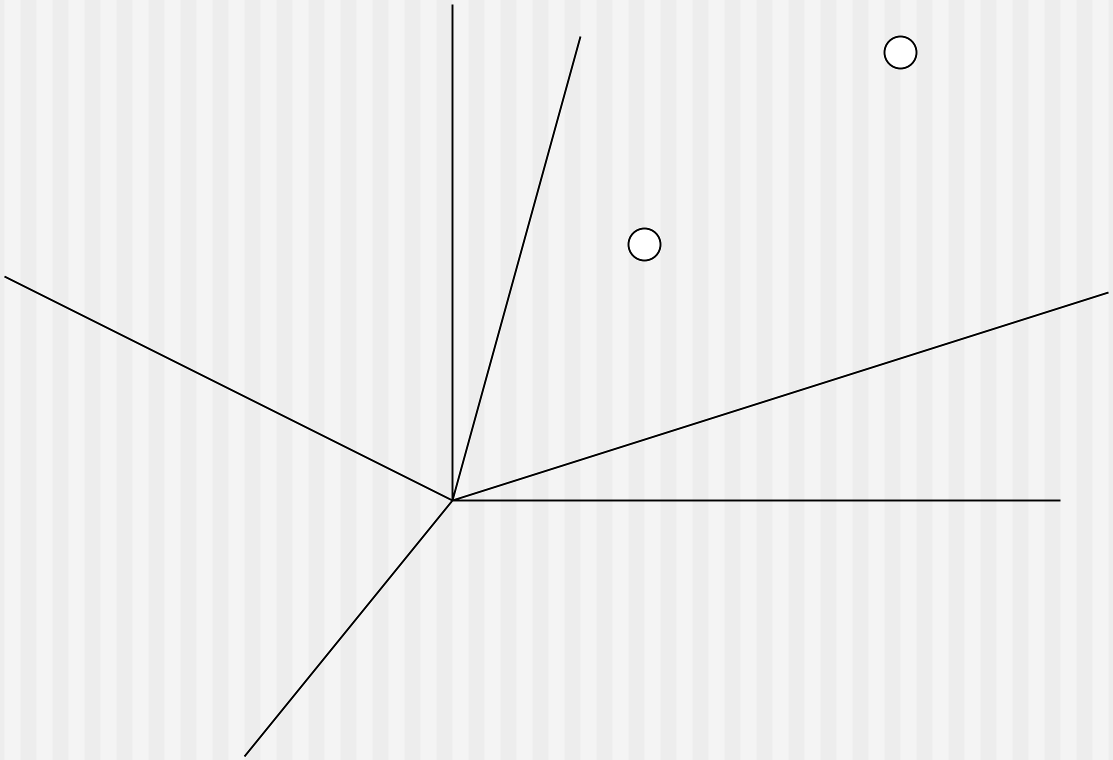
Remarks

- Curse of Dimensionality



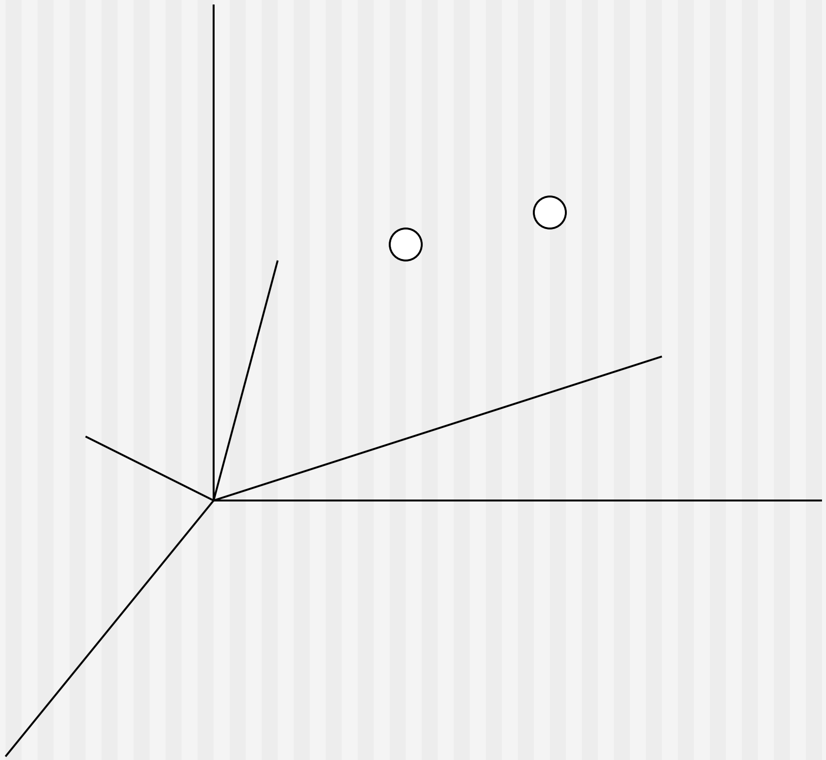
Remarks

- Curse of Dimensionality



Remarks

- Curse of Dimensionality



Remarks

- Efficient memory indexing
    - To retrieve the stored training examples (kd-tree)