

# 特征选择与KNN聚类 实验设计

人工智能1903班 王萱

人工智能1904班 魏良俊 吴洪涛 石纹昌

# 一、实验内容

- 1. 利用智能优化算法对wine数据集进行特征选择，并分析不同算法的效果。
- 2. 对KNN方法中K的取值进行灵敏性分析。

## 二、使用模拟退火进行特征选择

### ■ 1. 数据编码

共13个特征，将解编码成：

[0,1,1,0,0,1,0,1,1,0,1,0,0]

1表示选择这个特征，0表示不选择这个特征

## 二、使用模拟退火进行特征选择

### ■ 2. 将解进行解码、归一化并进行K近邻分类

```
13
14 #分割数据集, 比例为 训练集: 测试集 = 4:6
15 X_train,X_test,y_train,y_test=train_test_split(wine_dataset['data'],wine_dataset['target'],test_size=0.6,random_state=0)
16 #构建knn分类模型, 并指定 k 值
17 KNN=KNeighborsClassifier(n_neighbors=8)
18 #使用训练集训练模型
19 KNN.fit(X_train,y_train)
20 #评估模型的得分
21 score=KNN.score(X_test,y_test)
22 print(score)
```

## 二、使用模拟退火进行特征选择

### ■ 3. 是否转移和生成新解

生成新解：在当前解的基础上随机选择一位进行反转，不包括上次反转的解。

相当于带了一个小禁忌表。

## 二、使用模拟退火进行特征选择

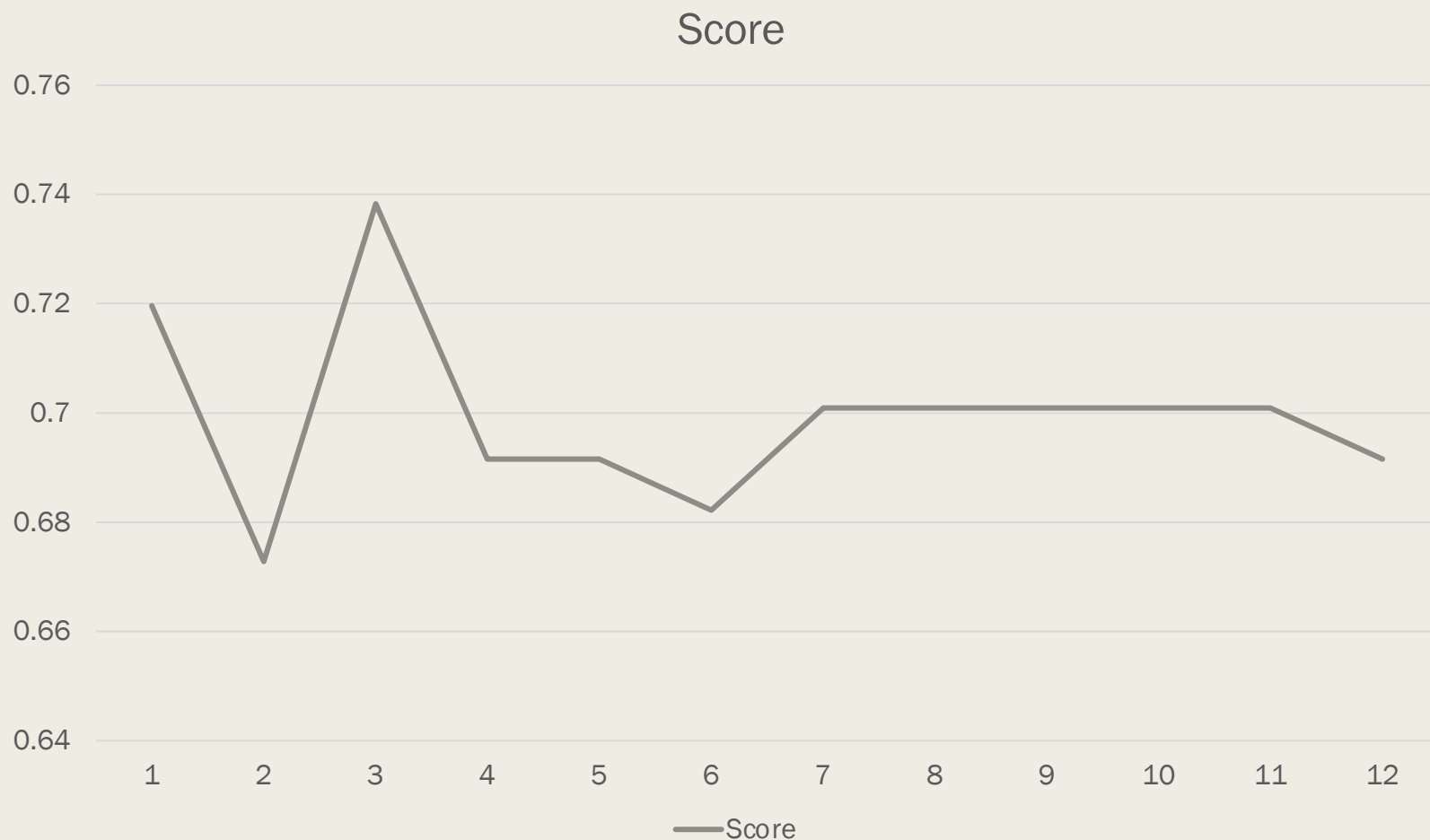
### ■ 4. 使用模拟退火进行特征选择和使用所有特征对比

[1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1]

Score	带特征选择	所有特征
K=3	0.9389	0.7383
K=8	0.8813	0.7009

### 三、不同K值对比

K值	Score
1	0.7196
2	0.6728
3	0.7383
4	0.6915
5	0.6915
6	0.6822
7	0.7009
8	0.7009
9	0.7009
10	0.7009
11	0.7009
12	0.6915



### 三、不同K值对比

如果选择较小的K值，就相当于用较小的领域中的训练实例进行预测，“学习”近似误差会减小，只有与输入实例较近或相似的训练实例才会对预测结果起作用，与此同时带来的问题是“学习”的估计误差会增大，换句话说，K值的减小就意味着整体模型变得复杂，容易发生过拟合。

如果选择较大的K值，就相当于用较大领域中的训练实例进行预测，其优点是可以减少学习的估计误差，但缺点是学习的近似误差会增大。这时候，与输入实例较远（不相似的）训练实例也会对预测器作用，使预测发生错误，且K值的增大就意味着整体的模型变得简单。



THANK YOU