## 数据挖掘Practice部分

#### 算法选择与实现

实现的数据挖掘算法:线性分类器,分别采用梯度下降法,最小平方误差准则与Fishe 线性判别准则SVM进行实现,并进行可视化。

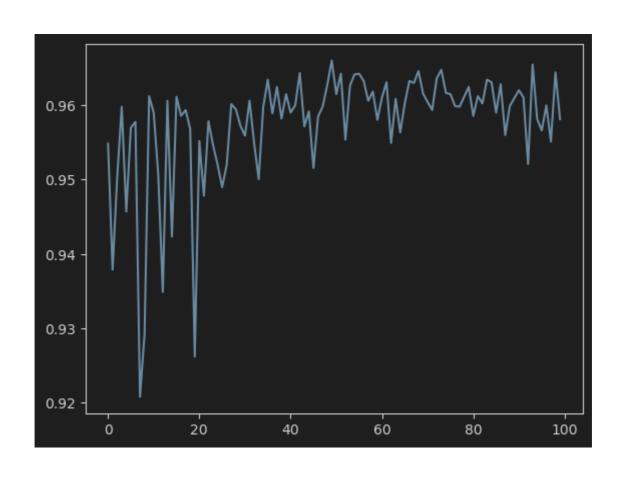
## 数据集与评估指标选择

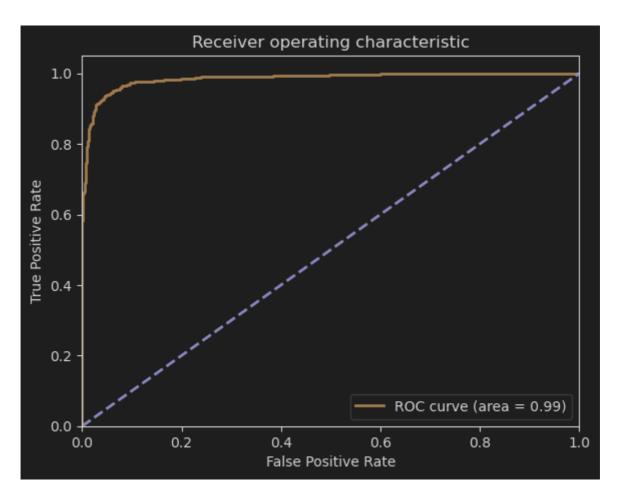
选择mnist手写数字数据集,二分类时只选择5,8进行二分类,二分类选择ROC曲线进行评估,多分类时选混淆矩阵进行评估。

### 测试结果

#### 梯度下降

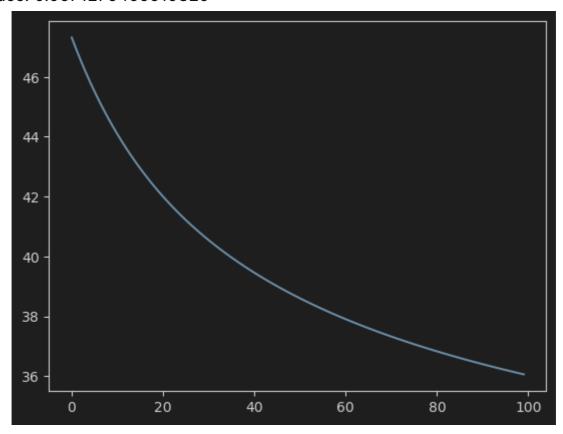
Train acc: 0.9581263307310149 Test acc: 0.9421221864951769

