

機器學習 HW3

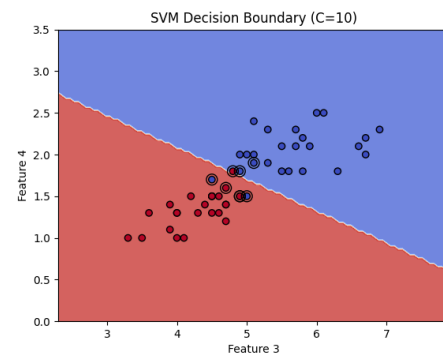
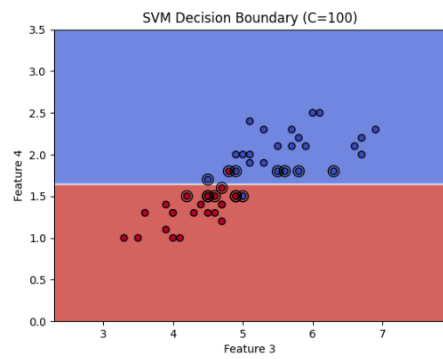
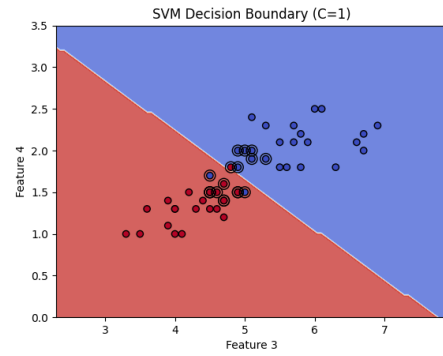
系所:電控碩一

學號:313512022

姓名:呂靖樑

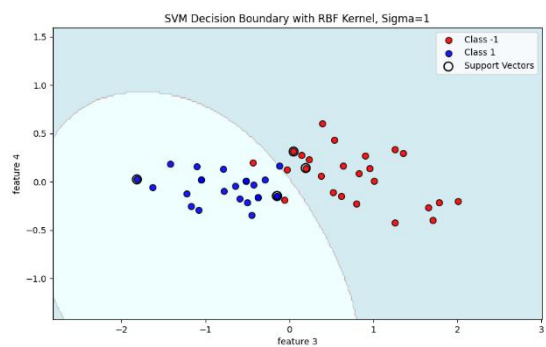
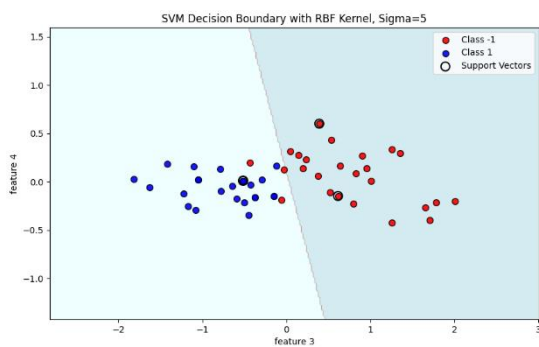
Part1

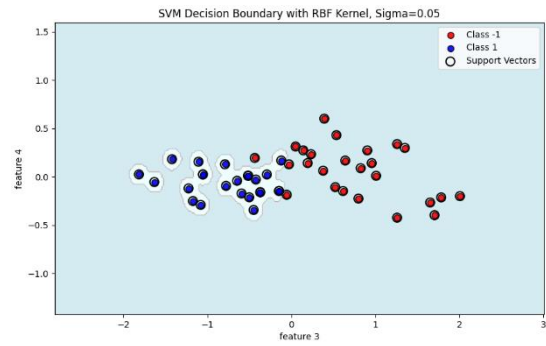
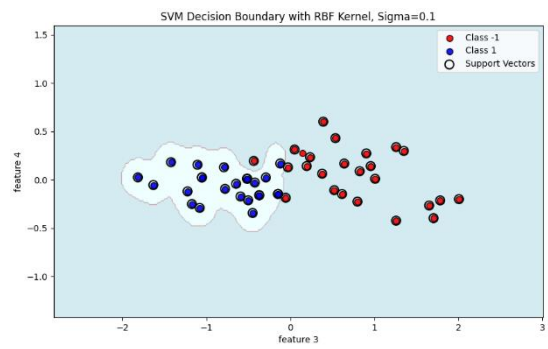
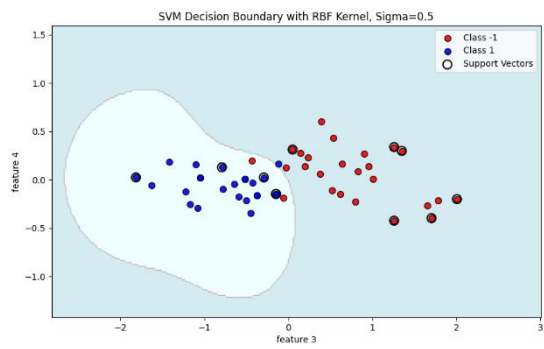
Linear SVM			
c = 1			
alpha	total alpha	b	CR
1 0.0667 1 0 1	14.4001	10.5376	94%
c = 10			
alpha	total alpha	b	CR
0 0 9 0 0	76	15.14	94%
c = 100			
alpha	total alpha	b	CR
0 0.0056 44.4307 0 0.0076	577.7776	11.1071	92%



