

進捗報告

水野泰旭

November 14, 2022

弘前大学工学部電子情報工学科 4 年

k 分割交差検証

混同行列のヒートマップ

まとめ

k 分割交差検証

Listing 1: kseparate_train.py

```
k = 10
n = len(images) // k
label_size = len(set(labels))
validation_scores = []
for fold in range(k):
    print("The fold is :", fold)
    # SEPARATE TRAIN DATA
    validation_images = images[n * fold: n * (fold + 1)]
    validation_labels = labels[n * fold: n * (fold + 1)]
    train_images = np.concatenate([images[:n * fold], images[n * (fold + 1):]])
    train_labels = np.concatenate([labels[:n * fold], labels[n * (fold + 1):]])
```

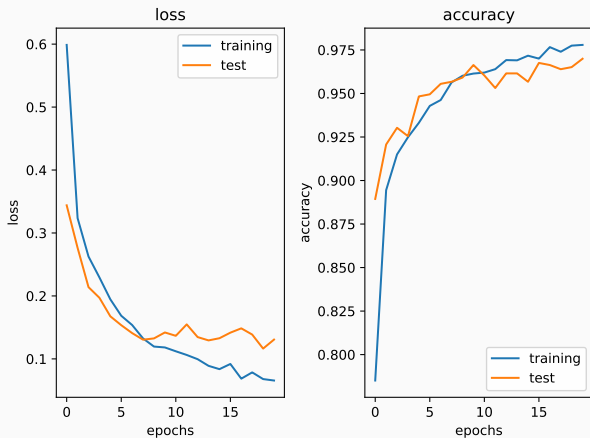
Table 1: k 分割交差検証結果（比率は無視）

fold	loss	accuracy
1	0.2004694938659668	0.9520807266235352
2	0.0869437754154205	0.9747793078422546
3	0.1595330983400344	0.9697352051734924
4	0.1649435311555862	0.9546027779579163
5	0.1361409425735473	0.9596469402313232
6	0.1505440622568130	0.9533417224884033
7	0.1733103841543197	0.9508196711540222
8	0.2493898719549179	0.9407314062118530
9	0.1259648948907852	0.9672130942344666
10	0.1510167270898819	0.9697352051734924
average	0.1598256722092628	0.9592686057090759

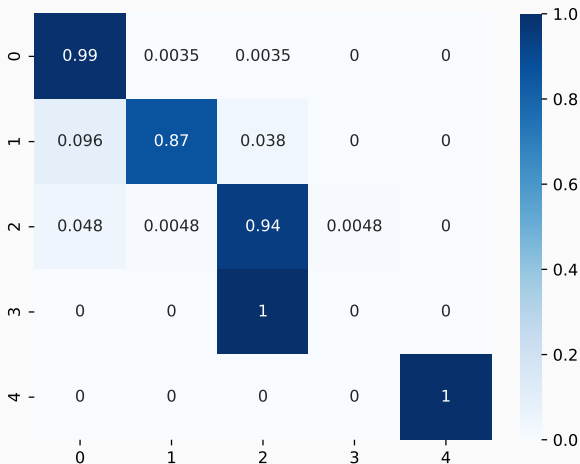
- ・ 配列をシャッフルして k 個に分割したため、訓練データと検証データのラベルの比率が毎回バラバラになっているため、結果が正確とは言えない
→sklearn を用いて、訓練データと検証データのラベルの比率を揃える

混同行列のヒートマップ

Test Accuracy: 0.9699519276618958

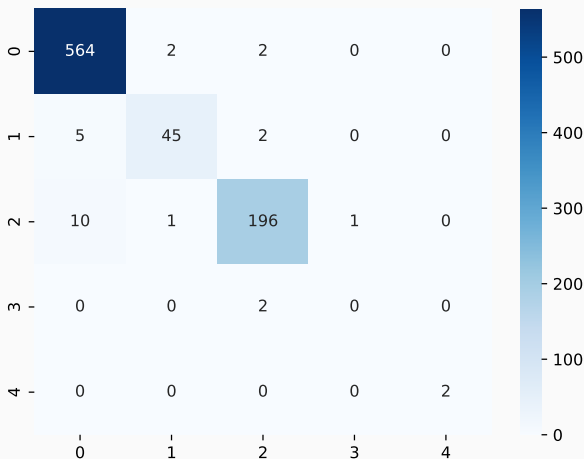


混同行列（割合）のヒートマップ¹



¹縦軸が正解ラベル、横軸が予測ラベル

混同行列（個数）のヒートマップ²



²縦軸が正解ラベル、横軸が予測ラベル

- ・ メモリが数値でわかりにくいので、ファミリー名で表す
- ・ x 軸と y 軸のラベルを入れる
- ・ 正解ラベルが 1 または 2 で、予測ラベルが 0 となって間違っているデータが多い
→ データの偏りが原因として考えられるので、水増しをしたら正解率上がるかも

まとめ

- ・ k 分割交差検証を sklearn を用いて訓練データと検証データのラベルの比率を 9:1 に揃える
- ・ 混同行列のヒートマップのラベルをファミリー名に直す
- ・ 卒論を書く