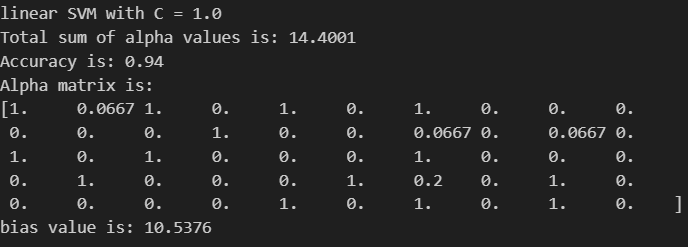
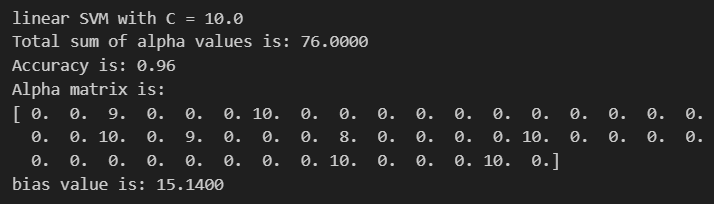
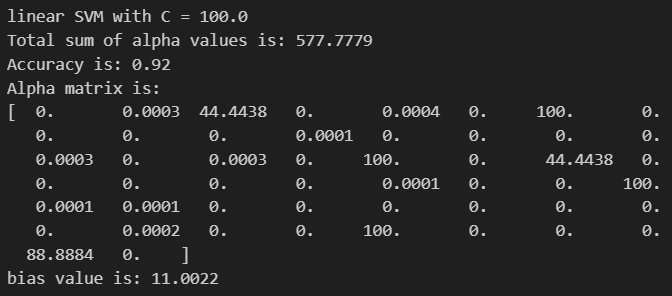
機器學習 HW3 312512011 李效賢

* Linear Kernel

1.C=1  


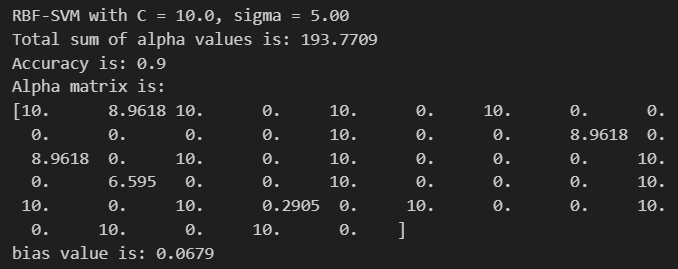
2.C=10  


3.C=1

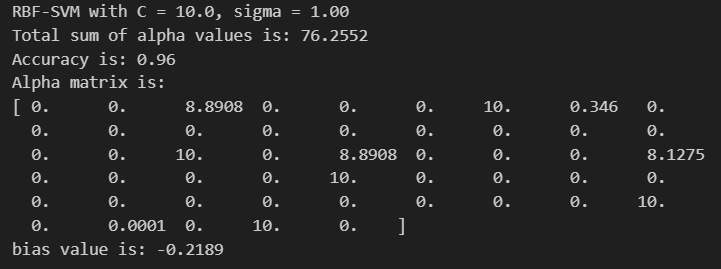
|  |  |  |  |
| --- | --- | --- | --- |
| C value | Bias | Alpha sum | Accuracy |
| 1 | 10.5376 | 14.4001 | 94% |
| 10 | 15.1400 | 76.0000 | 96% |
| 100 | 11.0022 | 577.7779 | 92% |

* RBF Kernel (Gaussian function) ( C=10固定)

