北京交通大学

图像处理与机器学习

Digital Image Processing and Machine Learning

主讲人：黄琳琳

电子信息工程学院

基本概念

- 举例：手写字符识别 （Handwritten Character Recognition）

$$y_k(\vec{x}) = y_k(\vec{x}, \vec{w})$$ 贝叶斯决策

$$y_k(\vec{x}) = \boxed{p(\omega_k \mid \vec{x})}$$ ？

第五章　贝叶斯决策

- 基本概念
- 贝叶斯决策
- 判别函数
- 概率密度估计

判别函数

- 判别函数 $y_k(\vec{x}) = p(\omega_k \mid \vec{x})$

$$y_k(\vec{x}) = p(\vec{x} \mid \omega_k)\, p(\omega_k) \qquad y_k(\vec{x}) = -R(\alpha_i \mid \vec{x})$$

✓ 先验概率密度估计

✓ 类条件概率密度估计

如何估计概率密度？

判别函数

- ✓ 先验概率密度估计

  -- 从训练样本估计或者假设

- ✓ 类条件概率密度估计

  -- 参数法（parametric）

  -- 非参数法（Non-parametric）

  -- 半参数法（Semi-parametric）

判别函数

◆ 参数法

✓ 假定概率密度函数形式，估计函数中的参数

$$p(\vec{x}\mid\omega_i) = p(\vec{x}\mid\theta_i) \qquad \textit{Gaussian, Gamma, Bernouli}$$

✓ 参数估计的方法

-- 极大似然估计 （ Maximum-Likelihood Estimation ， ML）

-- 贝叶斯估计 （Bayesian Estimation， BE）

概率密度估计

- 极大似然估计（Maximum Likelihood estimation,ML）

  -- 假设已知概率密度函数形式 $p(\omega|\vec{x},\theta)$，需要估计其中的参数 θ

  -- 设 $\vec{x}_1,\ldots\vec{x}_n$ 为数据集D中 随机抽取的n个相互独立的样本。

$$p(D|\theta)=p(\vec{x}_1,\vec{x}_2\ldots\vec{x}_n|\theta) \qquad p(D|\theta)=\prod_{j=1}^{n}p(\vec{x}_j|\theta)$$

$p(D|\theta)$ 为待估计参数 θ 的似然函数

极大似然估计：寻求 θ 估计值 $\hat{\theta}$ ，使得似然函数最大

概率密度估计

◆ 极大似然估计 （ Maximum likelihood estimation ）

$$Maximize \quad p(D|\theta) \qquad p(D|\theta) = \prod_{j=1}^{n} p(\vec{x}_j|\theta)$$

$$l(\theta) \equiv \ln p(D|\theta) \qquad l(\theta) = \sum_{j=1}^{n} \ln p(\vec{x}_j|\theta)$$

$$Maximize \quad l(\theta) \qquad \nabla_\theta l(\theta) = 0$$

$$\nabla_\theta l(\theta) = \sum_{j=1}^{n} \nabla_\theta \ln p(\vec{x}_j|\theta) = 0$$

$$\downarrow$$

$$\hat{\theta}$$

概率密度估计

◆ ML算法举例：高斯分布情况（The Gaussian case）

-- 已知概率密度函数为高斯分布，估计均值 $\vec{\mu}$ 和方差 Σ

$$p(\vec{x}\,|\,\vec{\mu}) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp[-\frac{1}{2}(\vec{x}-\vec{\mu})^t \Sigma^{-1} (\vec{x}-\vec{\mu})]$$

以一维为例，估计 μ 和方差 σ^2

$$\theta_1=\mu,\quad \theta_2=\sigma^2 \qquad \vec{\theta}=(\theta_1,\theta_2)$$

$$l(\vec{\theta}) = \ln p(x\,|\,\vec{\theta}) = -\frac{1}{2}\ln 2\pi\theta_2 - \frac{1}{2\theta_2}(x-\theta_1)^2$$

概率密度估计

$$l(\vec{\theta}) = \ln p(x|\vec{\theta}) = -\frac{1}{2}\ln 2\pi\theta_2 - \frac{1}{2\theta_2}(x-\theta_1)^2$$

$$\nabla_\theta l(\theta) = \sum_{j=1}^{n} \nabla_\theta \ln p(\vec{x}_j|\theta) = 0$$

$$\nabla_{\vec{\theta}}[\ln p(\vec{x}|\vec{\theta})] = \begin{pmatrix} \dfrac{1}{\theta_2}(\vec{x}-\theta_1) \\ -\dfrac{1}{2\theta_2} + \dfrac{(\vec{x}-\theta_1)^2}{2\theta_2^2} \end{pmatrix} = 0$$

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})^2$$

推广到高维，可得：

$$\boxed{\hat{\vec{\mu}} = \frac{1}{n}\sum_{j=1}^{n} \vec{x}_j}$$

$$\boxed{\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\vec{\mu}})(x_k - \hat{\vec{\mu}})^t}$$

概率密度估计

◆ 贝叶斯估计 （ Bayesian estimation ）

-- 假设参数的条件概率密度 $p(\vec{x}|\vec{\theta})$ 和 先验概率 $p(\vec{\theta})$ 已知

-- 将参数视为**随机变量**，估计参数的**后验概率**

$$p(\omega_k \mid \vec{x}) = \boxed{p(\vec{x} \mid \omega_k)}\, p(\omega_k)$$

同一类的样本集 ***D*** $p(\vec{x} \mid D) = \int p(\vec{x}, \theta \mid D) d\theta$

$$p(\vec{x}, \theta \mid D) = \boxed{p(\vec{x} \mid \theta)} \cdot \boxed{p(\theta \mid D)}$$

概率密度估计

◆ 贝叶斯估计 （ Bayesian estimation ）

-- 假设参数的条件概率密度 $p(\vec{x}|\vec{\theta})$ 和 先验概率 $p(\vec{\theta})$ 已知

同一类的样本集 *D*

$$p(\vec{x},\theta|D)=p(\vec{x}|\theta)\boxed{p(\theta|D)}$$

$$p(\theta|D)=\frac{p(D|\theta)\boxed{p(\theta)}}{\int p(D|\theta)p(\theta)d\theta}$$

$$p(D|\theta)=\prod_{k=1}^{n}\boxed{p(x_k|\theta)}$$

$$p(\vec{x}|D)=\int p(\vec{x},\theta|D)d\theta$$

$$p(\vec{x}|\omega_k)$$

谢 谢

本课程所引用的一些素材为主讲老师多年的教学积累，来源于多种媒体及同事和同行的交流，难以一一注明出处，特此说明并表示感谢！