

第二次作业

第一部分：计算题和简答题

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 $p_n(x)$ 。
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。
3. 针对概率密度估计问题, 请简述 EM 算法的基本步骤。
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 $Q(\theta, \theta^{\text{old}})$ 的基本形式。
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。

第二部分：编程题

1. 现有一维空间的 50 个样本点 (实际上, 这些样本点是在 Matlab 中按如下语句生成的: `mu=5; std_var=1; X=mvnrand(mu, std_var, 50);`)。现需要采用 Parzen 窗方法对概率密度函数进行估计。请分别编程实现方窗和高斯窗情形下的概率密度函数估计; 请讨论窗宽的影响, 并画出几种不同窗宽取值下所估计获得的概率密度函数曲线。50 样本点如下:

4.6019,	5.2564,	5.2200,	3.2886,	3.7942,
3.2271,	4.9275,	3.2789,	5.7019,	3.9945,
3.8936,	6.7906,	7.1624,	4.1807,	4.9630,
6.9630,	4.4597,	6.7175,	5.8198,	5.0555,
4.6469,	6.6931,	5.7111,	4.3672,	5.3927,
4.1220,	5.1489,	6.5319,	5.5318,	4.2403,
5.3480,	4.3022,	7.0193,	3.2063,	4.3405,
5.7715,	4.1797,	5.0179,	5.6545,	6.2577,
4.0729,	4.8301,	4.5283,	4.8858,	5.3695,
4.3814,	5.8001,	5.4267,	4.5277,	5.2760