

SVM 算法实现实验报告

201250070 郁博文

一、数据集选取

为了实现二分类，我选取了 Ionosphere Dataset，它包含了 34 个特征和 1 个目标变量，用于雷达信号分类的二分类任务。我将数据集按照 8:2 划分训练集和测试集。

二、SVM 算法

SVM 算法的基本思想是通过在样本特征空间中构建一个最优的超平面，使得不同类别的样本能够在超平面上得到最大的间隔，并且分类错误的样本数最少。

这个最优的超平面即为分类边界，而位于分类边界上的样本点称为支持向量。

SVM 的目标是找到一个超平面，可以将两类样本完全分开，或者在无法完全分开时，使分类错误的样本数最小化。对于线性可分的情况，SVM 通过最大化间隔来选择最优的分类边界。间隔是分类边界到支持向量的距离，即支持向量机的目标是找到一个使得间隔最大化的超平面。

对于线性不可分的情况，SVM 引入了松弛变量 (slack variable) 来容忍一定的分类错误。通过调整松弛变量的值，可以控制分类错误的容忍度。SVM 的目标函数变为了最大化间隔的同时，尽量减小分类错误和松弛变量的总和。

SVM 通过求解一个凸优化问题来找到最优的分类边界。通过引入拉格朗日乘子，可以将 SVM 的优化问题转换为对拉格朗日乘子的求解。对拉格朗日乘子进行优化后，可以得到最优的分类边界和支持向量。

具体的实现代码，由于有 sklearn 库所以十分简单，如下图就是模型的构建和拟合代码：

```
# 创建SVM模型
svm_model = SVC(kernel='linear', random_state=42)

# 拟合模型
svm_model.fit(X_train, y_train)

# 对测试集进行预测
y_pred = svm_model.predict(X_test)
```

三、 实验结果

测试集结果如下图所示：

```
Confusion Matrix:
[[19  9]
 [ 0 43]]
Accuracy: 0.8732394366197183
Precision: 0.8269230769230769
Recall: 1.0
```