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

概率密度估计

- 非参数法 （Non-parametric）

  -- 可以估计任意概率密度函数

  - Parzen window
  - K-Nearest Neighbor

概率密度估计

- **概率**

  -- 特征空间中一定区域内样本的比率

假设 N 个样本之中的 k 个 落入特征矢量 x 的周边局部区域 R，则概率为： $P=\dfrac{k}{N}$

- **概率密度** $p(x)$

  -- 假设局部区域 R（体积为V，样本数为k）内等概率密度

$$P=\int_R p(x')dx' \approx p(x)\cdot V \qquad P=\frac{k}{N} \longrightarrow p(x)\approx\frac{k}{N\cdot V}$$

如何估计概率密度 $p(x)$？

概率密度估计

$$p(x) \approx \frac{k}{N \cdot V}$$

- 两种估计方法:

  -- 固定局部区域体积 V 大小， 参数 k 变化

  Parzen Window

  -- 固定参数 k 值，局部区域体积 V 变化

  K-nearest Neighbor

概率密度估计

- Parzen Window

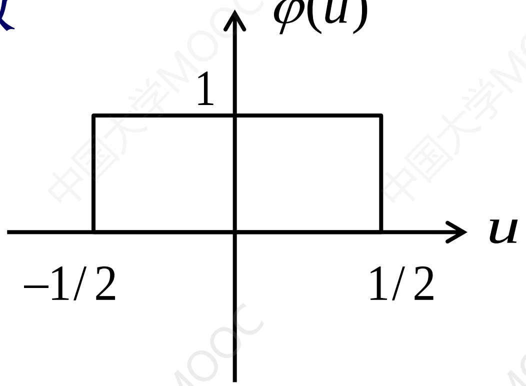
-- 设样本点 x 周边是一个边长为 h 的超立方体 （hypercube）

-- 局部区域体积 V 为 d-维空间： $V = h^d$

-- 单位超立方体 （unit hypercube），称为窗函数

$$\varphi(\vec{u}) = \begin{cases} 1, | u_j | < 1/2, \; j = 1, 2, \ldots d \\ 0, \quad else \end{cases}$$

$$\vec{u} = (u_1, u_2 \ldots u_d)^T$$



概率密度估计

➢ Parzen Window

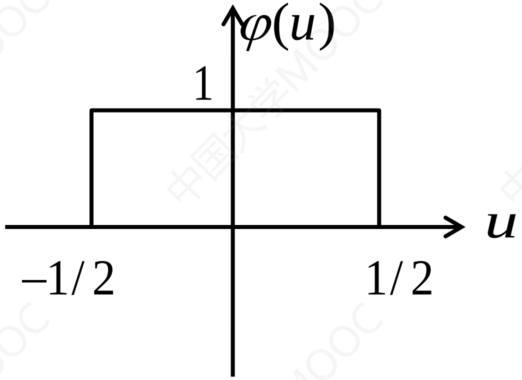
-- 训练数据集 $D:\{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_N\}$

-- 边长为 h的立方体的体积： $V = h^d$

-- 落入立方体的点数k：

$$k = \sum_{n=1}^{N} \varphi\left(\frac{\vec{x}-\vec{x}_n}{h}\right)$$

$$-\frac{1}{2} < \frac{\vec{x}-\vec{x}_n}{h} < \frac{1}{2}$$



$$\varphi(\vec{u}) = \begin{cases} 1, & |u_k| < 1/2, k = 1, 2, \ldots d \\ 0, & else \end{cases}$$

概率密度估计

- Parzen Window

$$p(x) \approx \frac{k}{N \cdot V} \qquad D:\{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_N\}$$

$$k = \sum_{n=1}^{N} \varphi\left(\frac{\vec{x} - \vec{x}_n}{h}\right) \qquad V = h^d$$

$$\boxed{\hat{p}(x) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{h^d}\varphi\left(\frac{\vec{x}-\vec{x}_n}{h}\right)}$$

概率密度估计

- Parzen Window

$$\varphi(\vec{u}) = \begin{cases} 1, |u_j| < 1/2, \ j = 1, 2, \dots d \\ 0, \quad else \end{cases}$$

$$\varphi(\vec{u}) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{\| \vec{u} \|}{2})$$

矩形窗

高斯函数（窗）

$$\hat{p}(x) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{h^d} \varphi(\frac{\vec{x} - \vec{x}_n}{h})$$

$$\hat{p}(x) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{(2\pi h^2)^{d/2}} \exp(-\frac{\left\| \vec{x} - \vec{x}_n \right\|^2}{2h^2})$$

概率密度估计

- 分类的例子

$$p(x) \approx \frac{k}{N \cdot V} \qquad D:\{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_N\}$$

