

---

# 天津大学

## 模式识别与深度学习课程

### 实验 1、SVM 分类算法实验报告



学 院 智能与计算学部  
专 业 计算机科学与技术  
学 号 3019244266  
姓 名 李润泽

---

## 1. 实验目标

本实验要求掌握 SVM 的基本原理，熟悉 SVM 并使用其解决鸢尾花和肿瘤的分类问题。首先在理解和掌握 SVM 鸢尾花分类算法原理的基础上，调节不同的参数进行实验并对不同参数所产生的影响进行分析，从而判断出哪些参数对实验结果影响较大并阐述原因；其次根据已有算法对肿瘤数据进行肿瘤分类的实验，调节参数分析训练集和测试集的准确率，对不用参数所产生的影响进行分析，判断出哪些参数对实验结果的影响比较大并阐述原因。

## 2. 实验一

### 2.1 算法实现及参数调节说明

#### 2.1.1 SVM 模型搭建算法

```
def classifier():
    clf = svm.SVC(C=0.5,                # 误差惩罚系数，默认 1
                  kernel='linear',      # 线性核 kernel='rbf':高斯核
                  decision_function_shape='ovo')  # 决策函数
    return clf
```

#### 2.1.2 SVM 模型训练算法

```
def train(clf, x_train, y_train):
    clf.fit(x_train,                    # 训练集特征向量
            y_train.ravel())           # 训练集目标值
```

#### 2.1.3 SVM 模型测试算法

```
def print_accuracy(clf, x_train, y_train, x_test, y_test):
    # 分别打印训练集和测试集的准确率
    # score(x_train, y_train):表示输出 x_train, y_train 在模型上的准确率
    print('training prediction:%.3f' % (clf.score(x_train, y_train)))
    print('test data prediction:%.3f' % (clf.score(x_test, y_test)))
```

#### 2.1.4 实验结果

本实验对 SVM 模型中的误差惩罚系数、kernel 和决策函数进行了不同程度的修改，实验结果如下：

C	kernel	decision_function_shape	Train acc	Test acc
0.1	linear	ovr	0.962	0.978
0.2			0.971	0.978
0.3			0.981	0.978
0.4			0.981	0.978
0.5			0.981	0.978
0.6			0.981	0.978
0.7			0.981	1
0.8			0.981	1
0.9			0.981	1
1			0.981	1
0.5	poly	ovo	0.981	1
0.5	rbf		0.962	0.978
0.5	sigmoid		0.352	0.289
0.5	linear	ovo	0.981	0.978

