

# Pattern Recognition

University of Chinese Academy of Sciences

Fall 2023

Shiming Xiang, Gaofeng Meng

---

## Homework 2

Chenkai GUO

2023.10.30

1. 设一维特征空间中的窗函数  $\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$ , 有  $n$  个样本  $x_i, i = 1, 2, \dots, n$ , 采用宽度为  $h_n$  的窗函数, 请写出概率密度函数  $p(x)$  的 Parzen 窗估计  $p_n(x)$

$$p_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} \varphi\left(\frac{x-x_i}{h_n}\right)$$

2. 给定一维空间三个样本点  $\{-4, 0, 6\}$ , 请写出概率密度函数  $p(x)$  的最近邻 1-NN 估计, 并画出概率密度函数曲线图。

$$p(x) = \begin{cases} \frac{1}{6|x+4|}, & x < -2 \\ \frac{1}{6|x|}, & -2 < x < 3 \\ \frac{1}{6|x-6|}, & x > 3 \end{cases}$$

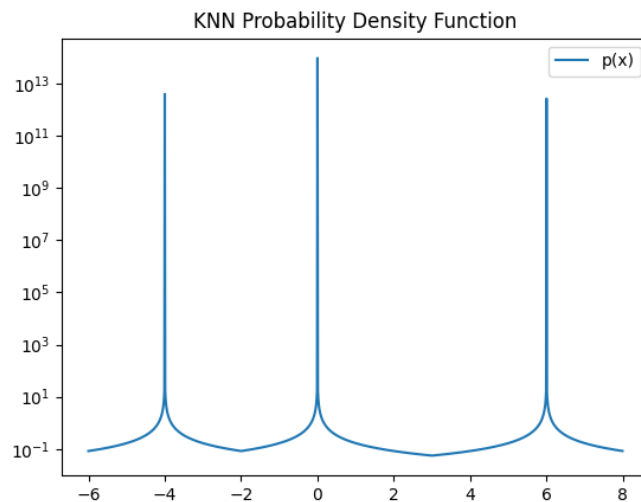


图 1: KNN 估计概率密度函数曲线图

3. 针对概率密度估计问题, 请简述 EM 算法的基本步骤。

- (1) 初始化  $\theta^{old}$
- (2)  $E$  step: 基于初始化的  $\theta^{old}$  和样本集, 估计隐变量的后验分布  $p(\mathbf{z}|\mathbf{x}, \theta^{old})$
- (3)  $M$  step: 基于当前估计的隐变量后验分布  $p(\mathbf{z}|\mathbf{x}, \theta^{old})$  更新参数  $\theta^{new}$

$$\begin{aligned}\theta^{new} &= \arg \max_{\theta} \sum_i E_{p(z_i|x_i, \theta^{old})} [\ln(p(z_i, x_i|\theta^{old}))] \\ &= \arg \max_{\theta} \sum_i \sum_{z_i} p(z_i|x_i, \theta^{old}) \ln(p(z_i, x_i|\theta^{old}))\end{aligned}$$

- (4)  $\theta^{old} \leftarrow \theta^{new}$ , 并重复 2-3 步骤, 直到达成输出条件

4. 对混合高斯模型参数估计问题, 在  $EM$  优化的框架下, 请给出其中的  $Q(\theta, \theta^{old})$  的基本形式。

$$Q(\theta, \theta^{old}) = \sum_i \sum_{z_i:1:k} (p(z_i|x_i, \theta^{old}) \ln(\pi_{z_i}) + p(z_i|x_i, \theta^{old}) \ln(\mathcal{N}(x_i|\mu_{z_i}, \Sigma_{z_i})))$$

5. 针对离散状态离和散观测情形的一阶  $HMM$ , 请描述其学习问题的基本任务。

### 隐马尔科夫模型 (HMM)

假设观测序列由一个不可见的马尔科夫链生成(隐变量), 其状态空间为  $\{z_1, z_2, \dots, z_n\}$ , 观测变量为  $\{x_1, x_2, \dots, x_n\}$ ,  $HMM$  需要估计三组参数:  $\theta = (\pi, \mathbf{A}, \mathbf{B})$ ,  $\mathbf{A}$  为状态转移概率矩阵 ( $\mathbf{A}_{i,j} = P(z_t = j|z_{t-1} = i)$ ),  $\mathbf{B}$  为观测概率矩阵 ( $\mathbf{B}_{i,j} = P(x_t = j|z_t = i)$ ),  $\pi$  为初始状态概率向量 ( $\pi_k = P(z_1 = k)$ )

- (1)  $E$  step: 基于初始化的  $\theta^{old}$  和样本集, 估计隐变量的后验分布  $p(\mathbf{z}|\mathbf{x}, \theta^{old})$

$$q(z_1, \dots, z_n) \equiv p_{\theta}(z_1, \dots, z_n|x_1, \dots, x_n)$$

- (2)  $M$  step: 基于当前估计的隐变量后验分布  $p(\mathbf{z}|\mathbf{x}, \theta^{old})$  更新参数  $\theta^{new}$

$$\theta = \arg \max_{\theta} \sum_z q(z_1, \dots, z_n) \ln(p_{\theta}(x_1, \dots, x_n, z_1, \dots, z_n))$$

- (3) 迭代直至收敛

## 6. 编程题。

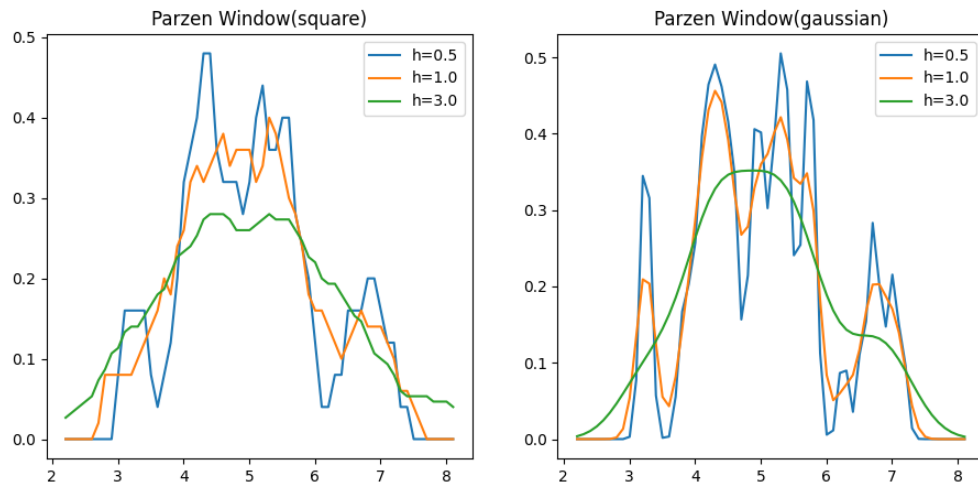


图 2: Parzen 窗方法估计结果 (方窗、高斯窗)

代码见附件(*parzen\_window.py*)