



本科实验报告

课程名称: 人工神经网络与深度学习基础

实验项目: 基于 SOM 网的植物聚类实验

实验地点: 教学楼 6209

专业班级: 计算机类 3 班 学号: 25020007105

学生姓名: 施家鑫

2025 年 12 月 2 日

学生姓名	施家鑫	实验成绩	
实验名称	基于 SOM 网的植物聚类实验		
实验目的和要求 (必填)			
<p>实验目的:</p> <p>在 Matlab 中通过 SOM 网实现对 10 种植物的聚类。</p> <p>实验要求:</p> <p>掌握 SOM 网的使用方法</p>			
实验内容和原理 (必填)			
<p>本实验利用自组织映射神经网络 (SOM) 对 10 种常见植物进行无监督聚类分析。通过给出的 10 种植物的 8 项生物学特征 (如是否开花、是否常绿、高度等级等)，构建特征向量并输入 SOM 网络，探索植物在特征空间中的自然分组规律，理解 SOM 在其中的应用机制。</p>			
主要仪器设备	台式或笔记本计算机		
实验记录(写出实验内容中的程序代码和运行结果) (可分栏或加页)			
<p>现给定特征数据及植物名称表如下:</p> <p>八项特征分别为:</p> <p>是否开花 是否常绿 是否木本 是否水生 是否耐旱 高度等级 1 - 10 是否具香气 寿命等级 1 - 10</p> <pre> D = [1 0 0 0 0 3 1 6; % 玫瑰 1 1 0 0 1 2 0 8; % 仙人掌 0 1 1 0 1 8 0 10; % 松树 1 1 0 0 0 6 0 7; % 竹子 1 0 0 1 0 5 0 4; % 荷花 1 0 0 0 0 1 0 2; % 小麦 0 1 0 0 0 2 0 5; % 蕨类 1 0 1 0 0 7 0 9; % 苹果树 1 0 0 0 0 3 1 6; % 薰衣草 0 1 0 0 1 1 0 3 % 苔藓]; </pre> <pre> plants = {'玫瑰', '仙人掌', '松树', '竹子', '荷花', '小麦', '蕨类', '苹果树', '薰衣草', '苔藓'}; </pre>			

将每个特征（每一行）线性缩放到 [0, 1] 区间，归一化处理。

```
P = normalize(D, 'range');
```

构建 SOM 网络并进行训练：

```
net = newsom(minmax(P), [3 3]);
net.trainParam.epochs = 200;
net = train(net, P);
```

将原始特征数据重新带入训练好的模型中仿真：

```
y = sim(net, P);
class_id = vec2ind(y);
unique_classes = unique(class_id);
for c = unique_classes
    fprintf("类别 %d: ", c);
    idx = find(class_id == c);
    for k = idx
        fprintf("%s ", plants{k});
    end
    fprintf("\n");
end
```

执行输出效果：

```
>> SOM_Classify
类别 1: 竹子 荷花
类别 2: 松树
类别 3: 蕨类 苔藓
类别 4: 苹果树
类别 6: 仙人掌
类别 8: 玫瑰 薰衣草
类别 9: 小麦
```

实验结果和分析

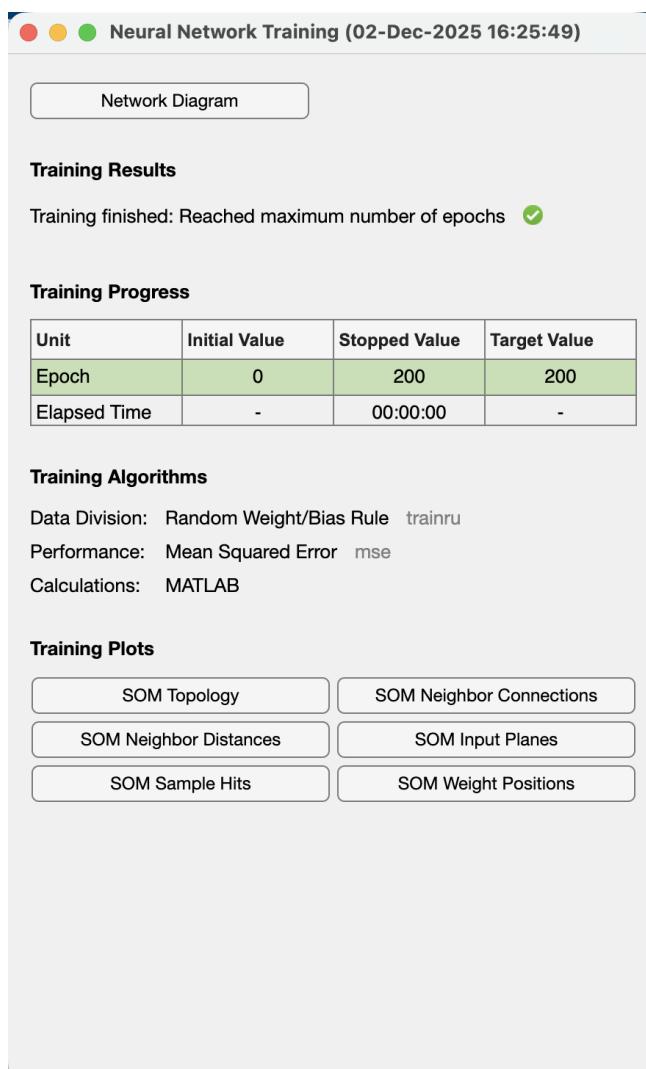
多次执行结果如图所示，每一次重新执行都会出现不同的结果，原因在于 SOM 网络的初始权值是随机初始化的：

```
>> SOM_Classify
类别 1: 竹子 荷花
类别 2: 松树
类别 3: 蕨类 苔藓
类别 4: 苹果树
类别 6: 仙人掌
类别 8: 玫瑰 薰衣草
类别 9: 小麦
```