表 1 不同参数的 SVM 模型的训练准确率和测试准确率

根据表 1 的数据可以得出如下的图表：

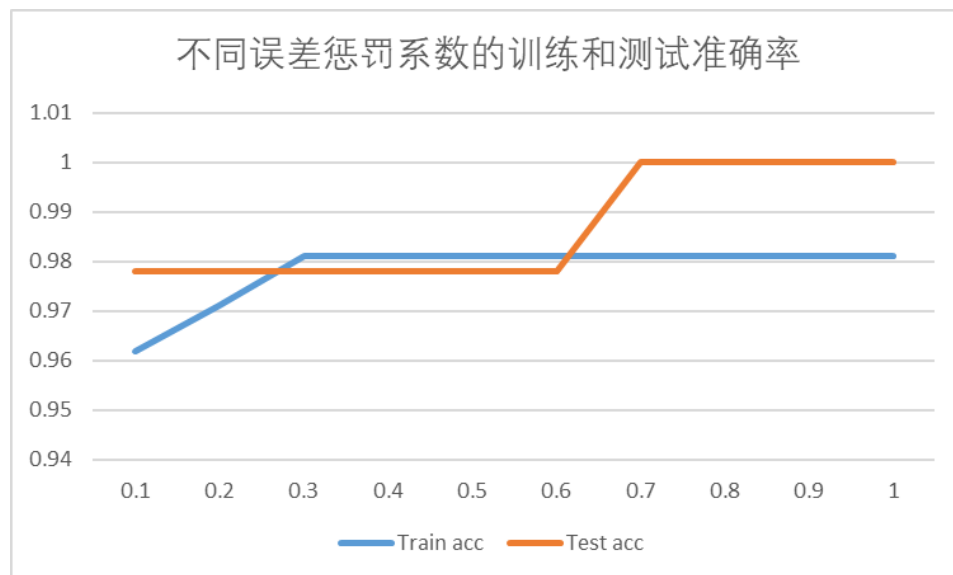


图 1 不同误差惩罚系数 C 的训练准确率和测试准确率（kernel=linear, decision\_function\_shape=ovr）

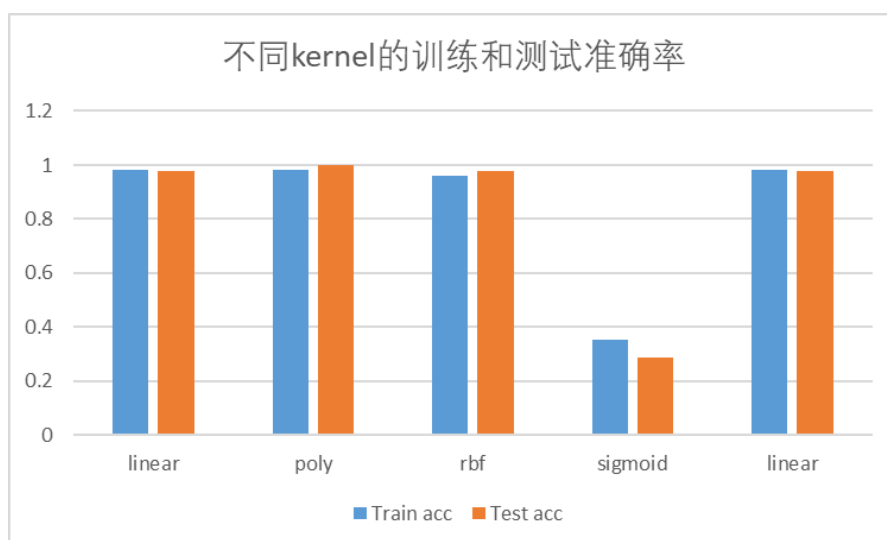


图 2 不同 kernel 的训练准确率和测试准确率 (C=0.5, decision\_function\_shape=ovr)

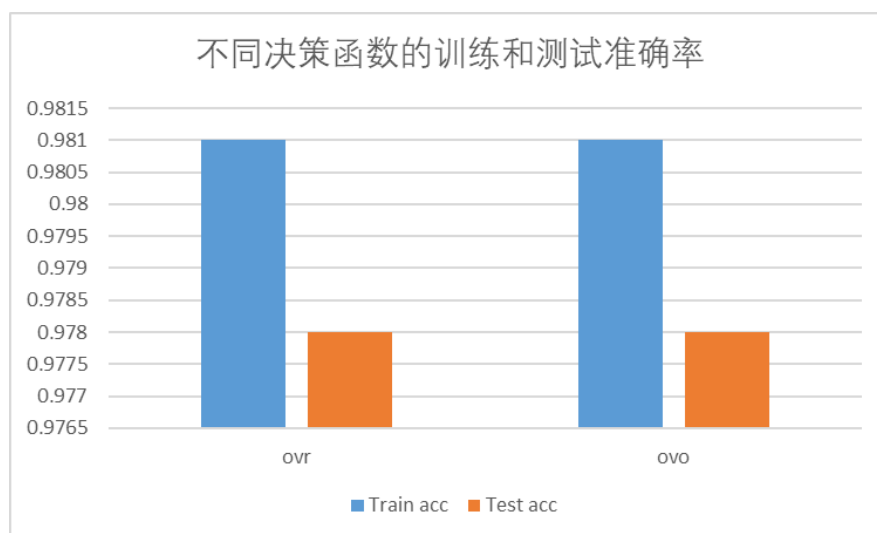


图 3 不同决策函数 decision\_function\_shape 的训练准确率和测试准确率 (C=0.5, kernel=linear)

## 2.2 结果分析

根据上述图表，可以看出以下结论：

- 1、误差惩罚系数对训练和测试准确率的影响比较大，其中， $C < 0.3$  时，训练准确率随着  $C$  的上升而上升，此后保持不变；当  $C = 0.7$  时，测试准确率会上升，此后保持不变；
- 2、kernel 对训练和测试准确率的影响不大，只有 kernel=sigmoid 的模型的准确率比其他 kernel 的模型低，而其他 kernel 的 SVM 模型的准确率相近；
- 3、决策函数对训练和测试准确率影响不大，两种决策函数 (ovr 与 ovo) 的准确率相同。

