《数据挖掘原理及实践》编程练习

1. 读取 matlab 数据, 通过 Scott 方法建立统计直方图, 随后利用 Parzen-window 非参数方法建立概率密度函数.

Scott 方法计算直方图箱格个数：

$$Bin = \frac{x^H - x^L}{3.49\, s\, n^{-1/3}}$$

其中，x^H 和 x^L 分别表示数据的最大值和最小值；s 表示数据标准差；n 表示数据样本个数。

Parzen-window 非参数方法估计概率密度函数：

$$p(x) = \frac{1}{M}\sum_{m=1}^{M}\kappa(x - x_m)$$

$$\kappa(x - x_m) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x - x_m)^2}{2\sigma^2}\right]$$

要求：计算 Scott 直方图 Bin 个数，画出直方图，并在同一图中做出估计出的 Parzen-window 概率密度函数。

实验数据下载地址：
http://mirel.xmu.edu.cn/course/DM/DataMiningAssignment_PDF_data.mat