

類神經網路簡介 HW4

Q36111281 張偉治

1. 原始資料處理

1.1 分類資料轉換

將七類的豆子名稱分別對應到數字 1 到 7

1.2 移除離群值

使用四分位數的方式移除離群值，其中參數 $n=3$ ，透過該方法總共移除 12% 的資料。

```
#剔除離群值
n=3
IQR = 0
count = []

#IQR = Q3-Q1
for i in range(len(train_full.iloc[0, :-1])):
    if(1):
        IQR = np.percentile(train_full.iloc[:, i], 75) - np.percentile(train_full.iloc[:, i], 25)

        count = []
        for j in range(len(train_full.iloc[:, i])):
            if( train_full.iloc[j, i] > np.percentile(train_full.iloc[:, i], 75)+n*IQR ):
                count.append(j)

            for j in range(len(train_full.iloc[:, i])):
                if( train_full.iloc[j, i] < np.percentile(train_full.iloc[:, i], 75)-n*IQR ):
                    count.append(j)

        for k in count:
            train_full = train_full.drop(labels=k, axis=0)
        train_full = train_full.reset_index(drop=True)
```

1.3 填補空值

```
train_full = train.fillna(0)  
train_full.isna().sum()
```

```
Area          0  
Perimeter     0  
MajorAxisLength  0  
MinorAxisLength  0  
AspectRatio    0  
Eccentricity   0  
ConvexArea     0  
EquivDiameter  0  
Extent         0  
Solidity       0  
roundness      0  
Compactness    0  
ShapeFactor1   0  
ShapeFactor2   0  
ShapeFactor3   0  
ShapeFactor4   0  
Class          0  
dtype: int64
```

2. SVM 調整參數

```
model_SVM1 = svm.SVC(kernel='linear', C=2, gamma='auto')  
model_SVM2 = svm.SVC(kernel='linear', C=5, gamma='auto')  
model_SVM3 = svm.SVC(kernel='rbf', gamma=0.1, C=1)
```

分別使用兩種 SVM 模型：linear 和 rbf，其中 linear 的部分又使用兩種不同的 C 值，區分不同模型的複雜度。

最後對應的準確率分別為 0.903、0.902、0.300 可以發現 linear 模型的 SVM 對該資料集具有較好的精確度。

3. 其他分類器

這次除使用 NN 外，另外還使用 bagging 的技巧，最後分別使用較寬與較深的 NN 模型對 bagging 後的資料進行預測，發現較寬的 NN2 對 bagging 具有相對較好的準確度。