Part2

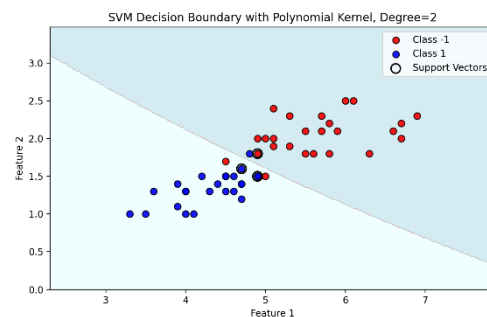
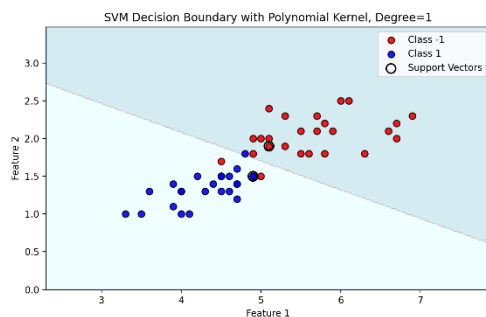
RBF Kernel SVM (C = 10)							
sigma = 5				sigma = 0.1			
alpha	total alpha	b	CR	alpha	total alpha	b	CR
10	193.7714	0.0678	90%	0.2801	36.0574	-0.1594	88%
8.9619				0.3292			
0				1.2411			
10				0.3631			
0				0.0704			
sigma = 1				sigma = 0.05			
alpha	total alpha	b	CR	alpha	total alpha	b	CR
0	76.2554	-0.2189	96%	0.5377	42.7547	-0.1106	82%
0				0.3213			
8.8907				0.6269			
0				0.5453			
0				0.9401			
sigma = 0.5							
alpha	total alpha	b	CR				
0	62.7532	-0.2439	94%				
0							
5.9852							
0							
0							

