

第二次作业（在网上提交电子版）

（本次作业覆盖第 4、5、6 章内容）

第一部分简述题

1. 请简述谱聚类的算法思想，写出一种谱聚类算法的计算步骤，指出哪些因素会影响谱聚类的结果。
2. 请指出数据聚类存在哪些挑战性问题。
3. 关于前向神经网络，请描述误差反向传播算法的原理；结合三层前向神经网络，给出权重更新公式。
4. 请描述 CNN 网络结构，并结合前向计算过程描述网络待学习的参数。
5. 请给出多维缩放（MDS）算法的问题描述（Problem Formulation）、学习准则和详细的算法推导过程。
6. 请给出 LLE、LE 和 Isomap 算法的异同点，在此基础上给出流形学习的统一学习模型，并指出流形学习所存在的挑战性问题。
7. 请简述局部投影保持(LPP)算法的思想，写出算法的计算步骤。

第二部分：计算机编程

第一题

首先，使用 Matlab 生成如下数据集 (注意:与第一次的作业之数据略有不同):

```
close all;
Sigma = [1, 0; 0, 1];
mu1 = [1, -1];
x1 = mvnrnd(mu1, Sigma, 200);
mu2 = [5, -4];
x2 = mvnrnd(mu2, Sigma, 200);
mu3 = [1, 4];
x3 = mvnrnd(mu3, Sigma, 200);
mu4 = [6, 4];
x4 = mvnrnd(mu4, Sigma, 200);
mu5 = [7, 0.0];
x5 = mvnrnd(mu5, Sigma, 200);

% Show the data points
```

```

plot(x1(:,1), x1(:,2), 'r. '); hold on;
plot(x2(:,1), x2(:,2), 'b. ');
plot(x3(:,1), x3(:,2), 'k. ');
plot(x4(:,1), x4(:,2), 'g. ');
plot(x5(:,1), x5(:,2), 'm. ');

```

现在假定 x_1 中的 200 个点属于第一类， x_2 中的 200 个点属于第二类， x_3 中的 200 个点属于第三类， x_4 中的 200 个点属于第四类， x_5 中的 200 个点属于第五类。请写出一个 K 均值聚类算法（给出代码），完成对数据的聚类，报告聚类精度和 NMI 指标。

第二题

数据同第一题。

现在假定 x_1 中的 200 个点属于第一类， x_2 中的 200 个点属于第二类， x_3 中的 200 个点属于第三类， x_4 中的 200 个点属于第四类， x_5 中的 200 个点属于第五类。请编写一个 Matlab 程序（其它代码也可以），采用三层前向神经网络来实现分类。

要求：

- (1) 每类随机采用 60% 的数据做训练样本，采用剩余的 40% 样本作测试。
- (2) 当结构固定时，画出测试精度随学习率 η 变化的曲线。
- (3) 画出测试精度随隐含层结点数变化的曲线。

第三题

本次编程主要涉及到流形学习方法的实现。首先，采用如下 matlab 代码生成一个瑞士蛋卷数据集，它包含 1000 个三维样本点：

```

N=1000;
%加高斯噪声
noise = 0.001*randn(1,N);
%标准的 swiss roll 数据
tt = (3*pi/2)*(1+2*rand(1,N)); height = 21*rand(1,N);
X = [(tt+ noise).*cos(tt); height; (tt+ noise).*sin(tt)];

%显示数据
point_size = 20;
figure(1)
cla

scatter3(X(1,:),X(2,:),X(3,:), point_size,tt,'filled');
view([12 12]); grid off; axis off; hold on;
axis off;
axis equal;
drawnow;

```

数据显示如下图所示：



请编程实现 Isomap、LLE 和 LE 算法，并将上述数据降低至二维空间，其中，数据图采用 K -近邻方式进行构造。请显示出 K 取不同值时上述三种方法所对应的降维结果，并给出总结性评述。