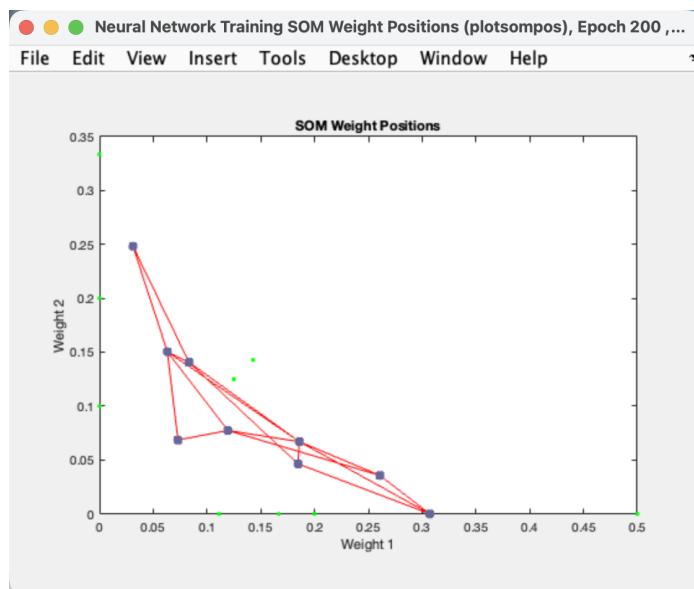
```
>> SOM_Classify
类别 1: 松树 竹子 苹果树
类别 3: 蕨类
类别 5: 玫瑰 薰衣草
类别 6: 仙人掌 苔藓
类别 7: 荷花
类别 8: 小麦
```

```
类别 1: 松树 苹果树
类别 3: 玫瑰 仙人掌 薰衣草
类别 4: 竹子
类别 6: 蕨类
类别 7: 荷花
类别 9: 小麦 苔藓
```

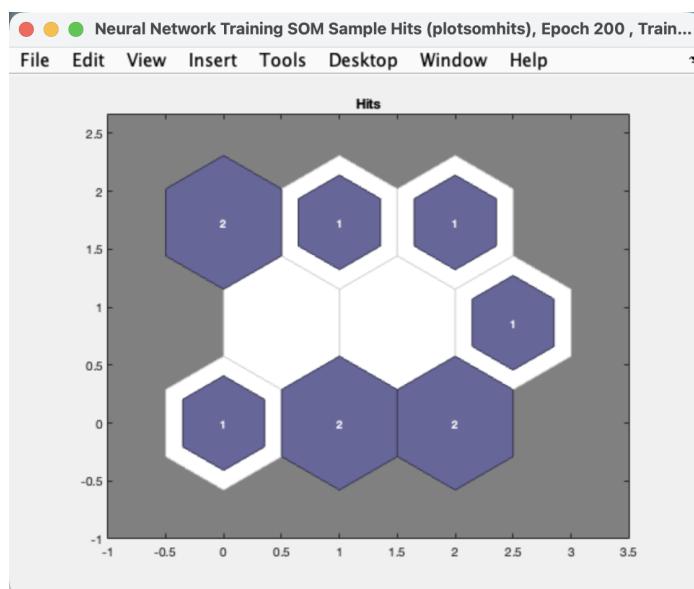
得到如下结果：迭代终止条件为达到了设定的最大迭代次数。



其中一次的权重图：



SOM Sample Hits: SOM 采样命中，即各类别中植物的数量：



心得体会（遇到的问题和解决方法）

在本次基于 SOM 网的植物聚类实验中，我不仅加深了对无监督学习和神经网络聚类原理的理解，也在实践中遇到了若干关键问题，通过查阅资料、调试代码和反复验证，逐渐对这些问题有了进一步的认识，并且明确了可能的优化方向。

1. 聚类结果不稳定，每次运行结果不同

最初运行代码时，发现每次执行后植物的分组结果都不一致。例如，有时“玫瑰”

和“薰衣草”被分到同一类，有时却被分到不同类别，这严重影响了实验的可重复性和结论的可信度。

原因分析：

由于 SOM 网络的初始权值是随机生成的，而 SOM 的训练过程对初始值较为敏感，尤其在样本量较少（仅 10 种植物）的情况下，微小的初始差异可能导致最终聚类结构显著不同。

2. 聚类结果不符合植物学常识

我发现“玫瑰”和“薰衣草”（二者均开花且具香气）并未被聚为一类，反而“仙人掌”和“松树”被分在一起，而“荷花”（唯一水生植物）有时竟与“小麦”归为一类，这明显违背常识。

原因分析：

仔细检查数据后发现，原始特征中“高度等级”和“寿命等级”为 1-10 的数值，而其他特征是 0/1 二值变量。最初未对特征数据表进行归一化处理，导致 1-10 数值的特征项权重占比过大，影响结果。

解决方法：

对输入数据进行按特征归一化处理，使用 `normalize(D, 'range')` 将每个特征（每一行）线性缩放到 [0, 1] 区间。

3. 输出神经元数不合理导致聚类过于稀疏

使用 4×4 (16 个神经元) 的 SOM 网格时，常出现“一个神经元只对应一个植物”的情况，导致类别过多、聚类意义不强。

解决方法：

尝试更小的网格结构，如 3×3 (9 类) 或 3×2 (6 类)，聚类结果更加紧凑且具有实际意义。