



清华大学

Tsinghua University

基于SVM的文本特征分类器

汇报人：李沐晟 指导老师：杨毅



清华大学

Tsinghua University

目录

contents

- 1 / 背景介绍与摘要
- 2 / 数据生成与处理
- 3 / 模型训练与判决
- 4 / 实验总结与思考

1-背景介绍与摘要

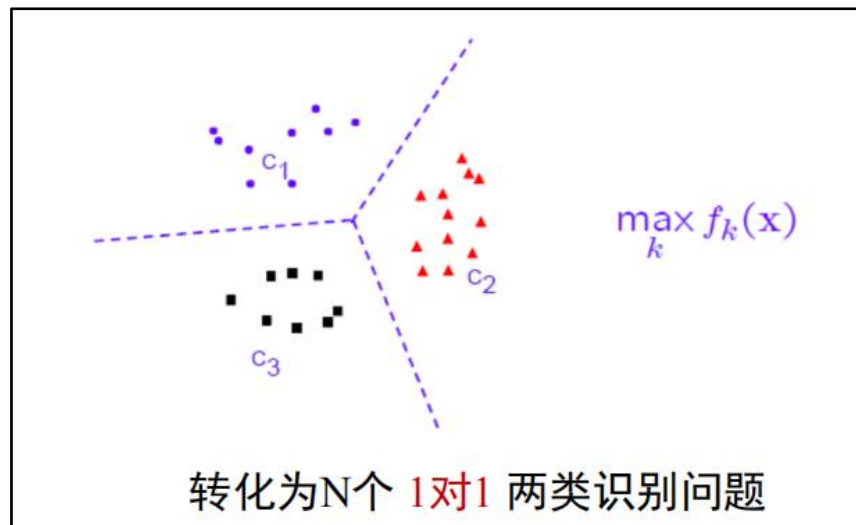
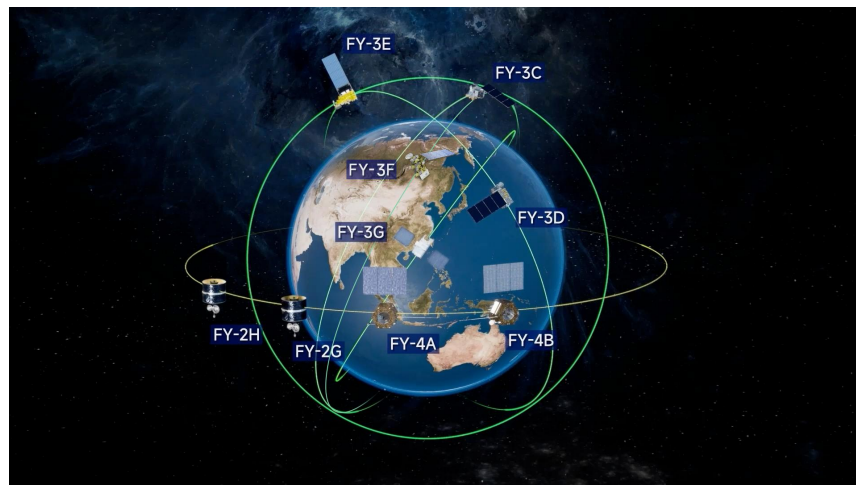


背景介绍

在当代，**太空飞行器**的应用已经成为人类探索宇宙、提升科技水平、推动经济发展的重要手段。然而，随着太空活动的增加，国际社会面临着日益严峻的挑战，确保外太空的和平与可持续发展变得至关重要。为此，一些重要的**国际公约**应运而生。这些公约为各国在外太空的活动设定了基本规则和行为准则。

实验概述

在本次实验中，我们以外太空探索中的**国际公约**为场景，筛选出其中的4条作为公约，利用大语言模型生成合成数据集。我们使用文本嵌入模型将文本转化为特征向量，并使用主成分分析(PCA)方法对其特征降维到2维。划定训练集、测试集与验证集后，我们通过训练支持向量机(SVM)以实现分类任务，并取得了非常良好的分类结果。为了实现多分类任务，我们采用一对一(OvO)方法，得到最终的判决结果。本方案在测试集中得到了100%分类准确率结果。

