Lecture 23

Clustering (4)

Outline

- Clustering as Density Estimation
- Mixture Density Estimate
- Non-parametric density estimate: Parzen Windows

(C) 2001 by Yu Hen Hu

Clustering and Density Estimate

- Probability density function p(x) describes the distribution of samples in the feature space.
- The clustering results $w_m(t) \propto p(x)$
- p(x) can also be estimated using
  - Parametric method: mixture density model
  - Non-parametric method: parzen window

(C) 2001 by Yu Hen Hu

Mixture Density Estimation

$$p(x|\theta) = \sum_{i=1}^{c} p(x|\theta(i),i)p(i)$$

- Assume # of clusters "c" is known. Want to estimate the <u>Prior Probabilities</u> $\{p(i); 1 \le i \le c\}$, and $\theta = [\theta(1), \cdots, \theta(c)]$.
- Let $\{x_k; 1 \le k \le n\}$ be drawn from the mixture density, <u>Log-Likelihood</u> $L(x) = \log\{\prod p(x_k | q)\} = \sum_{k=1}^{n} \log p(x_k|\theta)$

$$\nabla_{\theta(i)} L(x) = \sum_{k=1}^{n} \frac{1}{p(x_k | q)} \{\nabla_{\theta(i)} p(x_k|\theta(i),i)p(i)\}$$

$$= \sum_{k=1}^{n} \frac{p(x_k | q(i),i)\, p(i)}{p(x_k | q)} \nabla_{\theta(i)} [\log p(x_k|\theta(i),i)]$$

(C) 2001 by Yu Hen Hu

Mixture of Gaussian Densities

- $p(x|\mu(i), \Sigma(i), i) = (2\pi|\Sigma|)^{1/2}\exp\{-[|x-\mu(i)|^T\Sigma^{-1}|x-\mu(i)|]/2\}$
- θ (μ and Σ) can be estimated as:

$$\hat{\mu}(i) = \frac{\hat{p}(i \mid x_k, \hat{\mu}(i), \hat{\Sigma}(i))}{\sum_{k=1}^{N} \hat{p}(i \mid x_k, \hat{\mu}(i), \hat{\Sigma}(i))} \mathrm{x}_k$$

$$\hat{\Sigma}(i) = \frac{\hat{p}(i \mid x_k, \hat{\mu}(i), \hat{\Sigma}(i))}{\sum_{i=1}^{N} \hat{p}(i \mid x_k, \hat{\mu}(i), \hat{\Sigma}(i))} (x_k - \hat{\mu}(i))(x_k - \hat{\mu}(i))^T$$

- Parameters appear in both sides of the equations $\Rightarrow$ iterative solution is required.

(C) 2001 by Yu Hen Hu

ML Estimates of Mixture Parameters

- ML estimate of class i prior probability:

$$\hat{p}(i) = \frac{1}{n}\sum \hat{p}(i \mid \mathrm{x}_k, \hat{\mu}(i), \hat{\Sigma}(i));$$

$$\hat{p}(i \mid \mathrm{x}_k, \hat{\mu}(i), \hat{\Sigma}(i)) = \frac{p(\mathrm{x}_k \mid i, \hat{\mu}(i), \hat{\Sigma}(i))\,\hat{p}(i)}{\sum_{j=1}^{c} p(\mathrm{x}_k \mid i, \hat{\mu}(i), \hat{\Sigma}(i))\,\hat{p}(i)}$$

$$= \frac{|\hat{\Sigma}(i)|^{-1/2}\exp\{-\frac{1}{2}[|\mathrm{x}_k - \hat{\mu}(i)|^T\ \hat{\Sigma}^{-1}(i)\,|\mathrm{x}_k - \hat{\mu}(i)|]\}\,\hat{p}(i)}{\sum_{j=1}^{c} |\hat{\Sigma}(i)|^{-1/2}\exp\{-\frac{1}{2}[|\mathrm{x}_k - \hat{\mu}(i)|^T\ \hat{\Sigma}^{-1}(i)\,|\mathrm{x}_k - \hat{\mu}(i)|]\}\,\hat{p}(i)}$$

(C) 2001 by Yu Hen Hu

Parzen Windows

- A non-parametric method using interpolation functions
- Let φ(u) be such that φ(u) ≠ 0 and $\int \varphi(u)du = 1$

  {x(k); 1 ≤k ≤ N} are drawn from unknown density p(x). Set $V_N = h_N^d$ where h_N is a smoothing parameter and d is the dimension of x. Then the estimate of p(x) is:

$$\hat{p}_N(\mathrm{x}) = \frac{1}{NV_N}\sum_{k=1}^{N}\varphi(\mathrm{x}-\mathrm{x}(k))/h_N$$

- Conditions: $\lim_{n\to\infty} V_N = 0$, and $\lim_{n\to\infty} NV_N = \infty$.
- Example of window function: $\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$