

**提示：**如果问题的陈述不能完全满足解题条件，那么你可以做出某些假设。但是，这些假设必须是合理的，而且你也需要在作业中对这些假设给出清晰地说明和解释。

**原则：**你必须**独立完成**本课程的所有作业，除了课程设计内容之外，本课程没有需要小组协作完成的作业。一般来说，你可以与同学们讨论完成作业过程中遇到的问题，但是作业的具体解决方法（包括作业本身）必须是自己独立完成的。

### 问题 1: (40 分)

在数据文件 PRHW#3Data.mat 中包含六个矩阵  $c_1$ 、 $c_2$ 、 $c_3$ 、 $t_1$ 、 $t_2$  和  $t_3$ ，每个矩阵的维数均为  $500 \times 2$ ，即 500 个样本，每个样本 2 维特征。 $c_i$  矩阵中为第  $i$  类的训练样本集合， $t_i$  矩阵中为第  $i$  类的测试样本集合。使用  $k$  近邻分类器解决该模式分类问题

- 1) 给出对应最佳分类结果的  $k$  值和总分类错误率。
- 2) 给出并解释当  $k=1$ 、最佳  $k$  值和 50 时分类器的分类错误率。

### 问题 2: (60 分)

考虑简单的最近邻编辑算法：

```

1 begin initialize  $j = 0, \mathcal{D} = \text{data set}, n = \#\text{prototypes}$ 
2   construct the full Voronoi diagram of  $\mathcal{D}$ 
3   do  $j \leftarrow j + 1$ ; for each prototype  $x'_j$ 
4     Find the Voronoi neighbors of  $x'_j$ 
5     if any neighbor is not from the same class as  $x'_j$  then mark  $x'_j$ 
6   until  $j = n$ 
7   Discard all points that are not marked
8   Construct the Voronoi diagram of the remaining (marked) prototypes
9 end
```

- 1) 通过反例证明该算法不会产生最小的点集（提示：考虑一个问题，来自每个类别的样本点均被约束在二维笛卡尔网格的交叉点上）
- 2) 建立一个顺序编辑算法。在算法中，依次考虑每个样本点，并且在考虑下一个样本点之前决定是否保留该样本点。证明你所给出的算法是否依赖于所考虑样本点的顺序。