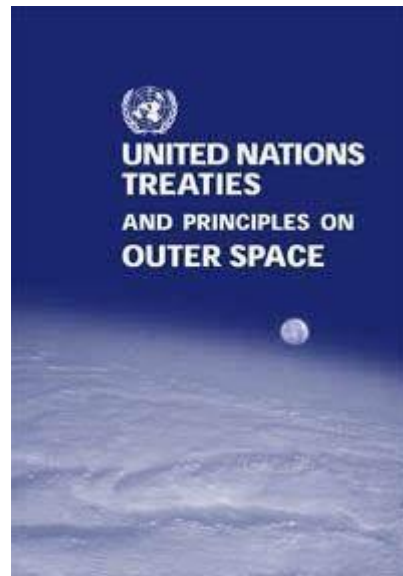


2-数据生成与处理



通过查阅网络相关文献以及大语言模型，了解到在外太空探索方面的国际公约有《外层空间条约》（Outer Space Treaty, 1967）、《月球协定》（Moon Agreement, 1984）、《空间物体造成损害的国际责任公约》（Liability Convention, 1972）、《外空物体登记公约》（Registration Convention, 1976）、《国际电信联盟公约》（ITU Convention）等等，经过选择与总结，将上述规则的内容整理为如下几条公约：

- 1、禁止在外太空部署核武器或其他大规模杀伤性武器。
- 2、外太空，不得被任何一个国家通过宣示主权、占领或其他手段据为己有。
- 3、各国应采取措施，减少和管理太空垃圾，以确保外太空环境的可持续利用。
- 4、各国必须向联合国登记其发射的航天器，以确保外太空活动的透明度。



随后利用大语言模型(ChatGPT-4o)针对上述公约**分别**生成70条违反的描述（如“某国在某情况下违反某条约”等），其中训练数据40条，验证数据10条，测试数据20条。对于**全部不违反**的描述类别，生成训练数据40条，验证数据20条，测试数据20条。

2-数据生成与处理



清华大学
Tsinghua University

部分数据预览

1.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

某国在紧张的国际局势下，将一颗携带核弹头的卫星发射到地球轨道。
某国为了威慑邻国，将多枚生物武器导弹部署在月球基地上。
某国在与敌对国家的冲突中，将化学武器送入近地轨道，作为战略威慑。
某国秘密将一个装有放射性物质的无人飞船送入火星轨道，以展示其军事实力。
某国在与另一国家的太空竞赛中，将一个携带氢弹的卫星送入地球同步轨道。
某国在国际军备竞赛中，发射了一颗装有神经毒气的卫星，进入低地轨道。
某国为了报复敌国，将一枚装有核弹头的导弹发射到国际空间站附近的轨道。
某国在战争期间，将几颗装有核武器的卫星送入太空，威胁对手国家。
某国为测试新型武器，将一个装有高能激光武器的卫星发射到地球轨道。
某国在隐秘军事行动中，将多个装有化学武器的无人飞行器部署在外太空。

3.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

某国在发射卫星时没有采取适当措施，导致火箭第一级残骸留在近地轨道上。
某国在太空中进行反卫星导弹试验，导致大量碎片散布在轨道中。
某国的老旧卫星失效后，没有采取任何清除措施，任其成为太空垃圾。
某国在发射探测器时，不采取减速和销毁轨道器的措施，导致探测器残骸留在轨道上。
某国进行大规模卫星发射，未考虑轨道拥堵问题，增加了太空垃圾风险。
某国在太空中进行爆炸性试验，产生了大量轨道碎片。
某国的火箭二级未能正常脱离轨道，成为新的太空垃圾。
某国发射的卫星发生碰撞，未采取任何后续处理措施。
某国的老旧通信卫星未及时退役，留在轨道上成为太空垃圾。

杀伤
武器

1

宣示
主权

2

太空
垃圾

3

登记
注册

4

2.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

某国在火星上插上国旗，宣布对该星球的所有权。
某国将其月球基地周围的区域划为军事禁区，禁止其他国家进入。
某国在与其他国家的太空竞赛中，宣布对地球同步轨道的控制权。
某国在火星上建立了多个科研站，并宣称该地区为其领土。
某国在月球南极发现水冰后，宣布对该地区的主权。
某国在国际空间站上宣布其控制的模块为国家领土。
某国在金星表面放置探测器后，宣示对该区域的主权。
某国在小行星带建立采矿基地，并宣布对该区域的控制权。
某国在建立月球基地后，将周围区域划为其专属经济区。
某国在与他国谈判中，声称对月球背面的所有权。
某国在火星上建立了殖民地，并宣布该地区为其领土。