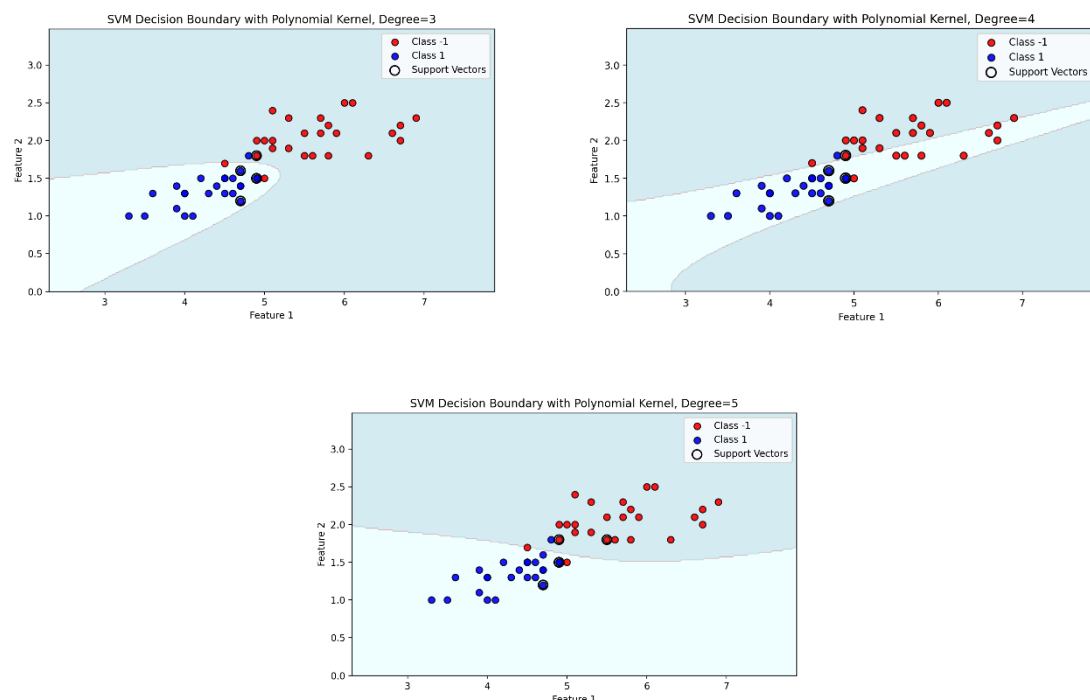




Part3

Polynomial Kernel SVM (C = 10)							
poly = 1				poly = 4			
alpha	total alpha	b	CR	alpha	total alpha	b	CR
0	76	15.14	94%	0	49.1488	1.2659	86%
0				0			
9				3.2054			
0				0			
0				0			
poly = 2				poly = 5			
alpha	total alpha	b	CR	alpha	total alpha	b	CR
0	54.2595	13.6664	94%	0	47.91	26.3713	92%
0				0			
3.8295				3.5334			
0				0			
0				0			
poly = 3							
alpha	total alpha	b	CR				
0	53.0684	6.3886	96%				
0							
3.8728							
0							
0							





Part4

1.

Linear SVM 找到的是一條直線或平面作為決策邊界，適合線性可分的數據；而 kernel-based SVM 使用核函數將數據映射到高維空間，找到非線性決策邊界，能處理更複雜的數據分佈。通過使用 PCA 等技術，我們可以將在高維空間中找到的決策邊界映射回二維或三維空間中，這樣可以更直觀地觀察邊界如何分隔數據，從而更好理解分類模型的效果。

2.

從 Part 2 和 Part 3 的結果可以看出，隨著核參數的變化，模型的支持向量數量、total alpha 以及分類率的變化情況。

在 Part 2 中，隨著 sigma 的下降，支持向量 SV 的數量增加，但 total alpha 反而減少。這表明模型變得更 overfitting，導致分類率下降。當 sigma 過小時，模型邊界過於靈活，使得模型對訓練數據過度適應，但無法很好地泛化到新的數據上，最終導致分類率不佳。而在 Part 3 中，隨著多項式次數 (poly) 的上升，模型的決策邊界變得更加複雜，total alpha 呈現下降趨勢，支持向量的影響也隨之減小。隨著多項式次數的增加，模型開始過度適應訓練數據，出現 overfitting 現象，分類率在 poly = 4 和 poly = 5 時明顯下降，這也表明模型在複雜數據上的泛化能力不足。

綜合來看，無論是 RBF 核中的 sigma，還是 polynomial 核中的 poly，核參數的選擇都會顯著影響模型的邊界複雜度和泛化能力。找到適當的核參數至關重要，能有效地平衡模型的複雜度，避免 overfitting 或 underfitting，從而提高分類效果。

3.

為了避免 SVM 模型中的超平面 overfitting，可以採取幾種方法來控制模型的複雜度。首先，可以調整參數 C ，該參數控制模型對誤分類的容忍度。較大的 C 使模型更加靈活，但容易 overfitting，而較小的 C 則會增加邊界的平滑性，防止 overfitting。其次，可以適當選擇核函數的參數，例如 RBF 核的 σ 。較小的 σ 使模型邊界過於靈活，容易 overfitting，適當增加 σ 值可以使決策邊界平滑，提升泛化能力。對於 polynomial 核，次數 degree 的增大也會導致 overfitting，因此需要謹慎選擇。

使用交叉驗證來尋找最佳的參數組合。交叉驗證能評估不同參數在訓練和測試數據上的表現，從而找到適合的參數。

增加訓練數據量、進行特徵選擇和使用標準化技術，都能幫助減少模型的複雜度，提升泛化能力。這些方法的結合可以有效防止 SVM 模型 overfitting，從而達到更穩健的分類效果。