

機械学習特論第 15 回レポート

1 sklearn と Stratified K-Folds cross-validation と Optuna と

sklearn と 層状 K 分割交差検証と（できれば Optuna も）を自分の研究に適用せよとのことでしたが, AttnGAN をはじめとした GAN にはこれらは適用できないため, 3 回生の時の情報工学実験での DCGAN で取り扱った MNIST に関連して「手書き文字 (digits) の分類問題」に適用しました.

カーネル関数は RBF とシグモイドの二択, C とガンマは 0.0001~1000 の間で探索, 1000 回試行にてパラメータチューニングをさせました.

得られた最適パラメータは次のようになりました.

カーネル: RBF

C: 0.9248595577937428

ガンマ: 0.0012640732529140759

```
from functools import partial

import optuna
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import cross_validate
from sklearn.svm import SVC
from sklearn import datasets

def objective(X, y, trial):
    params = {
        'kernel': trial.suggest_categorical('kernel', ['rbf', 'sigmoid']),
        'C': trial.suggest_loguniform('C', 1e-3, 1e+3),
        'gamma': trial.suggest_loguniform('gamma', 1e-3, 1e+3)
    }
    model = SVC(**params)

    kf = StratifiedKFold(n_splits=5, shuffle=True, random_state=0)
    scores = cross_validate(model, X=X, y=y, cv=kf)
    return 1.0 - scores['test_score'].mean()

dataset = datasets.load_digits()
X, y = dataset.data, dataset.target
f = partial(objective, X, y)
study = optuna.create_study()
study.optimize(f, n_trials=1000)
print('params:', study.best_params)
```