4.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

某国在紧急军事行动中，未向联合国登记其发射的侦察卫星。
某国在秘密军事试验中，发射了一颗未登记的实验卫星。
某国在太空竞赛中，为了保密，未登记其最新发射的探测器。
某国发射了一颗监视卫星，但未向联合国通报登记。
某国在发射商用通信卫星时，未进行任何国际登记。
某国发射了一颗用于科学研究的卫星，但未向联合国报告。
某国在太空军事演习中，发射了未登记的间谍卫星。
某国发射了一颗小型立方卫星，但未进行登记。
某国在秘密任务中，发射了未登记的导航卫星。
某国在发射地球观测卫星时，未向联合国登记。

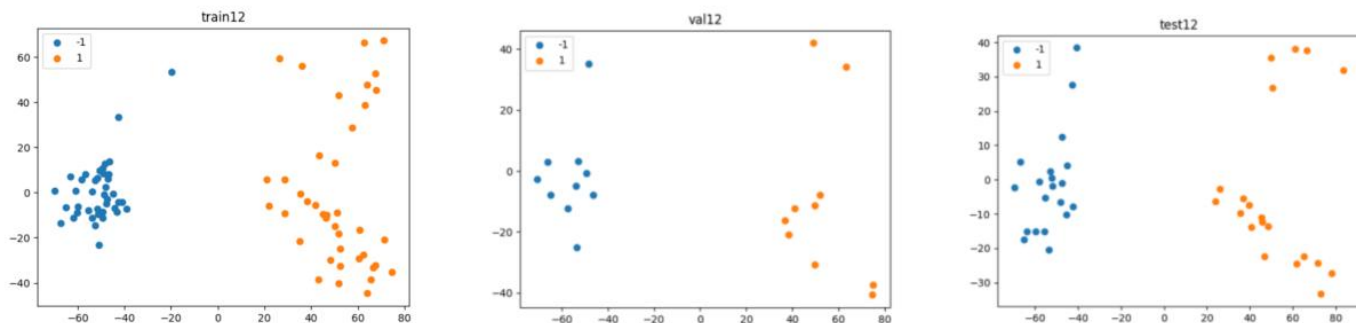
2-数据生成与处理



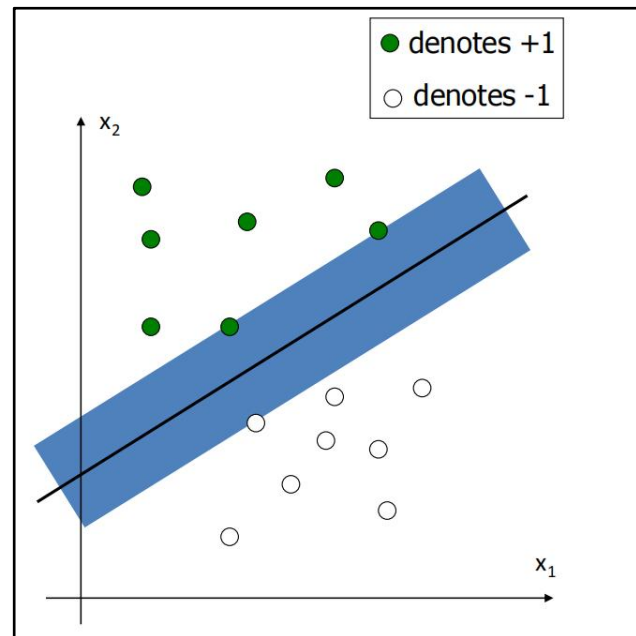
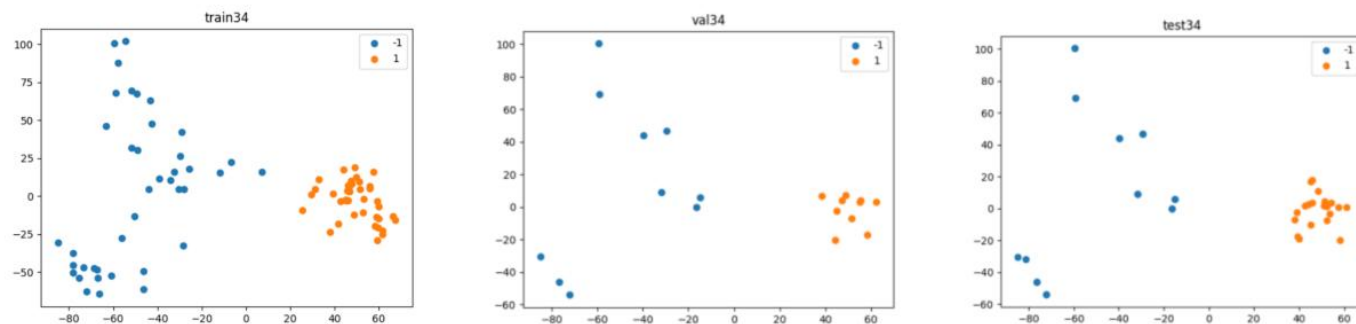
预处理可视化结果

