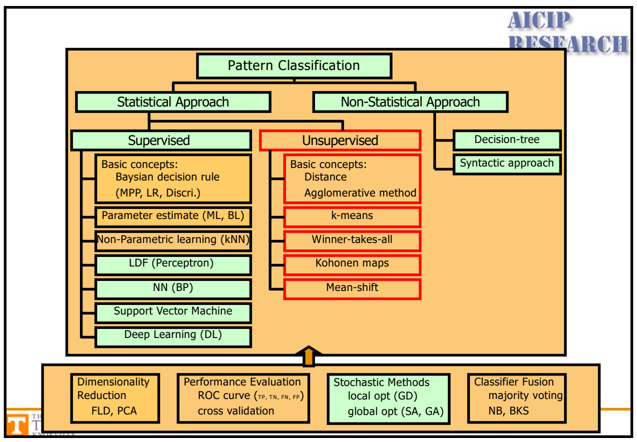
THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

ECE471-571 – Pattern Recognition

Lecture 12 – Unsupervised Learning (Clustering)

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu



AICIP RESEARCH

Review - Bayes Decision Rule

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

Maximum Posterior Probability: For a given x, if $P(\omega_1|x) > P(\omega_2|x)$, then x belongs to class 1, otherwise 2

Likelihood Ratio: If $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$, then x belongs to class 1, otherwise, 2.

Discriminant Function: The classifier will assign a feature vector x to class ω_i if $g_i(x) > g_j(x)$

- Case 1: Minimum Euclidean Distance (Linear Machine), $\Sigma_i=\sigma^2 I$
- Case 2: Minimum Mahalanobis Distance (Linear Machine), $\Sigma_i = \Sigma$
- Case 3: Quadratic classifier , Σ_i = arbitrary

Non-parametric kNN: For a given x, if $k_1/k > k_2/k$, then x belongs to class 1, otherwise 2

Estimate Gaussian (Maximum Likelihood Estimation, MLE), Two-modal Gaussian

Dimensionality reduction

Performance evaluation ROC curve

THE UNIVERSITY OF TENNESSEE

3

Unsupervised Learning

- What's unknown?
  - In the training set, which class does each sample belong to?
  - For the problem in general, how many classes (clusters) is appropriate?

Clustering Algorithm

- Agglomerative clustering
  - Step1: assign each data point in the training set to a separate cluster
  - Step2: merge the two "closest" clusters
  - Step3: repeat step2 until you get the number of clusters you want or the appropriate cluster number
- The result is highly dependent on the measure of cluster distance

Distance from a Point to a Cluster

- Euclidean distance
- City block distance
- Squared Mahalanobis distance

$$d_{euc}(x, A) = \|x - \mu_A\|$$

$$d_{mah}(x, A) = (x - \mu_A)^T \Sigma_A^{-1} (x - \mu_A)$$

Distance between Clusters

- The centroid distance

$$d_{mean}(A,B) = \|\mu_A - \mu_B\|$$

- Nearest neighbor measure

$$d_{\min}(A,B) = \min_{a,b} d_{euc}(a,b) \quad \text{for } a \in A, b \in B$$

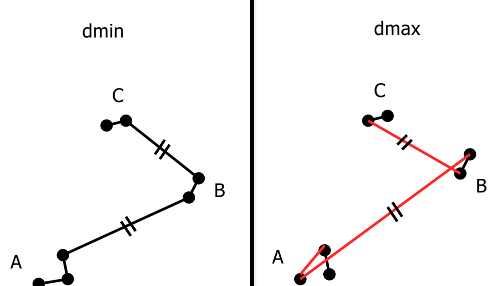
- Furthest neighbor measure

$$d_{\max}(A,B) = \max_{a,b} d_{euc}(a,b) \quad \text{for } a \in A, b \in B$$

THE UNIVERSITY OF TENNESSEE

7

Example



THE UNIVERSITY OF TENNESSEE

8

Minimum Spanning Tree

- Step1: compute all edges in the graph
- Step2: sort the edges by length
- Step3: beginning with the shortest edge, for each edge between nodes u and v, perform the following operations:
  - Step3.1: A = find(u) (A is the cluster where u is in)
  - Step3.2: B = find(v) (B is the cluster where v is in)
  - Step3.3: if (A!=B) C=union(A, B), and erase sets A and B

THE UNIVERSITY OF TENNESSEE

9

Comparison of Shape of Clusters

AICIP RESEARCH

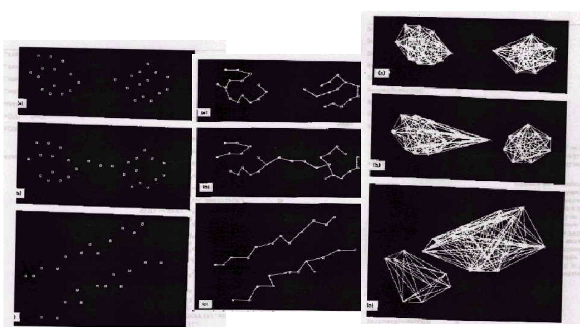
- dmin tends to choose clusters which are ??
- dmax tends to choose clusters which are ??

