**《数据挖掘与统计决策课程》BP神经网络作业内容与要求**

**一、作业题：**

借助Python软件对数据集**建立BP分类模型**

（1）按以下代码产生4个人工数据集：blobs，circles，moons和blocks，每个数据集都是二分类问题（标签用0和1表示）。在同一平面的4个不同图中描绘4个数据集的散点图，观察各个数据集的特点。

def generateData(n):

np.random.seed(12046)

blobs = make\_blobs(n\_samples=n, centers = [[-2, -2], [2, 2]])

circles = make\_circles(n\_samples=n, factor=.4, noise=.05)

moons = make\_moons(n\_samples=n, noise=.05)

blocks = np.random.rand(n, 2) - 0.5

y = (blocks[:, 0] \* blocks[:, 1] < 0) + 0

blocks = (blocks, y)

# 由于神经网络对数据的线性变换不稳定，因此将数据做归一化处理

scaler = StandardScaler()

blobs = (scaler.fit\_transform(blobs[0]), blobs[1])

circles = (scaler.fit\_transform(circles[0]), circles[1])

moons = (scaler.fit\_transform(moons[0]), moons[1])

blocks = (scaler.fit\_transform(blocks[0]), blocks[1])

return blobs, circles, moons, blocks

（2）利用三层神经网络[4,4,2]结构（第一隐藏层4个节点，第二隐藏层4个节点，输出层2个节点），隐藏层采用sigmoid激活函数，分别对以上4个数据集进行训练，并对数据集进行预测，得出分类正确率。在同一平面的4个不同图中描绘：预测的分类结果（错分的数据点可以采用不同颜色，以加以区分）。

（3）将训练过程的模型损失记录下来，在同一坐标系中画出4个数据集，模型损失随着迭代次数的变化曲线，并进行分析。

**交作业时间：6月13日（星期四）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（BP神经网络）.doc(x)**

**双学位-序号-姓名-作业（BP神经网络）.doc(x)**

**应数-序号-姓名-作业（BP神经网络）.doc(x)**

**统计-序号-姓名-作业（BP神经网络）.doc(x)**

**信管-序号-姓名-作业（BP神经网络）.doc(x)**

**信计-序号-姓名-作业（BP神经网络）.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)**。**