将数据两两类别之间进行数据预处理（后续也将两两直接训练SVM分类器），可视化结果如下：

1-2两类



3-4两类



明显的“聚类”特征
能够进行分类

3-模型训练与判决



模型框架

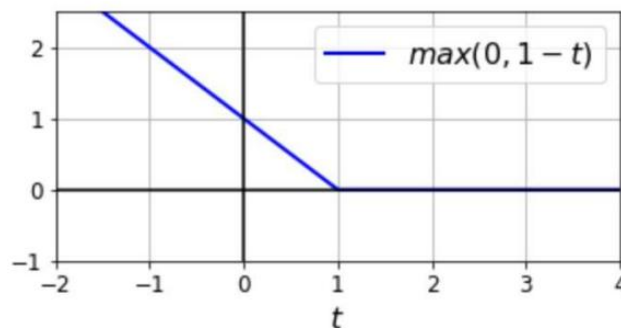
在本次实验中，我们采用SVM作为分类器，其网络结构在`svm_hw.py`中定义，主体由简单的线性层加上Hinge Loss功能构成。

```
19 class LinearFunction(torch.autograd.Function):
20     ...
21     ... we will implement the linear function:
22     ...  $y = xW^T + b$ 
23     ... as well as its gradient computation process
24     ...
25
26     ... @staticmethod
27 > ... def forward(ctx, x, W, b): ...
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44     ... @staticmethod
45 > ... def backward(ctx, grad_output): ...
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72 class Hinge(torch.autograd.Function):
73
74     ... @staticmethod
75 > ... def forward(ctx, output, W, label, C): ...
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96     ... @staticmethod
97 > ... def backward(ctx, grad_loss): ...
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
```

4.2 Hinge Loss图像

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \max(1 - y_i(w^T x_i + b), 0)$$

$$\text{hinge}(t) = \max(1 - t, 0)$$



```
117 class SVM_HINGE(nn.Module):
118
119 > ... def __init__(self, in_channels, C): ...
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138 > ... def forward(self, x, label=None):
139     ... # SVM calculation
140     ... output = LinearFunction.apply(x, self.W, self.b)
141     ... if label is not None:
142     ...     loss = Hinge.apply(output, self.W, label, self.C)
143     ... else:
144     ...     loss = None
145     ... output = (output > 0.0).type_as(x) * 2.0 - 1.0
146     ... return output, loss #返回SVM输出结果
```

