

网格搜索 (Grid Search)

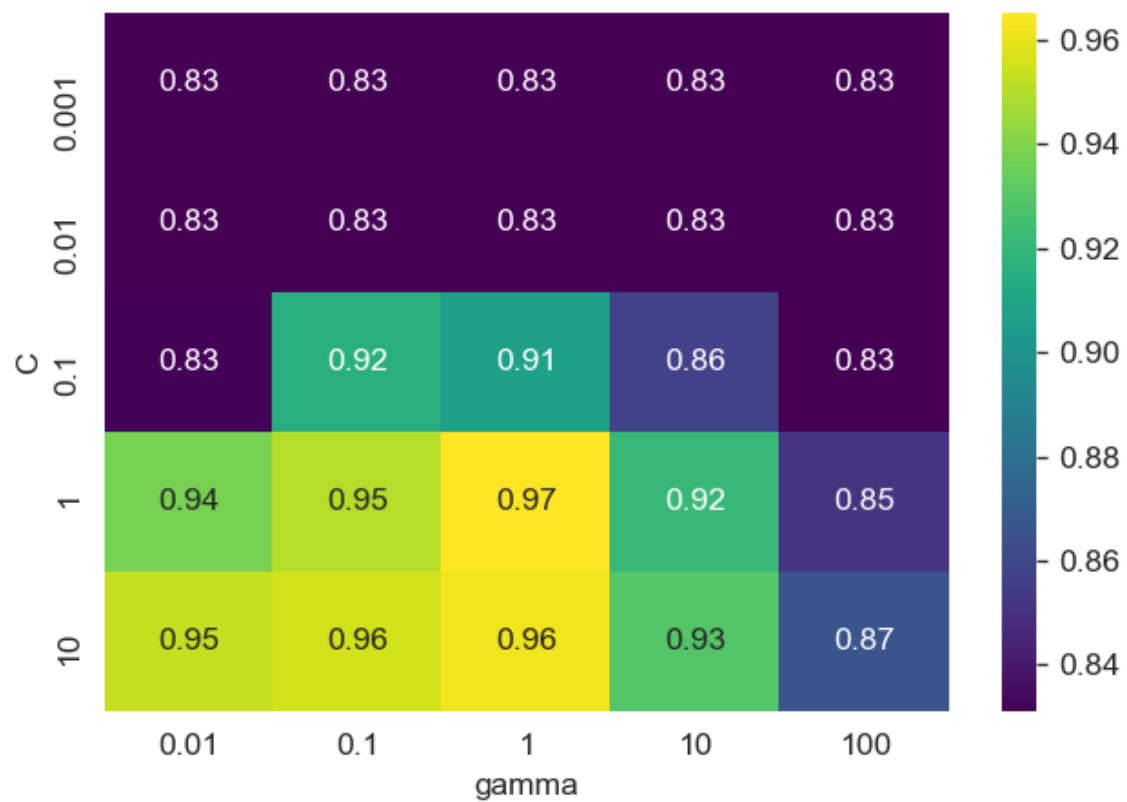
网格搜索

- SVM模型涉及两个参数：C和gamma
- 网格搜索：就是用循环对多个参数组合进行尝试
- 代码示例：

```
for gamma in [0.001,0.01,0.1,1,10]:  
    for C in [0.01,0.1,1,10,100]:  
        svm=SVC (gamma=gamma,C=C)  
        svm.fit(X_train,y_train)  
        score=svm.score(X_test,y_test)  
  
        .....  
  
svm=SVC (**best_parameters)  
svm.fit(X_train,y_train)
```

网格搜索

- 以gamma作为x，以C作为y，绘制网格搜索精度的热图



考虑过拟合的网格搜索

- 前面的代码中，我们用测试集进行了调参，又用测试集来评估模型，这种方式会有问题。



- 应该用一个独立的数据集来评估模型，就是一个在建立模型时没有用到的数据集，所以需要
将数据集做如下划分。



Training Data: 训练模型, Valid Data: 评估参数, Test Data: 评估模型

带有交叉验证的网格搜索

1、



2、



网格搜索：GridSearchCV类

- 由于带交叉验证的网格搜索是一种常用的调参方法，因此sklearn提供了一个GridSearchCV类，这个类可以帮助我们完成带有交叉验证的网格搜索。
- 用法如下：

```
param_grid={'C':[0.001,0.01,0.1,1,10],  
            'gamma':[0.01,0.1,1,10,100]}  
  
kfolder=KFold(n_splits=5,shuffle=True,random_state=0)  
grid_search=GridSearchCV(SVC(),param_grid,cv=kfolder)  
grid_search.fit(X_train,y_train)
```

交叉验证的结果可视化

1. 通过`grid_search`的`cv_results_`接口获取到网格搜索的结果`results`
2. 从`results`的`mean_test_score`属性获取到每一次的平均泛化精度
3. 用`seaborn`绘制热力图，以平均泛化精度作为`data`，以`gamma`为`x`轴，以`C`为`y`轴