THE UNIVERSITY OF TENNESSEE

10

Example

AICIP RESEARCH



THE UNIVERSITY OF TENNESSEE

11

The k-means Algorithm

AICIP RESEARCH

- Step1: Begin with an arbitrary assignment of samples to clusters or begin with an arbitrary set of cluster centers and assign samples to nearest clusters
- Step2: Compute the sample mean of each cluster
- Step3: Reassign each sample to the cluster with the nearest mean
- Step4: If the classification of all samples has not changed, stop; else go to step 2.

THE UNIVERSITY OF TENNESSEE

12

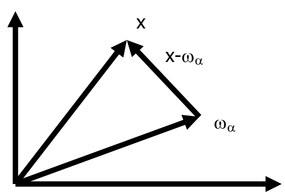
Winner-take-all Approach

AICIP RESEARCH

- Begin with an arbitrary set of cluster centers ω_i
- For each sample **x**, find the nearest cluster center ω_α, which is called the winner.
- Modify ω_α using $\omega_\alpha^{new} = \omega_\alpha^{old} + \varepsilon(\mathbf{x} - \omega_\alpha^{old})$
  - ε is known as a "learning parameter".
  - Typical values of this parameter are small, on the order of 0.01.

Winner-take-all

AICIP RESEARCH



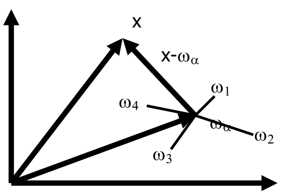
*Kohonen Feature Maps (NN)

AICIP RESEARCH

- An extension of the winner-take-all algorithm. Also called **self-organizing feature maps**
- A problem-dependent topological distance is assumed to exist between each pair of the cluster centers
- When the winning cluster center is updated, so are its neighbors in the sense of this topological distance.

AICIP RESEARCH

*SOM – A Demo



THE UNIVERSITY OF TENNESSEE

16

AICIP RESEARCH

SOM – The Algorithm

- The winning cluster center and its neighbors are trained based on the following formula

$$\omega_r^{k+1} = \omega_r^k + \varepsilon(k)\Phi(k)\left(x - \omega_r^k\right)$$

ω_r are the cluster centers

g_{ω_r} are the coordinate of the cluster centers

$g_{\omega_{winner}}$ is the coordinate of the winner

Learning Rate (as k inc, ε dec, more stable)

$$\varepsilon(k) = \varepsilon_{\max}\left(\frac{\varepsilon_{\min}}{\varepsilon_{\max}}\right)^{\frac{k}{k_{\max}}}$$

$$\Phi(k) = \exp\left(-\frac{\left\|g_{\omega_r} - g_{\omega_{winner}}\right\|^2}{2\sigma^2}\right)$$

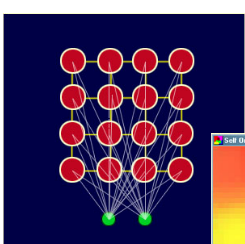
The closer (topological closeness) the neighbor, the more it will be affected.

THE UNIVERSITY OF TENNESSEE

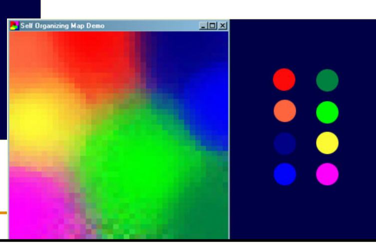
17

AICIP RESEARCH

SOM - An Example



http://www.ai-junkie.com/ann/som/som1.html



THE UNIVERSITY OF TENNESSEE

*Mean Shift Clustering

AICIP RESEARCH

- Originally proposed in 1975 by Fukunaga and Hostetler for mode detection
- Cheng's generalization to solve clustering problems in 1995
- Non-parametric – no prior knowledge needed about number of clusters
- Key parameter: window size
- Challenging issue:
  - How to determine the right window size?
  - Slow convergence

THE UNIVERSITY OF TENNESSEE KNOXVILLE 19

Mean Shift Clustering

AICIP RESEARCH

- 1. Initialization: Choose a window/kernel of size *h*, e.g., a flat kernel, and apply the window on each data point, x

$$K(x) = \begin{cases} 1 & \text{if } ||x|| \le h \\ 0 & \text{if } ||x|| > h \end{cases}$$

- 2. Mean calculation: Within each window centered at x, compute the mean of data, where Ω_x is the set of points enclosed within window h

$$m(x) = \frac{\sum_{s \in \Omega_x} K(s-x)s}{\sum_{s \in \Omega_x} K(s-x)}$$

- 3. Mean shift: Shift the window to the mean, i.e., x=m(x), where the difference m(x)-x is referred to as the mean shift.
- 4. If $||m(x)-x||>\varepsilon$, go back to step 2.

THE UNIVERSITY OF TENNESSEE KNOXVILLE 20

AICIP RESEARCH

Original | RGB plot
k=10, 24, 64 (kmeans)
Init (random | uniform), k=6
Mean shift (h=30 → 14 clusters)
Mean shift (h=20 → 35 clusters)

21