**问题1**

**数据集信息：**

MN DoT ATR站301的每小时94号州际公路西行交通量，大约在明尼阿波利斯和明尼苏达州圣保罗之间。包括每小时天气特征和节假日，以影响交通量。

**属性信息：**

假日分类美国国定假日加明尼苏达州州节假日  
数值  
凯尔文降雨的平均平均温度 \_1h每小时  
下雪时发生的降雨的毫米数（毫米）\_1h 每小时 中下雪时发生的毫米的毫米数的毫米  
\_all数值云量的百分比  
weather\_main分类当前天气的简短文字说明  
weather\_description分类当前天气的 较长文字说明  
date\_time DateTime在本地CST时间中收集的数据的小时数  
traffic\_volume数字每小时I-94 ATR 301报告的西行流量

**问题2**

数据集信息：

在[Cortez and Morais，2007]中，首先使用ln（x + 1）函数对输出“区域”进行了转换。

然后，应用了几种数据挖掘方法。拟合模型后，对输出进行

ln（x + 1）逆变换的后处理。使用了四种不同的输入设置

。使用10倍（交叉验证）x 30次运行进行实验。

测量了两个 回归指标：MAD和RMSE。高斯支持向量机（SVM）

仅接受4种直接天气条件（温度，RH，风和雨），获得了最佳MAD值：

12.71±0.01（使用t型学生分布的均值和置信区间在95％以内）。将

最好的RMSE获得了由幼稚的平均预测。回归误差曲线分析

（REC）表明，SVM模型可以在较低的允许误差范围内预测更多示例。实际上，

SVM模型可以预测较小的火灾，而后者是大多数情况。

属性信息：

有关更多信息，请阅读[Cortez and Morais，2007]。

1. Montesinho公园地图内的X-x轴空间坐标：1到9

2. Montesinho公园地图内的 Y-y轴空间坐标：2到9

3.每年的月份-月：“ jan”到“ dec'4.

每天-星期几：从“周一”到“星期日”

5.

FFMC

- FWI系统的FFMC指数：18.7至96.20 6. DMC-FWI系统的DMC指数：1.1至291.3 7. DC- FWI系统的DC指数：7.9至860.6

8. ISI-FWI系统的ISI指数：0.0至56.10

9. temp- 摄氏温度：2.2至33.30

10. RH-相对湿度（％）：15.0至100

11。风-以km / h为单位的风速：0.40至9.40

12.雨量-外部雨量，单位为mm / m2：0.0到6.4

13.面积-森林的烧毁面积（以公顷为单位）：0.00到1090.84

（此输出变量非常偏向0.0，因此

使用对数变换）。

**问题3**

数据集信息：

提取是由Barry Becker从1994年人口普查数据库中进行的。使用以下条件提取了一组合理的干净记录：（（AAGE> 16）&&（AGI> 100）&&（AFNLWGT> 1）&&（HRSWK> 0））

预测任务是确定一个人的年收入是否超过50K年。

属性信息：

属性列表：

> 50K，<= 50K。

年龄：连续。

工作类别：私人，自营非收入，自营收入，联邦政府，地方政府，州政府，无薪，从未工作过。

fnlwgt：连续。

教育程度：学士，部分大学，11年级，高中毕业生，教授学校，Assoc-acdm，Assoc-voc，9、7-8-8、12，硕士，1-4至10，博士学位，5-6至学前班。

教育人数：连续。

婚姻状况：已婚公民配偶，离婚，未婚，分居，丧偶，已婚配偶缺席，已婚AF配偶。

职业：技术支持，工艺修理，其他服务，销售，执行管理，专业教授，搬运清洁工，机器操作检查，行政助理，农家捕鱼，运输，私人住宅serv，保护serv，武装部队。

关系：妻子，独生子女，丈夫，亲戚，其他亲戚，未婚。

种族：白色，亚洲人-帕斯岛，亚洲印第安人-爱斯基摩人，其他，黑人。

性别：女，男。

资本收益：连续。

资本损失：连续。

每周小时：连续。

祖国：美国，柬埔寨，英国，波多黎各，加拿大，德国，美国外围地区（关岛-USVI等），印度，日本，希腊，南美，中国，古巴，伊朗，洪都拉斯，菲律宾，意大利，波兰，牙买加，越南，墨西哥，葡萄牙，爱尔兰，法国，多米尼加共和国，老挝，厄瓜多尔，台湾，海地，哥伦比亚，匈牙利，危地马拉，尼加拉瓜，苏格兰，泰国，南斯拉夫，萨尔瓦多，特立尼达和多巴哥，秘鲁，洪，荷兰霍兰。

**问题4**

数据集信息：

消费者参与度的可变性通过主成分分析进行了分析，突出了使用Facebook Live引起的变化。通过研究不同时间段（每小时，每天和每月）不同参与度指标的平均值来分析季节性成分。最后，我们确定统计异常值职位，并根据其销售方式和活动对其进行定性分析。

属性信息：

status\_id

status\_type

status\_published

num\_reactions

num\_comments

num\_shares

num\_likes

num\_loves

num\_wows

num\_hahas

num\_sads

num\_angrys

问题5