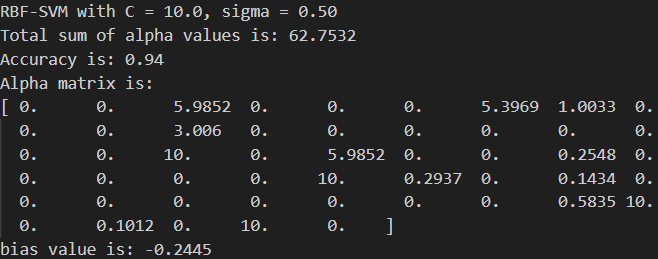
1. Sigma=5.0



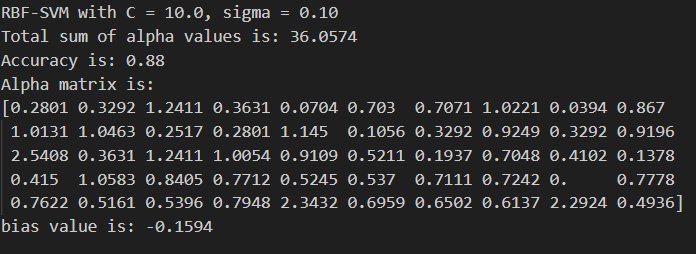
2. Sigma=1.0



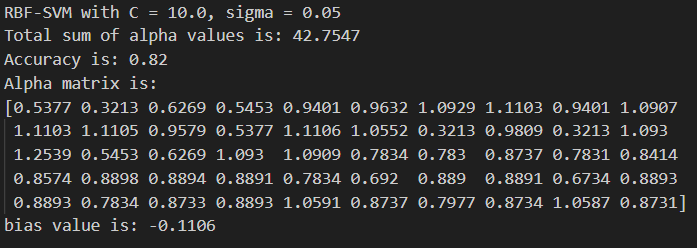
3. Sigma = 0.5



4. Sigma = 0.1



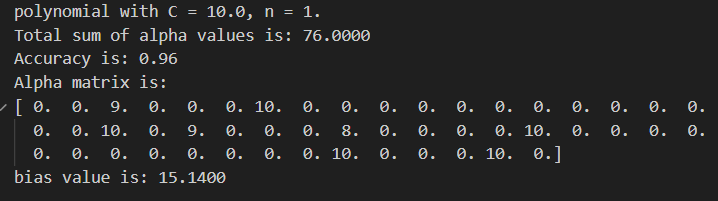
5. Sigma = 0.05



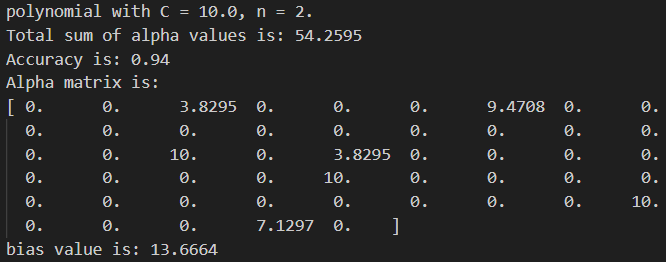
|  |  |  |  |
| --- | --- | --- | --- |
| Sigma value | Bias | Alpha sum | Accuracy |
| 5.0 | 0.0679 | 193.7709 | 90% |
| 1.0 | -0.2189 | 76.2552 | 96% |
| 0.5 | -0.2445 | 62.7532 | 94% |
| 0.1 | -0.1594 | 36.0574 | 88% |
| 0.05 | -0.1106 | 42.7547 | 82% |

* Polynomial Kernel (向量內積的n次方，C值固定為10)

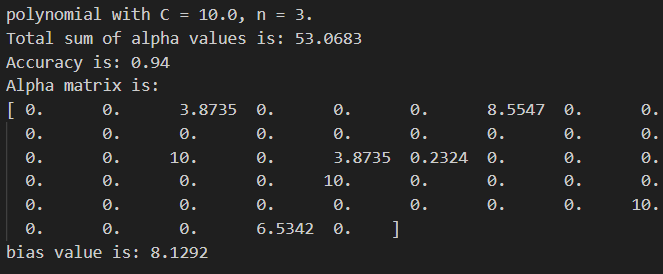
1. N=1 (計算alpha時意同linear kernel)