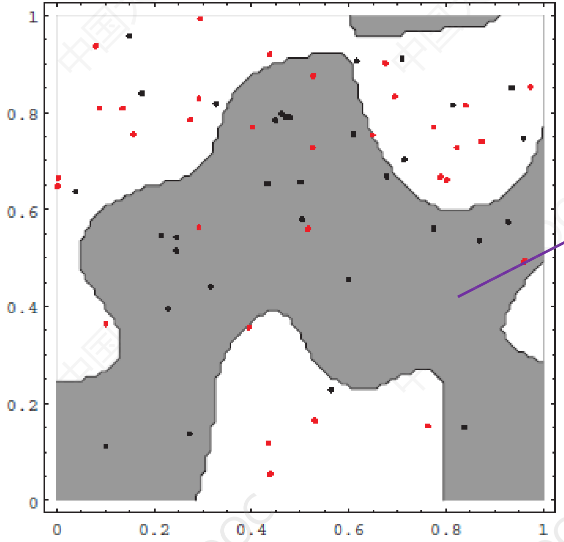
第 **i** 类条件概率密度: $$p(x \mid \omega_i) \approx \frac{k_i}{N_i \cdot V}$$

$$\max_i \ p(x \mid \omega_i) p(\omega_i)$$

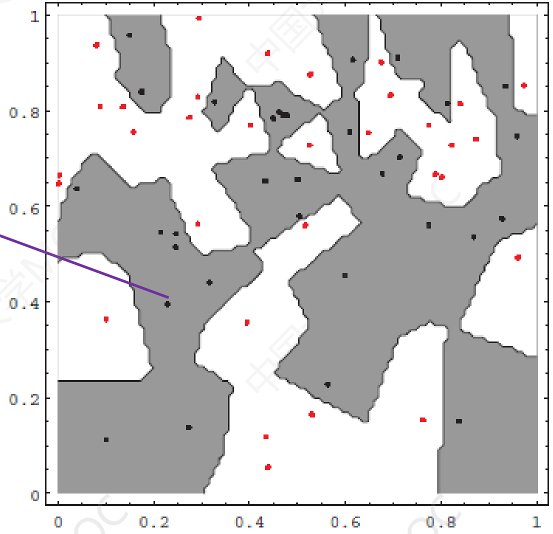
概率密度估计

◆ 分类的例子 $p(x|\omega_j) \approx \dfrac{k_j}{N_j \cdot V}$ $\max_i \; p(x|\omega_i)p(\omega_i)$

Large h



Decision regions



Small h

Large h: low variability, under-fitting

Small h: high variability, over-fitting

概率密度估计

$$p(x) \approx \frac{k}{N \cdot V}$$

- 两种估计方法:

-- 固定局部区域体积 V 大小， 参数 k 变化

-- 固定参数 k 值，局部区域体积 V变化

K-nearest Neighbor

概率密度估计

- K 近邻 （K-nearest Neighbor ）

  -- 固定参数 k 值，局部区域体积 V变化

对于来自于ω_i的 N_i个样本, 扩大(缩小)窗口使得有 k_i 个点落入该窗口.

第 i 类条件概率密度:

$$p(x|\omega_i) \approx \frac{k_i}{N_i \cdot V}$$

K 近邻

$$p(x|\omega_i) \approx \frac{k_i}{N_i \cdot V} \qquad \sum_{i=1}^{M} k_i = k \qquad p(\omega_i) = \frac{N_i}{N} \qquad p(\omega_j) = \frac{N_j}{N}$$

贝叶斯公式

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{\sum_j^M p(x|\omega_j)p(\omega_j)} = \frac{(k_i/N_i)\cdot p(\omega_i)}{\sum_j^M (k_j/N_j)\cdot p(\omega_j)} = \frac{k_i}{\sum_j^M k_j} = \frac{k_i}{k}$$

$$p(\omega_i|x) = \frac{k_i}{k}$$

K 近邻

$$p(x \mid \omega_i) \approx \frac{k_i}{N_i \cdot V} \qquad \sum_{i=1}^{M} k_i = k \qquad p(\omega_i) = \frac{N_i}{N} \qquad p(\omega_j) = \frac{N_j}{N}$$

$$p(\omega_i \mid x) = \frac{k_i}{k}$$

贝叶斯决策： 分到后验概率最大的类，分类错误率最小

To select the class with *most nearest neighbors* among ***k*** will make minimum-error decision.

K 近邻

假设有a、b、c三个样类别的样本，在随机变量 X 的10-近邻中有 3个a类点， 2个b类点，5个c类点，应该将该随机变量归为哪一类分类错误最小？为什么？

$$p(\omega_i \mid x)=\frac{k_i}{k}$$

$$p(\omega_a \mid x)=\frac{3}{10} \qquad p(\omega_b \mid x)=\frac{2}{10} \qquad p(\omega_c \mid x)=\frac{5}{10}$$

根据贝叶斯决策，将此随机变量X归为 c 类，分类错误最小。

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！