T6 - Nested CV

March 6, 2021

1 Nested cross validation

1.1 Setting up

[ ]:
```
import pandas as pd
import numpy as np
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC

# Load data
dataObj = load_breast_cancer()
X = dataObj.data
y = dataObj.target

# Splitting data
X_train, X_test, y_train, y_test = train_test_split(X, y,
    stratify=y,
    test_size=0.20,
    random_state=1)

# Model
pipe_svc = Pipeline([('scl', StandardScaler()),
            ('clf', SVC(random_state=1))])
```

[ ]:
```
param_range = [0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0]

set1 = {'clf__C': param_range,
        'clf__kernel': ['linear']}

set2 = {'clf__C': param_range,
        'clf__gamma': param_range,
        'clf__kernel': ['rbf']}

param_grid = [set1, set2]
```

1.2 5x2 nested stratified k-fold cross-validation

[ ]:
```
from sklearn.model_selection import GridSearchCV

# Construct grid search estimator
gs = GridSearchCV(estimator=pipe_svc,
                  param_grid=param_grid,
                  scoring='accuracy',
                  cv=2)
```

[ ]:
```
from sklearn.model_selection import cross_val_score

# Putting grid search into straintify k-fold cross validator
scores = cross_val_score(gs, X_train, y_train, scoring='accuracy', cv=5)

print("\nCrossed-validation accuracy")
print(f"Mean:{np.mean(scores):6.3f}")
print(f"Std:{np.std(scores):6.3f}")
```