

任务一：数据分布存在差异

任务和主题

训练集使用 mnist，测试集使用 color mnist。在数据分布存在差异的情况下，SVM 和 CNN 的精度会降低，如何解决？

数据分布

训练集使用删减过后的 `mnist` 手写数字数据集，标注文件使用 `less_train_labs.txt`。训练集中包含 0~9 十个类别各 200 张图片。测试集使用 `color mnist` 的测试集，标注文件使用 `MNIST_color\testset\testset_label.txt`。

数据下载

本任务使用的标注文件在原始 `mnist` 数据集上进行了修改，详见数据集中的 `README.txt`。数据集下载链接：

<https://pan.baidu.com/s/1nDnZ7ofig7YZnkeB7ug2Rw> 提取码：zhe0

结果提交

要求给出 SVM 和 CNN 两种算法在测试集上每个类别的精度和平均精度，并将自己模型结果与下述两个 `baseline` 进行对比。

Baseline-SVM:

```
Class Accuracy:
Class 0: 0.0092
Class 1: 0.0000
Class 2: 0.9157
Class 3: 0.0020
Class 4: 0.0000
Class 5: 0.0504
Class 6: 0.0000
Class 7: 0.0000
Class 8: 0.0000
Class 9: 0.0000
Average Accuracy: 0.0977
```

Baseline-CNN:

```
Class Accuracy:
Class 0: 0.6653
Class 1: 0.1119
Class 2: 0.5475
Class 3: 0.0089
Class 4: 0.0132
Class 5: 0.0000
Class 6: 0.0031
Class 7: 0.1508
Class 8: 0.1253
Class 9: 0.0030
Average Accuracy: 0.1629
```


任务二：数据不平衡

任务和主题

训练集和测试集均使用 `mnist` 数据集，训练集各类别数据量不同，存在样本数量较少的类别。在这种数据不平衡的情况下，`SVM` 和 `CNN` 的精度会降低，如何解决？

数据分布

训练集使用不平衡的 `mnist` 数据集，标注文件使用 `imbalance_less_train_labels.txt`。训练集在任务一数据集的基础上，对“1”“3”“4”“6”“8”五类进行数量调整，分别保留 40，20，20，10，10 个样本；其他类别仍为 200 个样本。

测试集使用 `mnist` 原始的测试集。

数据下载

同任务一

结果提交

要求给出 `SVM` 和 `CNN` 两种算法在测试集上每个类别的精度和平均精度，并将自己模型结果与下述两个 `baseline` 进行对比。

Baseline-SVM

```
Class Accuracy:
Class 0: 0.9806
Class 1: 0.9427
Class 2: 0.9419
Class 3: 0.5535
Class 4: 0.4532
Class 5: 0.9428
Class 6: 0.5271
Class 7: 0.8794
Class 8: 0.1756
Class 9: 0.9167

Average Accuracy: 0.7313
```

Baseline-CNN

```
Class Accuracy:
Class 0: 0.9867
Class 1: 0.9330
Class 2: 0.9583
Class 3: 0.7257
Class 4: 0.5540
Class 5: 0.9910
Class 6: 0.4541
Class 7: 0.9183
Class 8: 0.2279
Class 9: 0.9346

Average Accuracy: 0.7684
```

任务三：数据存在标注错误

任务和主题

训练集和测试集均使用 `mnist` 数据集，但训练集中存在样本标注错误。在这种存在标注错误的情况下，`SVM` 和 `CNN` 的精度会降低，如何解决？

数据分布

训练集使用存在标注错误的 `mnist` 数据集，标注文件使用 `50%_incorrect_train_labs.txt`。训练集在任务一数据集的基础上，选择 50% 的数据，将其标签随机化为 0~9 中任一数字。测试集使用 `mnist` 原始的测试集。

数据下载

同任务一

结果提交

要求给出 `SVM` 和 `CNN` 两种算法在测试集上每个类别的精度和平均精度，并将自己模型结果与下述两个 `baseline` 进行对比。

Baseline-SVM

```
精确度：
Class 0: 0.9262948207171314
Class 1: 0.7873563218390804
Class 2: 0.9236276849642004
Class 3: 0.8459979736575481
Class 4: 0.8214643931795386
Class 5: 0.8023630504833512
Class 6: 0.8793814432989691
Class 7: 0.8141153081510935
Class 8: 0.8020361990950227
Class 9: 0.8102926337033299
Accuracy: 0.8385
```

Baseline-CNN

```
Class Accuracy:
Class 0: 0.9531
Class 1: 0.9877
Class 2: 0.9225
Class 3: 0.8911
Class 4: 0.9216
Class 5: 0.9406
Class 6: 0.9729
Class 7: 0.9037
Class 8: 0.7505
Class 9: 0.8553

Average Accuracy: 0.9099
```