Instance Based Learning

Based on Raymond J. Mooney's slides
University of Texas at Austin

Instance-Based Learning

- Unlike other learning algorithms, does not involve construction of an explicit abstract generalization but classifies new instances based on direct comparison and similarity to known training instances.
- Training can be very easy, just memorizing training instances.
- Testing can be very expensive, requiring detailed comparison to all past training instances.
- Also known as:
  - Case-based
  - Exemplar-based
  - Nearest Neighbor
  - Memory-based
  - Lazy Learning

Similarity/Distance Metrics

- Instance-based methods assume a function for determining the similarity or distance between any two instances.
- For continuous feature vectors, Euclidian distance is the generic choice:

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^{n} (a_p(x_i) - a_p(x_j))^2}$$

Where $a_p(x)$ is the value of the pth feature of instance x.

- For discrete features, assume distance between two values is 0 if they are the same and 1 if they are different (e.g. Hamming distance for bit vectors).
- To compensate for difference in units across features, scale all continuous values to the interval [0,1].

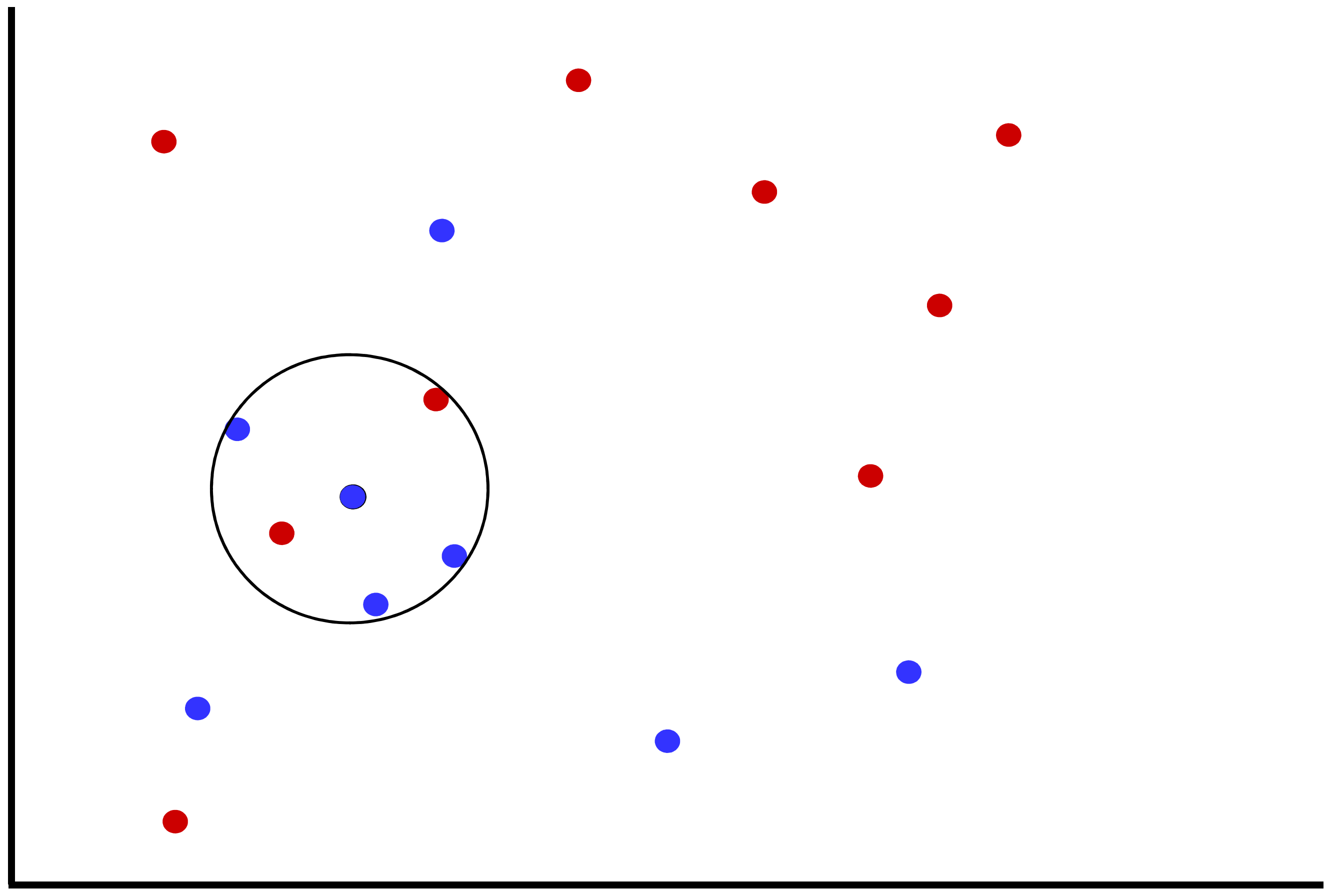
Other Distance Metrics

- Mahalanobis distance
  - Scale-invariant metric that normalizes for variance.
- Cosine Similarity
  - Cosine of the angle between the two vectors.
  - Used in text and other high-dimensional data.
- Pearson correlation
  - Standard statistical correlation coefficient.
  - Used for bioinformatics data.
- Edit distance
  - Used to measure distance between unbounded length strings.
  - Used in text and bioinformatics.

K-Nearest Neighbor

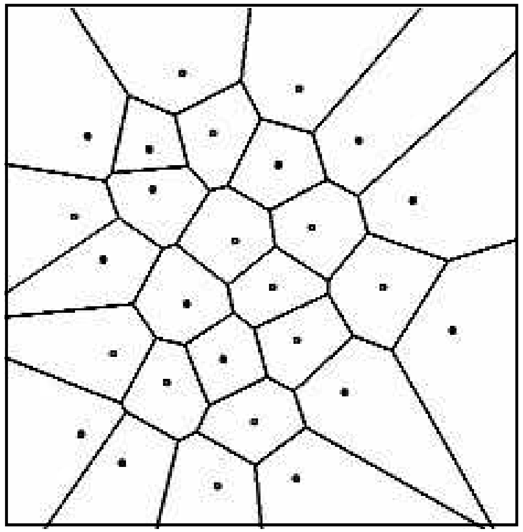
