# SVM

February 8, 2021

[1]:
```
import pandas as pd
import numpy as np
import warnings
from sklearn import svm
from sklearn.model_selection import RepeatedStratifiedKFold
from sklearn.model_selection import GridSearchCV

warnings.filterwarnings('ignore')
```

## 1 Feature Selection Accuracy

[27]:
```
data = pd.read_csv("../Dataset/fe_leaf.csv", delimiter=",")
X_, y_ = data.iloc[:, :-1], data.iloc[:, -1:]
num_features = len(X_.columns)
```

[3]:
```
cv_strat = RepeatedStratifiedKFold(n_splits=5, n_repeats=4, random_state=42)
param = {'C': [0.001, 0.1, 1, 100, 100000],'kernel':['linear', 'poly', 'rbf',␣
 ↪'sigmoid'], 'degree':[2, 3, 4, 5, 6, 7, 8], 'gamma':['scale', 'auto'],␣
 ↪'class_weight':['balanced', None], 'cache_size':[500]}
```

[40]:
```
gs_SVM = GridSearchCV(estimator = svm.SVC(), param_grid = param, cv=cv_strat,␣
 ↪scoring='accuracy')
gs_SVM.fit(X_, y_)
print(gs_SVM.best_params_)
print(gs_SVM.best_score_)
```

```
{'C': 100000, 'cache_size': 500, 'class_weight': 'balanced', 'degree': 2,
'gamma': 'scale', 'kernel': 'rbf'}
0.7330882352941176
```

[41]:
```
results_SVM = pd.DataFrame(gs_SVM.cv_results_['params'])
results_SVM['test_score'] = gs_SVM.cv_results_['mean_test_score']
results_SVM
```

[41]:

| | C | cache_size | class_weight | degree | gamma | kernel | test_score |
|---|---|---|---|---|---|---|---|
| 0 | 0.001 | 500 | balanced | 2 | scale | linear | 0.038971 |
| 1 | 0.001 | 500 | balanced | 2 | scale | poly | 0.040441 |
| 2 | 0.001 | 500 | balanced | 2 | scale | rbf | 0.033088 |
| 3 | 0.001 | 500 | balanced | 2 | scale | sigmoid | 0.033088 |
| 4 | 0.001 | 500 | balanced | 2 | auto | linear | 0.038971 |
| .. | … | … | … | … | … | … | … |
| 555 | 100000.000 | 500 | None | 8 | scale | sigmoid | 0.064706 |
| 556 | 100000.000 | 500 | None | 8 | auto | linear | 0.711029 |
| 557 | 100000.000 | 500 | None | 8 | auto | poly | 0.697794 |
| 558 | 100000.000 | 500 | None | 8 | auto | rbf | 0.722794 |
| 559 | 100000.000 | 500 | None | 8 | auto | sigmoid | 0.084559 |

[560 rows x 7 columns]

## 2 Full Dataset Accuracy

[6]:
```
data_all = pd.read_csv("../Dataset/leaf.csv", delimiter=",")
X_all, y_all = data_all.iloc[:, :-1], data_all.iloc[:, -1:]
num_features_all = len(X_all.columns)
```

[4]:
```
param = {'C': [0.001, 0.1, 1, 100, 100000],'kernel':['linear', 'poly', 'rbf',␣
 ↪'sigmoid'], 'degree':[2, 3, 4, 5, 6, 7, 8], 'gamma':['scale', 'auto'],␣
 ↪'class_weight':['balanced', None], 'cache_size':[500]}
gs_SVM_all = GridSearchCV(estimator = svm.SVC(), param_grid = param,␣
 ↪cv=cv_strat, scoring='accuracy')
gs_SVM_all.fit(X_all, y_all)
print(gs_SVM_all.best_params_)
print(gs_SVM_all.best_score_)
```

```
{'C': 100000, 'cache_size': 500, 'class_weight': 'balanced', 'degree': 2,
'gamma': 'scale', 'kernel': 'poly'}
0.7588235294117646
```

[5]:
```
results_SVM_all= pd.DataFrame(gs_SVM_all.cv_results_['params'])
results_SVM_all['test_score'] = gs_SVM_all.cv_results_['mean_test_score']
results_SVM_all
```

[5]:

| | C | cache_size | class_weight | degree | gamma | kernel | test_score |
|---|---|---|---|---|---|---|---|
| 0 | 0.001 | 500 | balanced | 2 | scale | linear | 0.039706 |
| 1 | 0.001 | 500 | balanced | 2 | scale | poly | 0.038971 |
| 2 | 0.001 | 500 | balanced | 2 | scale | rbf | 0.032353 |
| 3 | 0.001 | 500 | balanced | 2 | scale | sigmoid | 0.034559 |
| 4 | 0.001 | 500 | balanced | 2 | auto | linear | 0.039706 |
| .. | … | … | … | … | … | … | … |

```
555  100000.000         500         None       8  scale  sigmoid    0.106618
556  100000.000         500         None       8   auto   linear    0.757353
557  100000.000         500         None       8   auto     poly    0.712500
558  100000.000         500         None       8   auto      rbf    0.744118
559  100000.000         500         None       8   auto  sigmoid    0.105147

[560 rows x 7 columns]
```

[ ]: