

# Report

杨子董 521120910138

## 1.Linear SVM

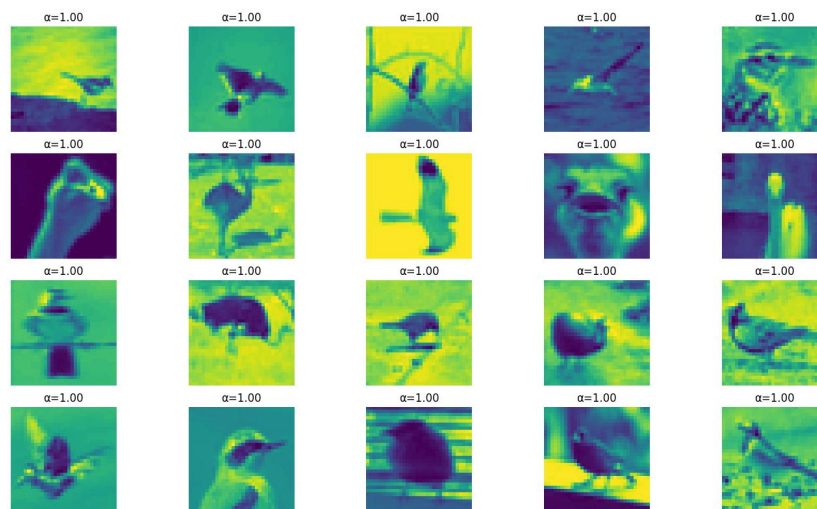
### 实现方式:

Linear SVM 我们在给定的读取和标准化数据后。只用调用 `sklearn.svm` 中的 `SVC` 类设定 `kernel='linear'`，然后在用 `fit` 拟合 `t` 就完成了。然后调用 `sklearn.metrics` 中的 `accuracy_score` 输出分类的正确率。使用 `.n_support_` 属性输出支持向量个数，用 `.dual_coef_[0]` 判断样本的符号，`np.sum` 输出个数。最后用 `np.argsort` 分别选出最大的 20 个正负样本下标，再调用 `matplotlib` 中的 `pyplot` 可视化对应图片即可。

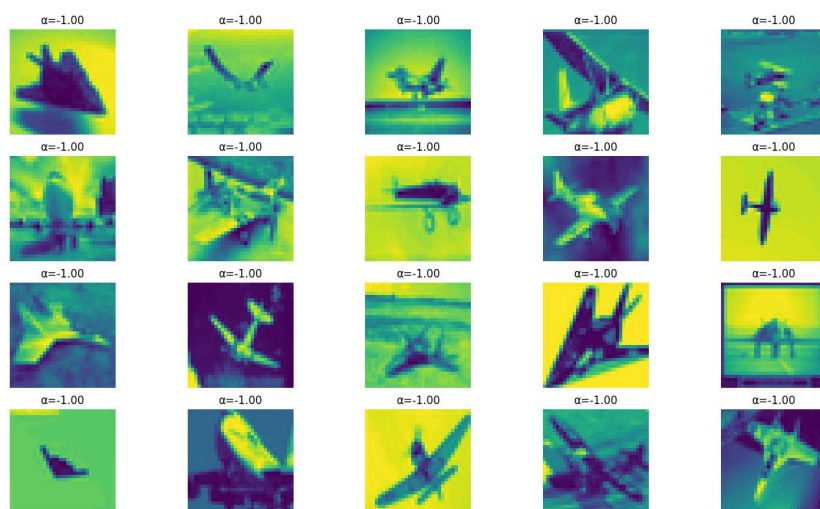
### 输出结果:

test\_accuracy: 84.60%  
support vectors: 3714  
positive support vectors: 1867  
negative support vectors: 1847

### TOP 20 Positive Samples:



### TOP 20 Negative Samples:



Linear SVM 的运行时间是最长的，分类效果是最差的。

初始的超参数正则化参数  $C=1$ 。

理论上  $C$  越大，对训练集分类精度会越细，但是分类边界会不平滑，容错低，运行时间也会越长。我测试了  $C=0.1$ ， $C=2$  和  $C=10$ ，输出结果分别为：

test\_accuracy: 84.30%  
 support vectors: 3767  
 positive support vectors: 1888  
 negative support vectors: 1879

test\_accuracy: 84.60%  
 support vectors: 3706  
 positive support vectors: 1865  
 negative support vectors: 1841

test\_accuracy: 84.60%  
 support vectors: 3705  
 positive support vectors: 1860  
 negative support vectors: 1845

$C=0.1$  时，运行时间短但是分类正确率些许下降， $C=2$  和  $10$  时分类正确率几乎没有变化，但是运行时间长了不少。因此取  $C=1$  是相对最合适的。

## 2.RBF kernel SVM

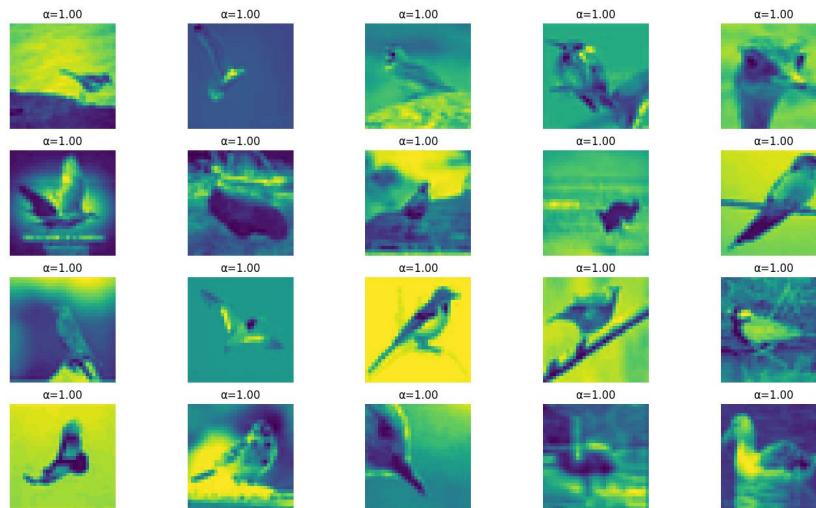
实现方式：

实现方式和 Linear SVM 的完全相同，使 `kernel='rbf'` 即可。

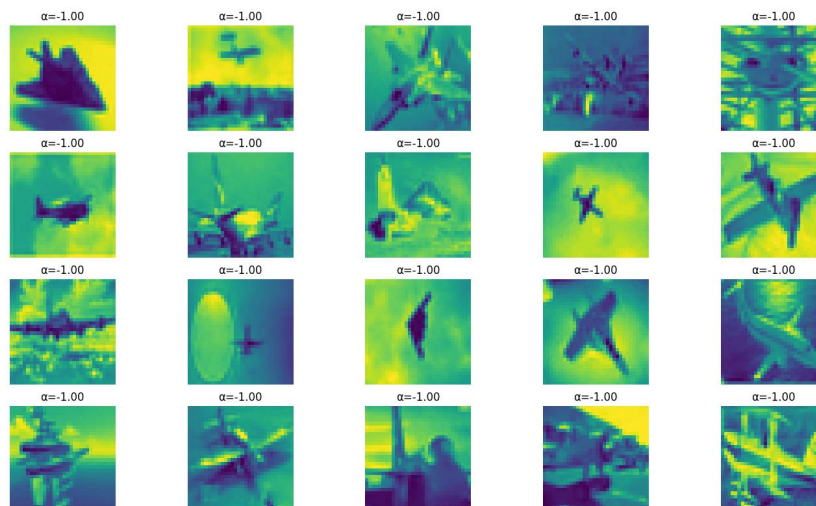
输出结果：

test\_accuracy: 87.95%  
support vectors: 4606  
positive support vectors: 2343  
negative support vectors: 2263

TOP 20 Positive Samples:



TOP 20 Negative Samples:



RBF kernel SVM 运行时间较好，分类效果最好。

初始的超参数正则化参数  $C=1$ ， $\gamma=\text{scale}=1/324$ 。

经过我的调优，分别测试了  $\gamma=1/340, 1/300, 1/280, 1/290, 1/285, 1/284, 1/286$  等； $C=0.5, 1, 2, 1.5, 2.5, 3$  等后。