- Calculate the distance between a test point and every training instance.
- Pick the *k* closest training examples and assign the test instance to the most common category amongst these nearest neighbors.
- Voting multiple neighbors helps decrease susceptibility to noise.
- Usually use odd value for *k* to avoid ties.

5-Nearest Neighbor Example



Implicit Classification Function

- Although it is not necessary to explicitly calculate it, the learned classification rule is based on regions of the feature space closest to each training example.
- For 1-nearest neighbor with Euclidian distance, the **Voronoi diagram** gives the complex polyhedra segmenting the space into the regions closest to each point.



Efficient Indexing

- Linear search to find the nearest neighbors is not efficient for large training sets.
- Indexing structures can be built to speed testing.
- For Euclidian distance, a **kd-tree** can be built that reduces the expected time to find the nearest neighbor to O(log n) in the number of training examples.
  - Nodes branch on threshold tests on individual features and leaves terminate at nearest neighbors.
- Other indexing structures possible for other metrics or string data.
  - Inverted index for text retrieval.

kd-tree

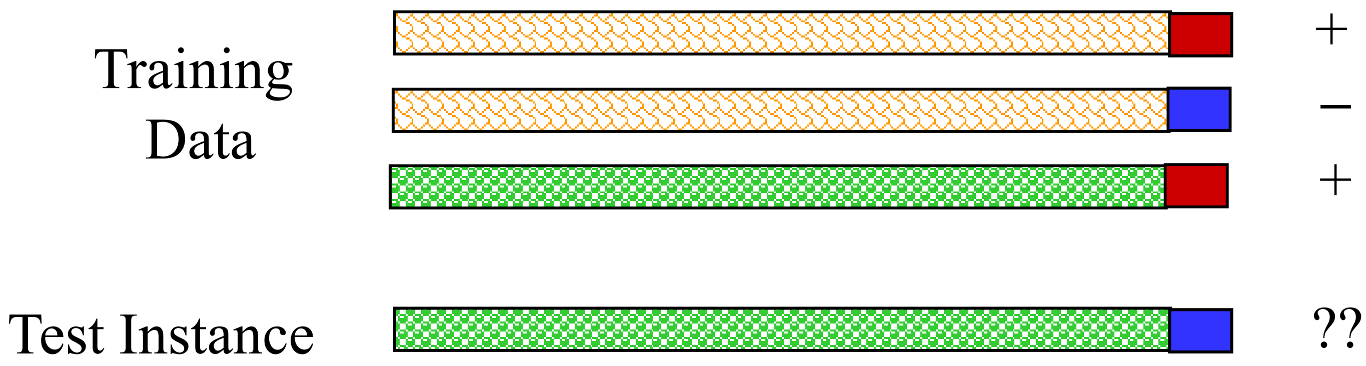
- The kd-tree is a binary tree in which every node is a k-dimensional point.
- Every non-leaf node generates a splitting hyperplane that divides the space into two subspaces.
- Points left to the hyperplane represent the left sub-tree of that node and the points right to the hyperplane by the right sub-tree.
- The hyperplane direction is chosen in the following way: every node split to sub-trees is associated with one of the k-dimensions, such that the hyperplane is perpendicular to that dimension vector.