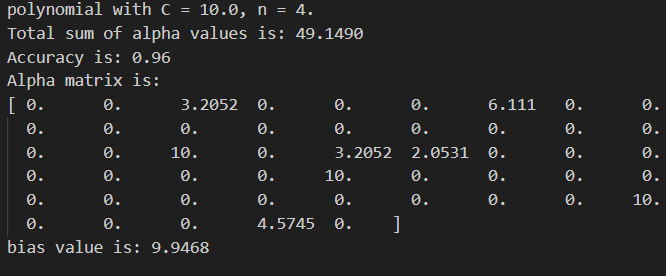
2. N=2



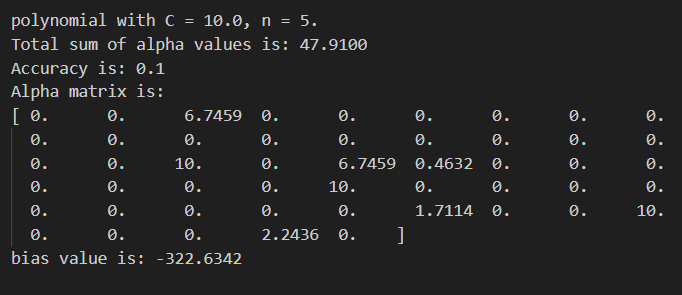
3. N=3



4. N=4



5. N=5



|  |  |  |  |
| --- | --- | --- | --- |
| N | Alpha sum | Bias | Accuracy |
| 1 | 76.0000 | 15.1400 | 96% |
| 2 | 54.2595 | 13.6664 | 94% |
| 3 | 53.0683 | 8.1292 | 94% |
| 4 | 49.1490 | 9.9468 | 96% |
| 5 | 47.9100 | -322.6342 | 10% |

問題與討論：

1. Linear SVM與kernel-based SVM所訓練的hyperplane有何差異?

Ans :

1.Linear SVM: 訓練出的Hyperplane為線性，decision boundary為線性，處理的是「線性可分離」(Linearly separable)或是近似線性可分離，具有較低的計算複雜度，適合處理相對簡單的數據分類問題

2.Kernel-based SVM:訓練的超平面為非線性，使用的Kernel function為RBF(高斯函數)或著多項式(內積的n次方)。可以把原空間中的特徵映射到較高維度進行分類，意味著此方式訓練出的模型，能夠處理原空間中非線性可分離的數據，然而具有較大的計算量和資源需求。

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。

Ans :

(1)改變RBF kernel的sigma值，表示將映射到不同的高維空間中進行分類。如果sigma較小，決策邊界會較複雜(kernel function的值衰減越快)，也就是有overfitting的問題；若sigma值較大，決策邊界會變得較簡單，也就是較傾向underfitting的一端。

Polynomial kernel則決定決策邊界的多項式次數，n越大意味著越高次方的多項式，也就是hyperplane會變得更加複雜，更傾向overfitting，反之亦然。

(2)Overfitting和Underfitting兩個狀況皆會使測試集的分類準確度下降，如何在兩者直接找到最佳的平衡點，亦是另一個AI演算法應用過程最佳化的重要議題，使用課堂所講述的格子搜尋法可以嘗試找到自己設定的組合中可能的最佳參數，但格子需要「切多細」，就需要依照當下電腦的性能、資源量，還有訓練測試數據等等特性做微調，也就是case by case。

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。

Ans :

在SVM中調整kernel function的參數以避免overfitting是其中一項重要議題(當然在所有的AI應用中都是)。

以下有幾種方法避免SVM函數中overfitting的現象：

1.k-fold Cross-validation(k折交叉驗證)：避免kernel將某些訓練集的特徵過度放大，選擇平均最高而非單一排列組合正確率最高者。

2.調整C ,sigma, n變數:參數C控制了SVM對於邊界的容忍度，較小的C值會導致更寬鬆的決策邊界，較大的C值會導致更嚴格的決策邊界。通過調整C的值，可以控制模型的複雜性，較小的C值通常有助於防止過度擬合。同樣的較小的sigma也意味著overfitting。而使用課堂上所教的網格搜尋法，透過不同的參數進行排列組合，可以找出最高分類率的組合，並且因為知道各個參數所代表的意義(背景知識)，可以比盲目亂試更有效率避免這樣的問題。

3.特徵選擇和降維：可以利用課堂上提及的Fishers等方式，選擇可分離度較大的特徵，或是選擇不同的特徵組合(就像前次作業4取1、4取2、4取3等等不同方式，嘗試出較佳的特徵組合。另一種方式可以使用課堂上提及的PCA等方式，先進行前處理後再進入演算法的訓練，降低複雜度的同時也可以避免overfitting

5.Big data:當使用數據量大的同時，可以避免小樣本數造成的統計偏差，這也是近年大數據AI火熱的原因之一，因為可以直接避掉許多統計上的極端值偏差。

6.使用不同的Kernel function:因為不同的Kernel function會將數據集映射到不同的高維空間，基於每組數據組的特性不同，不同的映射方式會讓分類率有所差異，可能A映射方式過度分類的問題，在B映射方式完全不會出現。