在正则化参数  $C=2.5$ ,  $\gamma=1/285$  时, 输出结果为:

test\_accuracy: 88.55%  
support vectors: 5232  
positive support vectors: 2673  
negative support vectors: 2559

### 3.polynomial kernel SVM

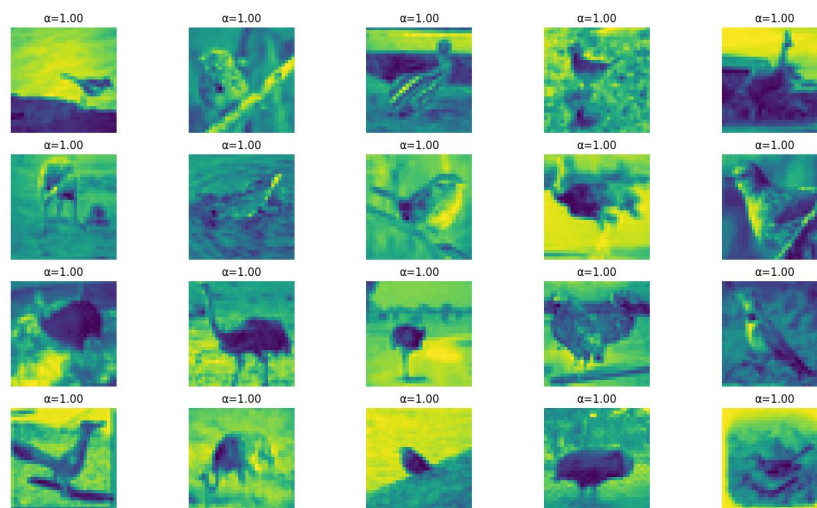
实现方式:

实现方式和 Linear SVM 的完全相同, 将 SVC 超参数设定为 `kernel='poly'` 即可。

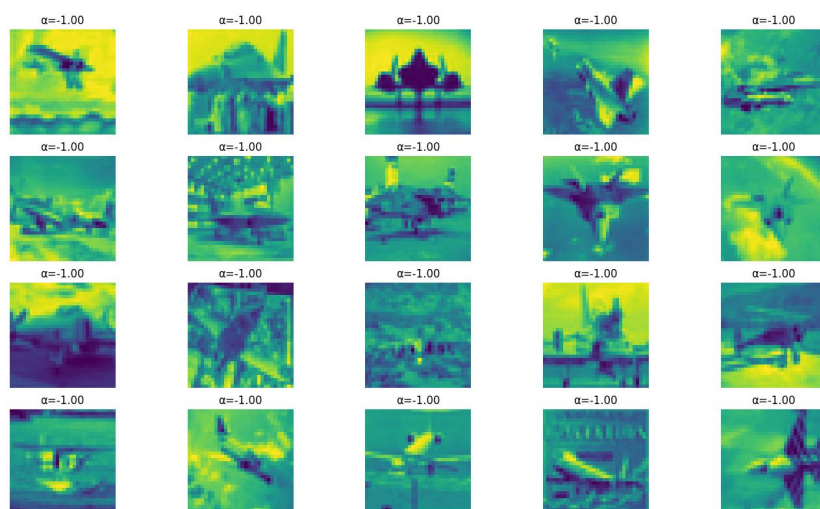
输出结果:

test\_accuracy: 86.80%  
support vectors: 6022  
positive support vectors: 3074  
negative support vectors: 2948

TOP 20 Positive Samples:



TOP 20 Negative Samples:



polynomial kernel SVM 运行时间最快，分类效果最好。

初始的超参数正则化参数  $C=1$ ， $\text{degree}=3$ ， $\text{gamma}=\text{scale}=1/324$ 。

经过我的调优，分别测试  $\text{degree}=10, 2, 3, 4$  等； $C=1, 2, 1.5, 0.5$  等； $\text{gamma}=1/330, 1/320, 1/300, 1/325, 1/329, 1/331, 1/332$  等后。

在正则化参数  $C=1.5$ ， $\text{degree}=3$ ， $\text{gamma}=1/330$  时，输出结果为：

test\_accuracy: 87.25%

support vectors: 5989

positive support vectors: 3063

negative support vectors: 2926