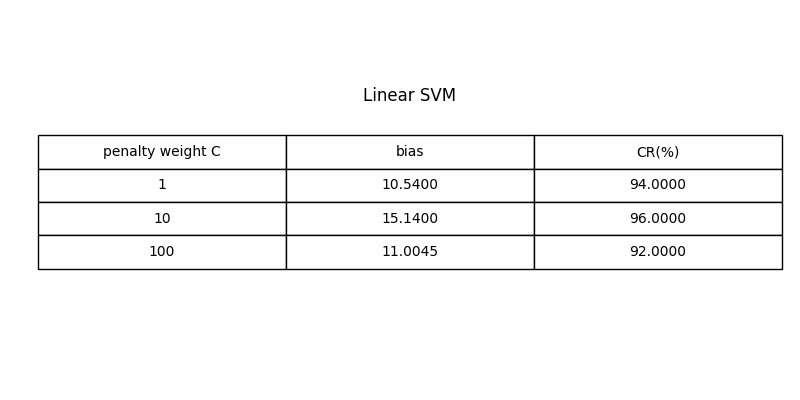
Machine Learning HW3

312512005 黃名諄

1. Linear SVM:

* alpha:

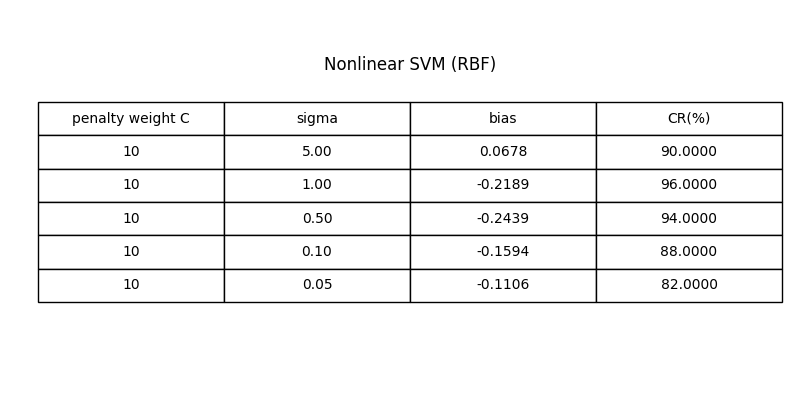
|  |  |
| --- | --- |
| C=1 |  |
| C=10 |  |
| C=100 |  |

* bias and CR(%) :

1. RBF kernel-based SVM (C=10):

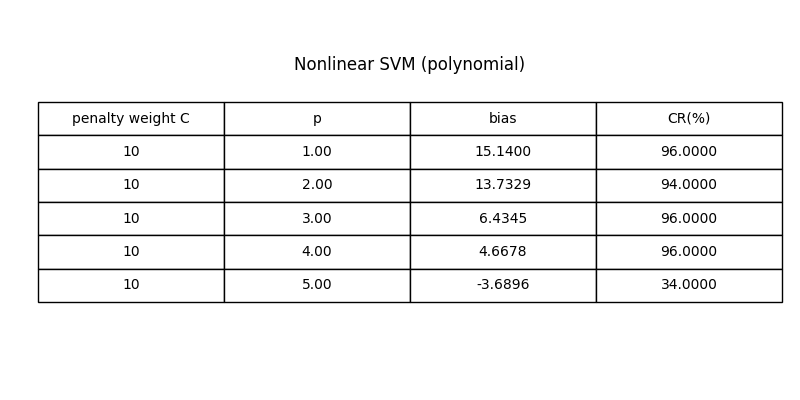
* alpha:

|  |  |
| --- | --- |
| sigma=5 |  |
| sigma=1 |  |
| sigma=0.5 |  |
| sigma=0.1 |  |
| sigma=0.05 |  |

* bias and CR(%) :

1. Polynomial kernel-based SVM (C=10):
   1. alpha:

|  |  |
| --- | --- |
| p=1 |  |
| p=2 |  |
| p=3 |  |
| p=4 |  |
| p=5 |  |

* bias and CR(%) :

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

Ans: Linear SVM是只在原特徵空間中去找hyperplane，而kernel-based SVM是將特徵映射到更高維空間中，再去找最佳分離hyperplane，在原特徵空間不好分離的兩類資料，映射到高維空間可能就能有較明顯的分離情況。

* 1. 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: 改變kernel parameter，相當於將映射到不同空間中找分離hyperplane，造成不同的decision function結果，使模型及分類率有所變化，具體來說，在RBF kernel情況下，sigma越大，標準差越大，RBF kernel更平緩而不集中，代表泛化能力較高，但分類能力可能就不是最好的；反之，sigma越小，其標準差越小，RBF kernel集中於局部，可能導致model過擬合，使其對test data無法有效分類，使CR降低，在實驗結果中確實有這樣的趨勢驗證此觀點。另一方面對於polynomial kernel，p越小表多項式次數較小，使model較簡單平滑，泛化性較高，但相對分類能力可能不是最好的；反之，p較大表多項式次數較大，model及決策邊界更複雜，能更貼近描述training data關係，但反而會使model過擬合，使在test data上分類率下降，像實驗結果中p=5之分類率就掉到34%，我還有試p=10，分類率剩28%，驗證了上述過擬合的情況。

* 1. 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: 可使用格子搜尋法及k-fold CV去尋找有最好分類表現的參數，避免某一參數訓練之模型發現過擬合情形，且由題2.之討論，也可大概知道若要避免過擬合發生，在RBF kernel情況下時，要避免 sigma太小；polynomial kernel情況下，則是要避免p過大。

* 1. 我和同學在實驗中發現，因助教有加上一段對alpha的處理及四捨五入至第6位，所以後續model計算我是以四捨五入後之alpha做的，這個四捨五入位數之影響很大，我試過取4位時，在polynomial kernel p=5下CR結果就差了20%，這是我發現的現象，對分類造成的影響似乎挺大的。