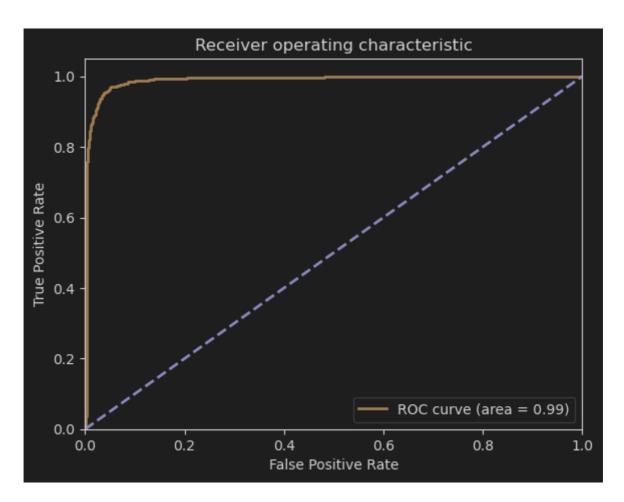




# 最小平方误差

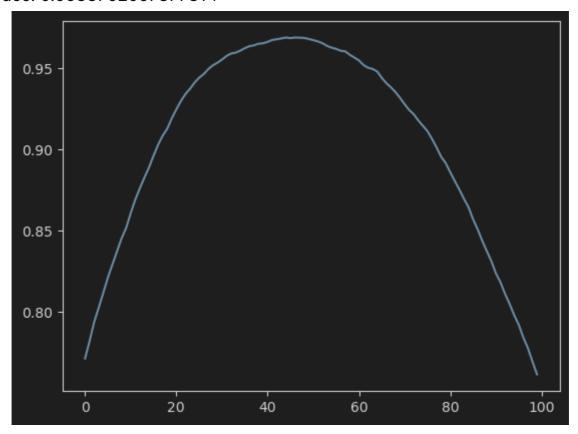
Train acc: 0.9744499645138396 Test acc: 0.9571275455519829

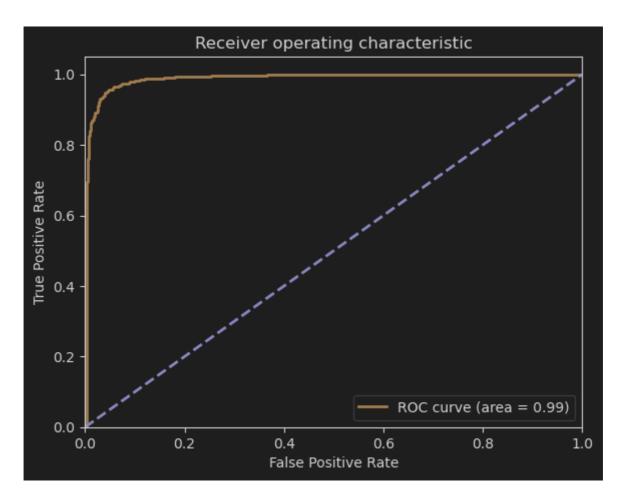




#### **Fisher**

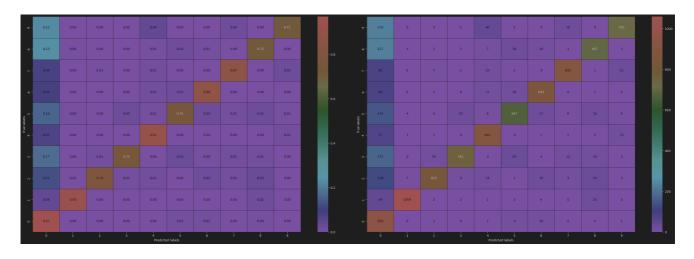
Train acc: 0.9687721788502484 Test acc: 0.9533762057877814





# **OVR** (based on fisher)

	precision	recall	f1-score	support	
0	0.44	0.97	0.61	980	
1	0.98	0.93	0.96	1135	
2	0.96	0.79	0.87	1032	
3	0.95	0.75	0.84	1010	
4	0.90	0.91	0.90	982	
5	0.90	0.75	0.82	892	
6	0.93	0.88	0.90	958	
7	0.94	0.87	0.90	1028	
8	0.89	0.72	0.79	974	
9	0.94	0.71	0.81	1009	
accuracy			0.83	10000	
macro avg	0.88	0.83	0.84	10000	
weighted avg	0.89	0.83	0.84	10000	



# **SVM**

Train starting. Train finished. Evaluate starting. Train acc: 0.9899166666666667 Test acc: 0.9792								
	precision	recall	f1-score	support				
0	0.98	0.99	0.99	980				
1	0.99	0.99	0.99	1135				
2	0.98	0.97	0.98	1032				
3	0.97	0.99	0.98	1010				
4	0.98	0.98	0.98	982				
5	0.99	0.98	0.98	892				
6	0.99	0.99	0.99	958				
7	0.98	0.97	0.97	1028				
8	0.97	0.98	0.97	974				
9	0.97	0.96	0.97	1009				
accuracy			0.98	10000				
macro avg	0.98	0.98	0.98	10000				
weighted avg	0.98	0.98	0.98	10000				