3-模型训练与判决

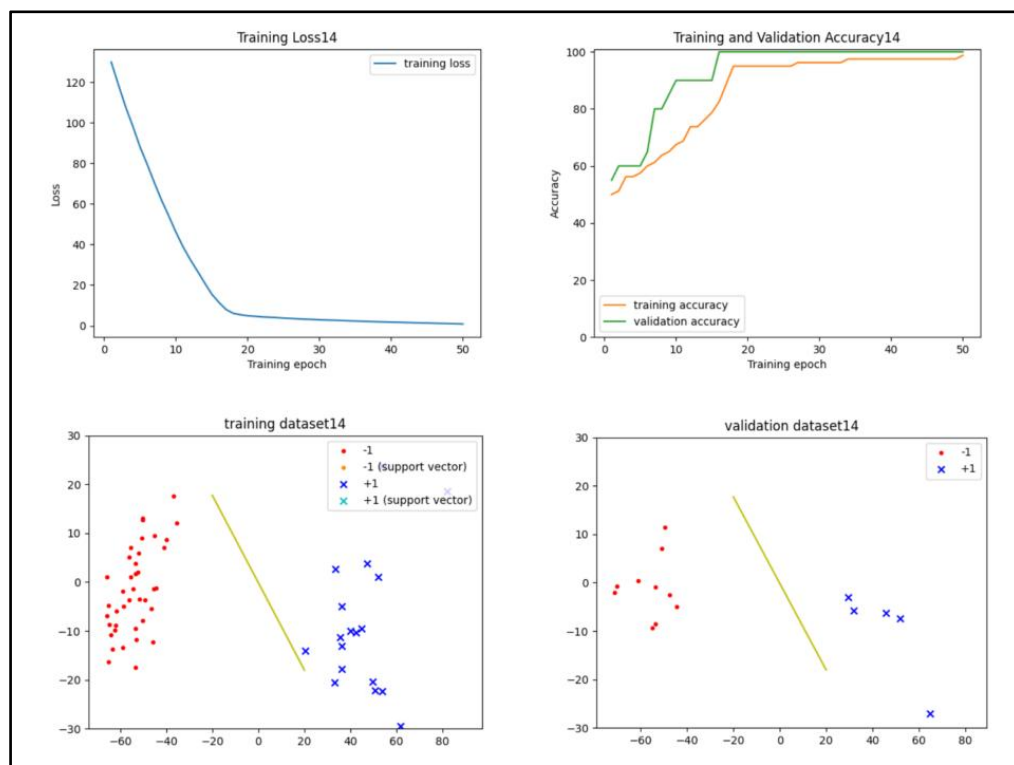
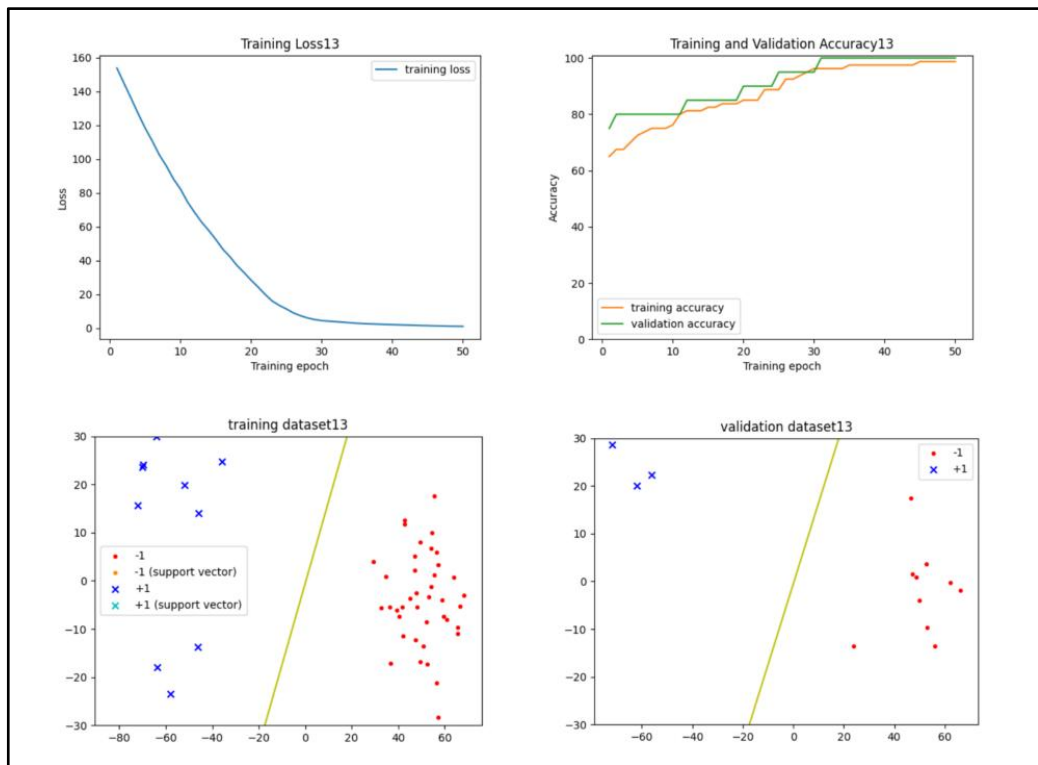