## 3. 实验二

### 3.1 算法实现说明

#### 3.1.1 SVM 模型搭建算法

```
def classifier():
    clf = svm.SVC(C=0.5,                # 误差惩罚系数，默认 1
                  kernel='linear',      # 线性核 kernel='rbf':高斯核
                  max_iter=10000,       # 最大迭代次数，默认-1
                  decision_function_shape='ovr') # 决策函数
    return clf
```

#### 3.1.2 SVM 模型训练算法

```
def train(clf, x_train, y_train):
    clf.fit(x_train,          # 训练集特征向量
            y_train.ravel()) # 训练集目标值
```

#### 3.1.3 SVM 模型测试算法

```
def print_accuracy(clf, x_train, y_train, x_test, y_test):
    # 分别打印训练集和测试集的准确率
    # score(x_train, y_train):表示输出 x_train, y_train 在模型上的准确率
    print('training prediction:%.3f' % (clf.score(x_train, y_train)))
    print('test data prediction:%.3f' % (clf.score(x_test, y_test)))
```

#### 3.1.4 实验结果

由于 max\_iter 默认为-1，而该情况下的程序运行时间过长，在建立该模型时添加了 max\_iter 参数。本实验对 SVM 模型中的误差惩罚系数、kernel、决策函数和最大迭代次数进行了不同程度的修改，实验结果如下：

C	kernel	decision_function_shape	max_iter	Train acc	Test acc
0.5	linear	ovr	10000	0.296	0.407
			1000	0.296	0.389
			100	0.344	0.352
			100000	0.288	0.407
			1000000	0.288	0.407
0.1			10000	0.296	0.407
0.2				0.296	0.407
0.3				0.296	0.407
0.4				0.296	0.407
0.6				0.296	0.407
0.7				0.296	0.407

0.8				0.296	0.407
0.9				0.296	0.407
1				0.296	0.407
0.5	poly			0.392	0.278
	rbf			0.48	0.333
	sigmoid			0.424	0.278
linear	ovo			0.296	0.407

表 2 不同参数的 SVM 模型的训练准确率和测试准确率

根据表 2 的数据可以得出如下的图表：

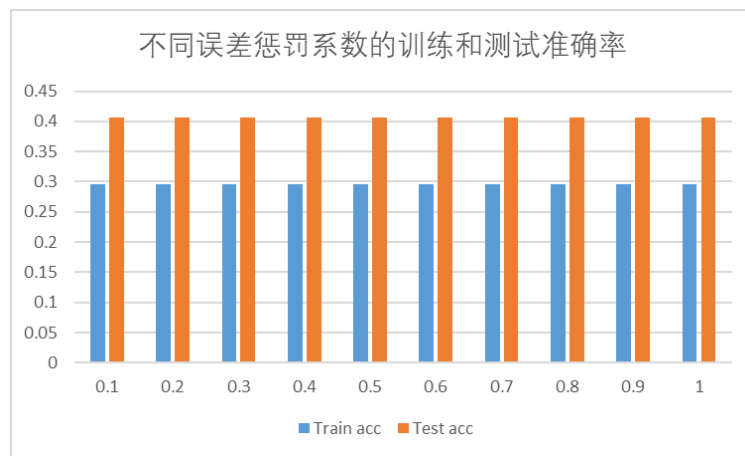


图 4 不同误差惩罚系数 C 的训练准确率和测试准确率（kernel=linear, decision\_function\_shape=ovr, max\_iter=10000）

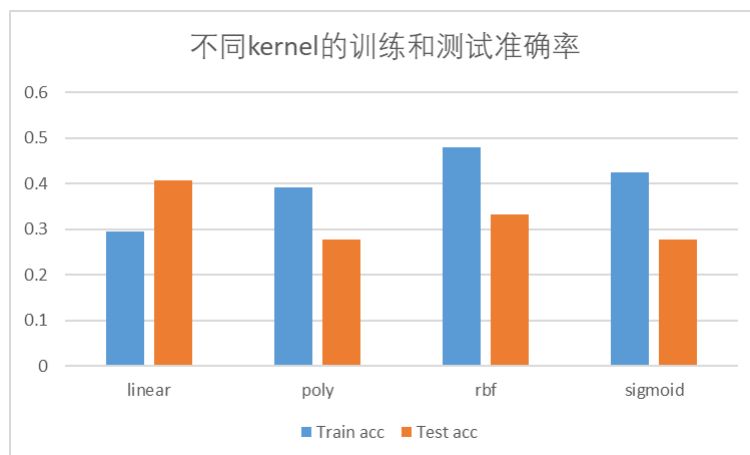


图 5 不同 kernel 的训练准确率和测试准确率（C=0.5, decision\_function\_shape=ovr, max\_iter=10000）