Nearest Neighbor Variations

- Can be used to estimate the value of a real-valued function (regression) by taking the average function value of the *k* nearest neighbors to an input point.
- All training examples can be used to help classify a test instance by giving every training example a vote that is weighted by the inverse square of its distance from the test instance.

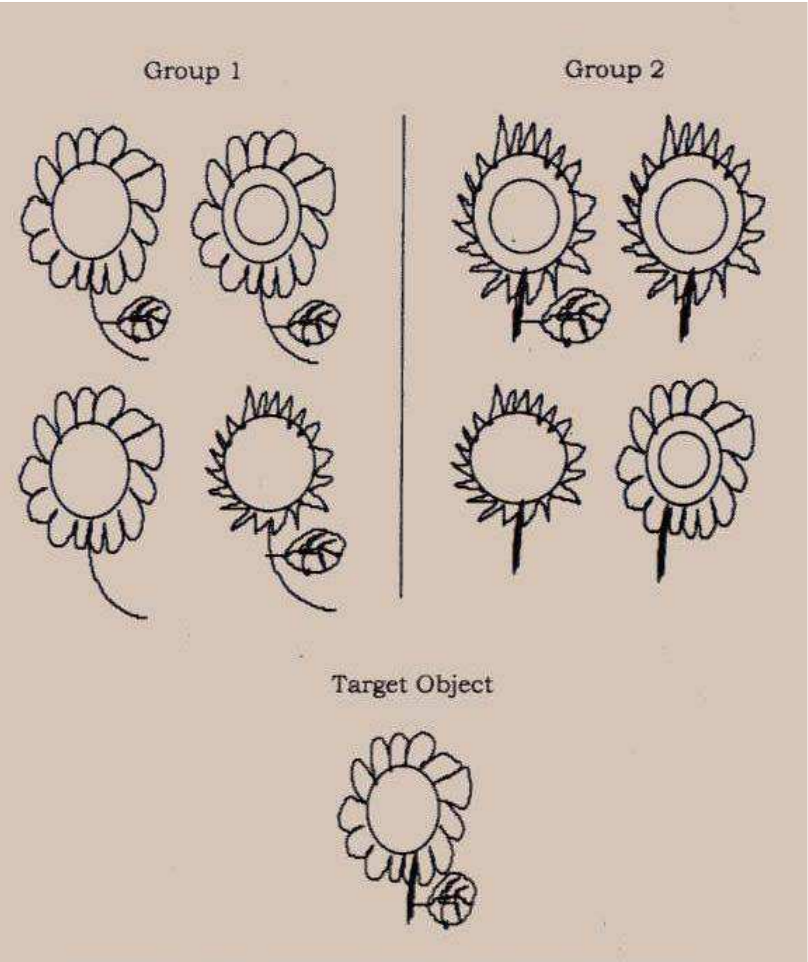
Feature Relevance and Weighting

- Standard distance metrics weight each feature equally when determining similarity.
  – Problematic if many features are irrelevant, since similarity along many irrelevant examples could mislead the classification.
- Features can be weighted by some measure that indicates their ability to discriminate the category of an example, such as information gain.
- Overall, instance-based methods favor global similarity over concept simplicity.



Rules and Instances in Human Learning Biases

- Psychological experiments show that people from different cultures exhibit distinct categorization biases.
- "Western" subjects favor simple rules (straight stem) and classify the target object in group 2.
- "Asian" subjects favor global similarity and classify the target object in group 1.



Example of item measuring whether judgments of similarity are based on family resemblance or rules.

Other Issues

- Can reduce storage of training instances to a small set of representative examples.
  - Support vectors in an SVM are somewhat analogous.
- Can hybridize with rule-based methods or neural-net methods.
  - Radial basis functions in neural nets and Gaussian kernels in SVMs are similar.
- Can be used for more complex relational or graph data.
  - Similarity computation is complex since it involves some sort of graph isomorphism.
- Can be used in problems other than classification.
  - Case-based planning
  - Case-based reasoning in law and business.

Conclusions

- IBL methods classify test instances based on similarity to specific training instances rather than forming explicit generalizations.
- Typically trade decreased training time for increased testing time.