训练过程

正如上文提到的，我们在两两类别之间分别训练出一个SVM分类模型，总共有6组分类模型。经过调试与判断，设置训练轮数(epoch)为50、C因子(regularization coefficient)为0.1、学习率为0.01能取得理想的训练结果。经过六次训练，分别得到六个两两分类的SVM模型，储存在./checkpoints文件夹中。

训练结果

训练结果可视化展示结果如下（以1-3分类与1-4分类为例），可见训练效果良好。此外所有测试结果均为100%准确率。



3-模型训练与判决

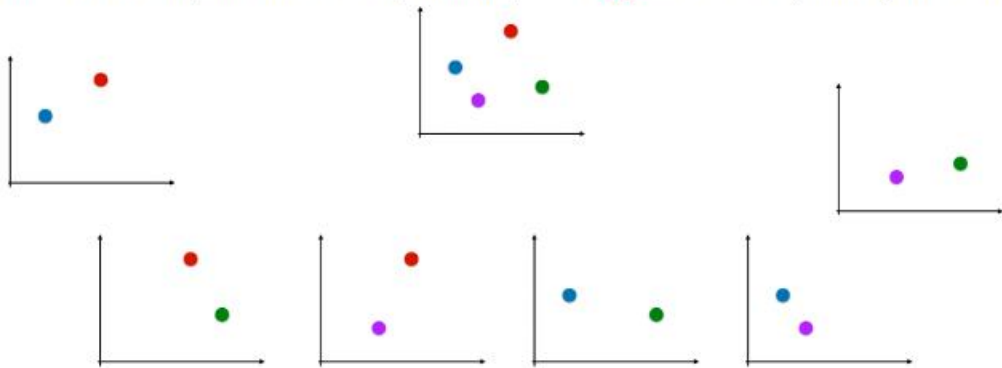


最终判决

为了对SVM进行扩展，实现多分类功能，我采取了一对一(OvO)方法，具体做法如下：对某一待分类文本，首先将其转化为特征向量并特征降维得到2维特征向量，随后使用每一个分类器(共6个)依次进行处理得到判决结果，最终使用“投票法”，统计出现次数最多的判决结果，作为最终的多分类判决结果(详见代码)。

2) OvO

- 思想：n 类样本中，每次挑出 2 种类型，两两结合，一共有 C_n^2 种二分类情况，使用 C_n^2 种模型预测样本类型，有 C_n^2 个预测结果，种类最多的那种样本类型，就认为是该样本最终的预测类型；



测试结果

针对4类违反不同公约的情景描述（共20*4=80条）进行分类测试，得到100%的准确率。这与分类界面清晰明确有关。

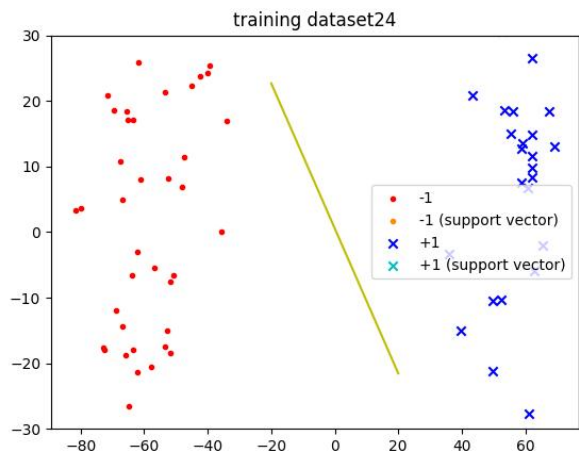
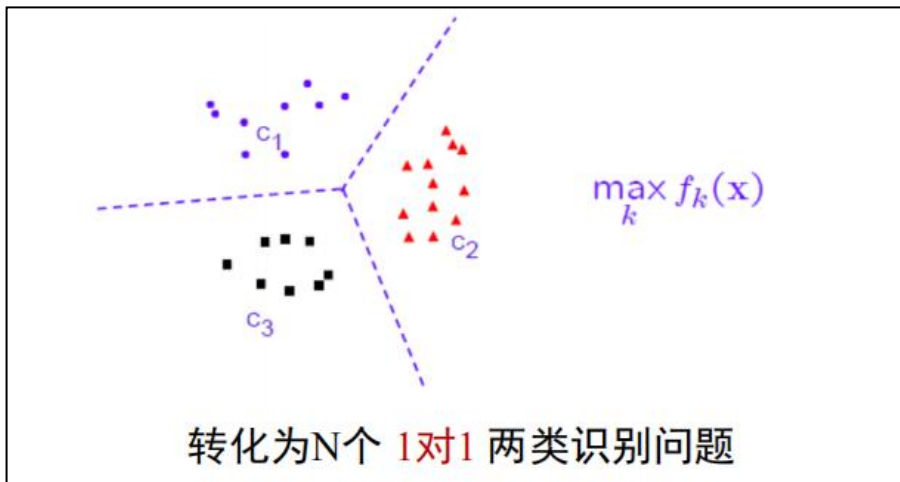
测试完毕...
Test accuracy = 100.0%

