**《数据挖掘与统计决策课程》SVM作业内容与要求**

**一、作业题：**

**借助Python软件熟悉SVM算法的原理。**

**1软间隔线性SVM**

1. **将data1.mat数据集可视化，观察数据的特点，是否存在异常点。**
2. **利用软间隔线性SVM对data1.mat数据集分类。取不同的参数，，比如，观察其对分类结果的影响。将10,50,100,1000对数据集的分类结果用图形展示（数据加决策曲线）。**

**2 基于高斯核函数的非线性SVM**

1. **将data2.mat数据集可视化，观察数据的特点，是否存在异常点。**
2. **利用基于高斯核函数的非线性SVM对data2.mat数据集进行分类，将分类结果用图形展示（数据加决策曲线）。**
3. **data3.mat数据集，其中变量X,y为训练集，Xval和yval为测试集。将data3.mat的训练集可视化，观察数据的特点，是否存在异常点。**
4. **对data3.mat数据集，在测试集上进行交叉验证，选取最好的参数和。要求遍历和的全部组合（,一共有64种不同组合），找出最好的和。根据最好的和，利用训练样本训练基于高斯核函数的非线性SVM，得到最优分类曲线。将训练样本的分类结果用图形展示（数据加决策曲线）。**

**交作业时间：5月12日（星期日）晚上之前。**

**二、作业要求：以Word（不压缩，如打不开文件，没有评定成绩）**

**文件的命名规则为：**

**强基-序号-姓名-作业（聚类）.doc(x)**

**双学位-序号-姓名-作业（聚类）.doc(x)**

**应数-序号-姓名-作业（聚类）.doc(x)**

**统计-序号-姓名-作业（聚类）.doc(x)**

**信管-序号-姓名-作业（聚类）.doc(x)**

**信计-序号-姓名-作业（聚类）.doc(x)**

**每个作业Word文件内容包括：**

**1）数据处理与分析**

**2）数学建模过程论述**

**3）软件辅助过程（代码，参数选择、数据项属性设置等，界面）**

**4）软件辅助结果分析（结果图形表示、结果数据分析）。**

**3.以班为单位，由学习委员收齐后发给老师邮箱：**[**liuxl@scut.edu.cn**](mailto:liuxl@scut.edu.cn)**。**