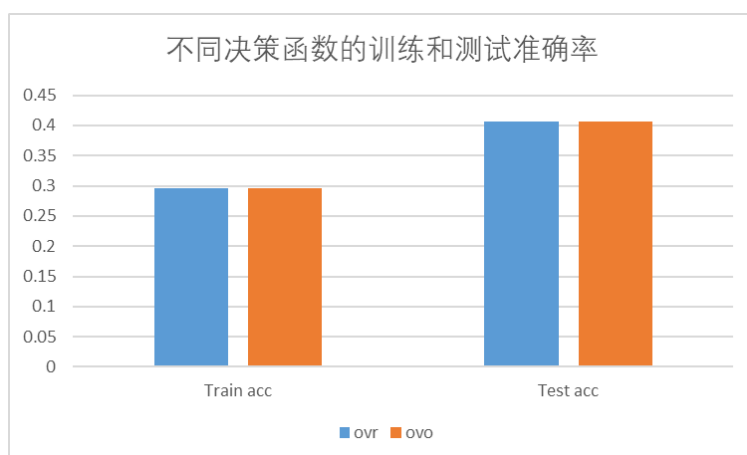


图 6 不同决策函数 `decision_function_shape` 的训练准确率和测试准确率（ $C=0.5$ , `kernel=linear`, `max_iter=10000`）

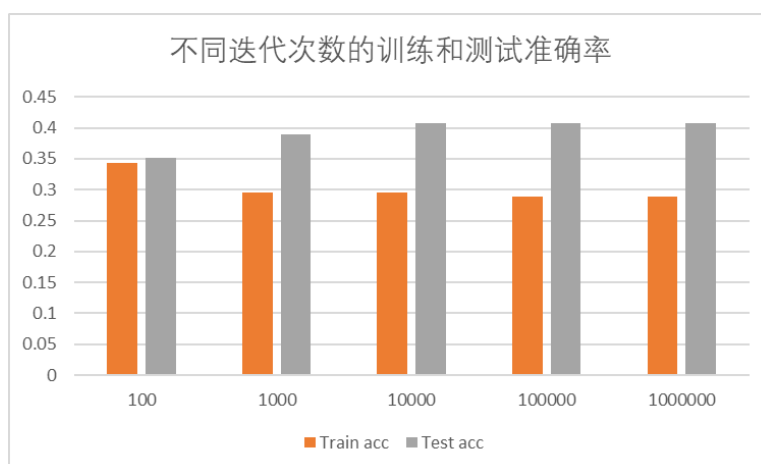


图 7 不同迭代次数 `max_iter` 的训练准确率和测试准确率（ $C=0.5$ , `kernel=linear`, `decision_function_shape=ovr`）

## 3.2 结果分析

根据上述图表，可以看出以下结论：

- 1、误差惩罚系数对训练和测试准确率的影响不大，无论  $C$  取多少，都不会影响模型的准确率；
- 2、`kernel` 对训练和测试准确率的影响较大，其中 `kernel=linear` 的模型测试准确率最高，而 `kernel=rbf` 的模型训练准确率最高；
- 3、决策函数对训练和测试准确率影响不大，两种决策函数（`ovr` 与 `ovo`）的准确率相同；
- 4、迭代次数对训练和测试准确率影响较大，随着迭代次数的上升，训练准确

---

率不断下降，而测试准确率在不断上升。

## 4. 总结

通过这次实验，我对 SVM 模型有了更多的理解。在我对鸢尾花分类算法的代码进行理解分析之后，成功以此进行了肿瘤分类代码的编写并顺利运行，从而提升了我的代码能力。针对老师在实验指导书上的各种问题我也进行了分析，最终顺利完成了本次实验，对于我在学习模式识别和深度学习的过程中会有更大的帮助。