4-实验总结与思考

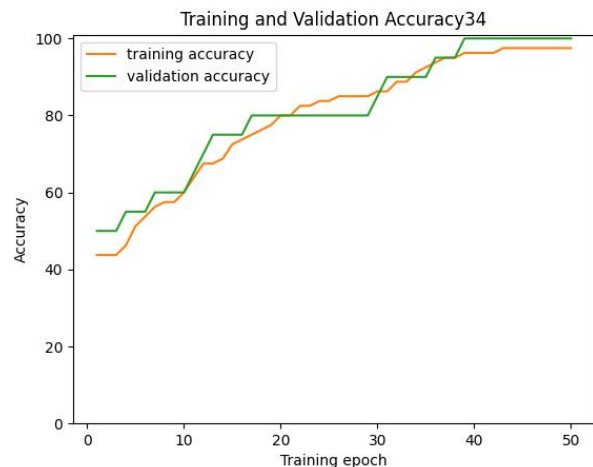
实验总结

在本次实验中，本人成功利用支持向量机（SVM）模型实现了文本特征的分类任务。总结而言，在数据生成与预处理方面，本实验通过查阅相关文献并利用大语言模型（如ChatGPT-4）生成了符合特定公约的违反描述，并将其嵌入为高维特征向量表示。随后，通过主成分分析（PCA）方法将特征降维为二维，这为后续的SVM分类奠定了基础；在模型训练与测试方面，采用一对一（OvO）策略，在每两类之间分别训练SVM模型，共训练了六组分类模型。通过合理设置超参数（如训练轮数、正则化系数等），得到了训练准确率和测试准确率均接近100%的分类结果；最终的多分类任务通过“逐一分类+投票法”的策略实现，测试结果显示所有测试数据均被正确分类，表现出非常高的分类性能。

	主成分分析PCA
优化问题	$\begin{aligned} \min & \{\varepsilon(M)\} \\ \text{s.t. } & \Phi_i^T \Phi_i - 1 = 0 \end{aligned}$



```
Epoch 47: loss = 0.556, training accuracy = 100.0%
Epoch 47: validation accuracy = 100.0%
Epoch 48: loss = 0.555, training accuracy = 100.0%
Epoch 48: validation accuracy = 100.0%
Epoch 49: loss = 0.554, training accuracy = 100.0%
Epoch 49: validation accuracy = 100.0%
Epoch 50: loss = 0.553, training accuracy = 100.0%
Epoch 50: validation accuracy = 100.0%
Model saved in checkpoints/svm24.pth
```



