

1. 目的：利用鳶尾花資料(Iris data set)來訓練 Support Vector Machine (SVM)，並分別比較線性與非線性 SVM 所訓練的超平面(hyperplane)有何差異。欲分析的數據集: Iris dataset (150 data points from 3 different classes)
2. 作業內容:

Part1: Linear SVM (Initialization: $C=1$)

Step1: 將Iris data set的變色鳶尾(Versicolor, label=2)以及維吉尼亞鳶尾(Virginica, label=3)分別設為positive class與negative class，並採用花瓣長度(3rd feature)與花瓣寬度(4th feature)作為特徵。

Step2: 將positive class與negative class的前25筆data設為training data，剩餘的50筆設為test data。

Step3: 實現SVM分類器（請勿直接使用開源的SVM套件）

補充：在求解對偶問題（Dual problem）時，可以利用python的開源程式(qpsolvers: <https://pypi.org/project/qpsolvers/>)，並請自行將訓練資料矩陣化並參閱前兩者的documentation調整輸入的形式。

Step4: 將training data送至SVM中訓練，求解Lagrange multiplier - alpha，再透過Kuhn-Tucker(KT) condition分析並求解bias。(紀錄底線標記之結果，數值請列到小數第四位)。

Step5: 將test data輸入至Step4所訓練的SVM decision function，對50筆test data分別進行分類決策，並記錄test data的分類率。

Step6: 將penalty weight C分別改為10以及100並重複Step4 – Step5。

Part2: RBF kernel-based SVM (Initialization: $C=10$, $\sigma=5$)

Step1 – Step5: 同Part1中Step1 – Step5流程，但採用RBF kernel。

Step6: 將sigma分別改為1、0.5、0.1以及0.05並重複Step4 – Step5。

Part3: Polynomial kernel-based SVM (Initialization: $C=10$, $p=1$)

Step1 – Step5: 同Part1中Step1 – Step5流程，但採用Polynomial kernel。

Step6: 將 p 分別改為 2、3、4、5 並重複 Step4 – Step5。

Part4 : Discussion and results presenting

請試著討論以下問題：

1. Linear SVM與kernel-based SVM所訓練的hyperplane有何差異？
2. 隨著kernel parameter的改變，RBF kernel與polynomial kernel所訓練的hyperplane可能有什麼變化？其與分類率的變化有何關聯？請嘗試解釋之。
3. 設定kernel parameter時，是否有方法避免hyperplane過度擬合(overfitting)的現象發生？若有請詳細討論。

將上列實驗之結果與討論事項整理後做詳細的討論，並以書面報告呈現。

3. 繳交期限: **2025/10/27(一) 23:59** (“Two” weeks from now)。
4. 繳交方式: 上傳至 E3 教學平台。
5. 注意事項:
 - 程式語言限以 python 撰寫。
 - 分類器請使用手刻方式完成，勿使用現成套件。
 - 遲交三天內該次作業分數打五折，遲交超過 72 小時以 0 分計算。
 - 作業報告請以書面 pdf 呈現，並將程式碼一併壓縮在一個壓縮檔中。
 - 壓縮檔名請符合下列格式：「作業_學號_姓名」(如：HW2_313512002_蕾潔)，否則作業成績打 8 折。