

Report of HW4

王焕宇 522030910212

Accuracy

使用 SVC(kernel='linear') 时，如果不设置 max_iter，模型会迟迟不收敛（可能是因为样本之间线性关系不明显），导致程序卡死。在设置了 max_iter 后，模型会警告 ConvergenceWarning: Solver terminated early，并且准确率明显较低，且准确率与 max_iter 大小较为相关，很不稳定。查阅资料后，我本想换用 LinearSVC(C=0.1, max_iter=100)，但是后者无法得到准确的支持向量，只能近似得到结果。因此，在测试集上的分类准确率中，我将汇报四种 SVM 的准确率，但是在后续讨论支持向量时，选取 SVC(kernel='linear', C=1, max_iter=5000) 作为分类器。

Method	Accuracy
SVC: Linear	96.20%
LinearSVC	98.25%
SVC: RBF	99.20%
SVC: Polynomial	99.10%

上述结果对应参数如下：

```
SVC(kernel='linear', C=1, max_iter=5000)
LinearSVC(C=0.1, max_iter=100)
SVC(kernel='rbf', C=2, gamma='scale')
SVC(kernel='poly', degree=2, C=1, gamma='scale', coef0=1)
```

P.S. 当设置 LinearSVC 中 max_iter = 5 时，会导致 ConvergenceWarning，但同时也会让准确率有 0.45% 的提升，但是没有采用这套参数。

Linear SVM

- 一共有 837 个支持向量参与 w 的计算，其中类别 0 包含 481 个，类别 1 包含 356 个。

```
len(classifier.support_)      # 837
classifier.n_support_         # [481, 356]
```

- 以下是上述支持向量的可视化，以及对应的权重（一页 420 张，总共 837 张，原图见代码压缩包）：