测试完毕...
Test accuracy = 100.0%

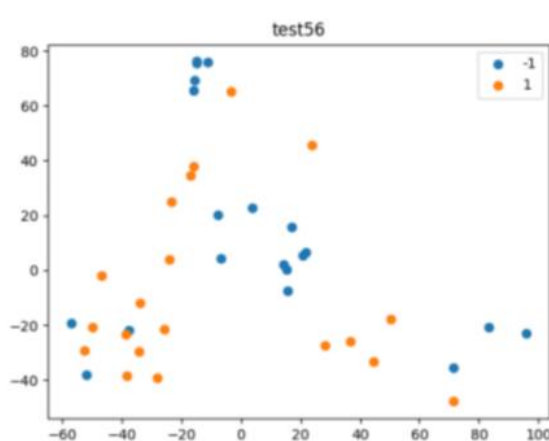
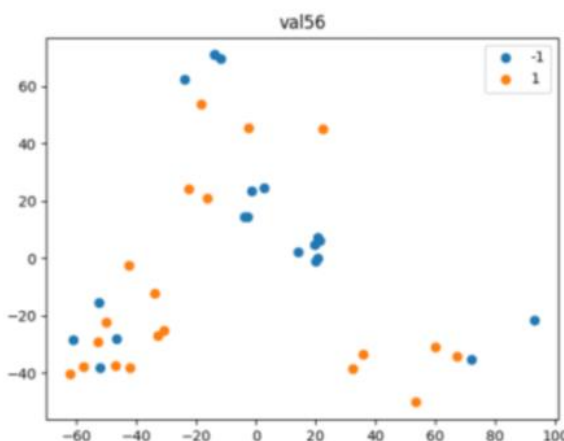
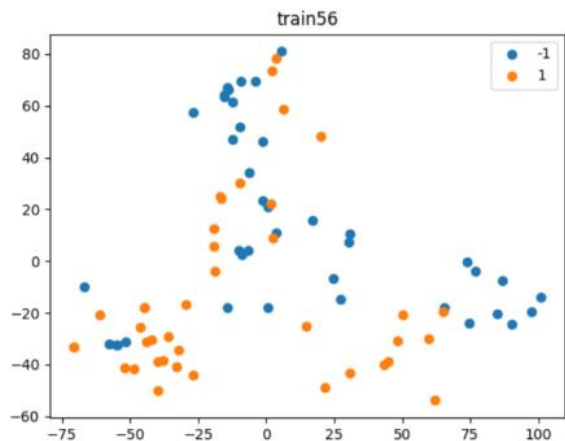
4-实验总结与思考



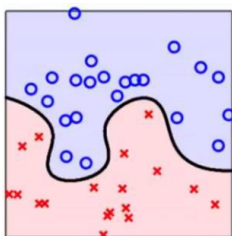
不足与后续改进

本次实验中的分类器与其说是针对“是否违反某一条公约”的分类器，实质上更是针对“围绕不同条公约的描述”的分类器，具体可以从处理“是/否违反公约”的情景描述数据预处理中看出。

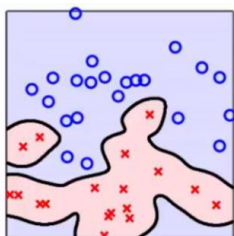
在下面可视化图像中，5类代表不违反任何一条公约的描述，6类表示违反四条公约中某一条的描述，实际上这样的数据是杂糅在一起的，“是否违反”这一语义特征被“描述主题”这一语义特征所掩盖。



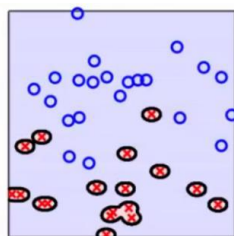
当然，我们也可以看出第5和第6类的样本点是存在一定**偏移**且存在**分离趋势**的，故理论上是可分的。后续我们可以借助核函数、MLP等等非线性方法实现进一步的分类（非线性分类）。



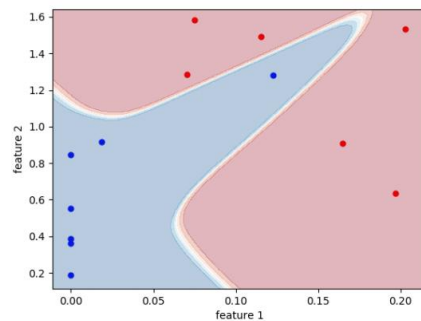
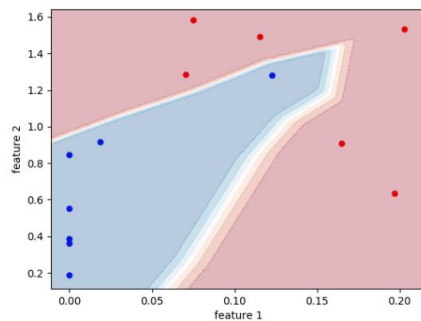
$$\exp(-1\|\mathbf{x} - \mathbf{x}'\|^2)$$



$$\exp(-10\|\mathbf{x} - \mathbf{x}'\|^2)$$



$$\exp(-100\|\mathbf{x} - \mathbf{x}'\|^2)$$





清华大学

Tsinghua University

感谢聆听，敬请批评指正

汇报人：李沐晟 指导老师：杨毅