

非参数统计第二次作业

作业提交截止日期：2014年11月19日

习题1 研究加拿大工人收入(age.income数据集)的年龄(age)和收入(income)的关系，该数据来源于R程序包“SemiPar”。总共调查了 $n = 205$ 个加拿大工人的年龄(age)和收入(income)，所有工人的都是高中毕业，取 $\log.income = \log(income)$ ，因变量为 $\log.income$ ，协变量为age. 分别运用局部多项式方法和penalized-splines方法来拟合age与 $\log.income$ 之间的函数关系。

习题2 研究ethanol数据集，该数据集来源于R程序包**locfit**，因变量为 NO_x ，协变量为E. 分别运用局部多项式方法和penalized-splines方法来拟合E与 NO_x 之间的函数关系。

- 作业内容要求：**
1. 采用word文档。
 2. 解释一下为什么采用非参数回归，而不是参数回归。
 3. 分别简要叙述一下局部多项式方法和penalized-splines方法的具体内容。
 4. 简单叙述一下所采用的选择smoothing parameter的方法。
 5. 简单分析一下最后拟合的结果。
 6. 文末附上主要的代码。