

Machine Learning HW4

312512005 黃名諄

1. SVM grid search result:

在做 grid search 時，會有某些參數組合做 one-against-one strategy 時，3 個模型中可能會有某幾個模型因沒有 alpha 在 0 和 C 之間，導致無法算出 bias，模型無法預測，因此我的處理方式是若當組參數組合的 3 個 model 中只要有一個沒有辦法算出 bias 就放棄此組 grid 參數，令此組參數 CR=nan 來處理；且最後有多組參數組合都能達到最佳的 CR，因此我都有印出來(row 是 sigma，column 是 C)

	1	5	10	50	100	500	1000
1.05^-100	54.00	54.00	54.00	54.00	54.00	54.00	54.00
1.05^-95	66.67	66.67	66.67	66.67	66.67	66.67	66.67
1.05^-90	49.33	49.33	49.33	49.33	49.33	49.33	49.33
1.05^-85	51.33	51.33	51.33	51.33	51.33	51.33	51.33
1.05^-80	51.33	51.33	51.33	51.33	51.33	51.33	51.33
1.05^-75	51.33	51.33	51.33	51.33	51.33	51.33	51.33
1.05^-70	41.33	41.33	41.33	41.33	41.33	41.33	41.33
1.05^-65	45.33	49.33	49.33	49.33	49.33	49.33	49.33
1.05^-60	54.00	56.00	56.00	56.00	56.00	56.00	56.00
1.05^-55	63.33	63.33	63.33	63.33	63.33	63.33	63.33
1.05^-50	65.33	66.00	66.00	66.00	66.00	66.00	66.00
1.05^-45	74.00	74.67	74.67	74.67	74.67	74.67	74.67
1.05^-40	84.67	85.33	85.33	85.33	85.33	85.33	85.33
1.05^-35	90.67	90.00	90.00	90.00	90.00	90.00	90.00
1.05^-30	94.00	92.67	92.67	92.67	92.67	92.67	92.67
1.05^-25	95.33	95.33	95.33	95.33	95.33	95.33	95.33
1.05^-20	95.33	94.67	95.33	95.33	95.33	95.33	95.33
1.05^-15	95.33	95.33	95.33	96.67	96.67	96.67	96.67
1.05^-10	94.67	95.33	95.33	96.00	96.67	96.67	96.67
1.05^-5	96.00	95.33	95.33	94.67	94.67	95.33	95.33
1.05^0	95.33	95.33	94.67	95.33	95.33	96.00	96.00
1.05^5	96.00	95.33	96.67	96.00	96.67	95.33	96.00
1.05^10	95.33	96.00	96.00	94.67	96.00	95.33	94.67
1.05^15	95.33	96.00	95.33	94.00	95.33	95.33	94.67
1.05^20	95.33	96.00	96.00	97.33	94.67	95.33	95.33
1.05^25	94.67	95.33	96.00	98.00	97.33	95.33	95.33
1.05^30	94.67	96.67	95.33	96.00	98.00	96.00	94.67
1.05^35	93.33	95.33	98.00	96.67	98.67	96.67	96.00
1.05^40	NaN	95.33	96.00	96.67	96.67	98.00	96.00
1.05^45	NaN	94.67	95.33	96.00	96.67	98.67	97.33
1.05^50	NaN	93.33	94.67	95.33	97.33	98.67	98.00
1.05^55	NaN	90.67	93.33	96.67	96.00	96.67	98.67
1.05^60	NaN	90.67	92.67	95.33	97.33	96.67	98.67
1.05^65	NaN	NaN	NaN	94.67	94.00	97.33	96.67
1.05^70	NaN	NaN	92.67	93.33	94.67	95.33	96.67
1.05^75	NaN	NaN	NaN	92.67	95.33	98.67	96.67
1.05^80	NaN	NaN	NaN	92.00	93.33	96.67	95.33
1.05^85	NaN	NaN	NaN	94.00	90.67	95.33	97.33
1.05^90	NaN	NaN	NaN	NaN	92.00	94.67	95.33
1.05^95	NaN	NaN	NaN	NaN	90.00	94.67	95.33
1.05^100	NaN	NaN	NaN	NaN	NaN	92.00	94.67
Optimal parameters and CR are : [C = 100, sigma = 1.05^35 , best CR = 98.67]							
Optimal parameters and CR are : [C = 500, sigma = 1.05^45 , best CR = 98.67]							
Optimal parameters and CR are : [C = 500, sigma = 1.05^50 , best CR = 98.67]							
Optimal parameters and CR are : [C = 1000, sigma = 1.05^55 , best CR = 98.67]							
Optimal parameters and CR are : [C = 1000, sigma = 1.05^60 , best CR = 98.67]							
Optimal parameters and CR are : [C = 500, sigma = 1.05^75 , best CR = 98.67]							

2. 結果討論:

1. 請問在 grid search 的結果中，C 的大小與分類率高低有何關係？

Ans: C 的意義是懲罰分類錯誤，所以越大表示模型對於 training data 的分類錯誤越重視，也就是可能會過擬合 training data；相反 C 越小泛化能力較大但也可能造成欠擬合發生。

從我做出來的結果來分析，可以發現其實 C 對分類率都是小幅度的影響，主要還是 sigma 影響較大，而在 C 由小到大排序下中後段的分類率較高，可驗證不能使 C 太大太小而過擬合或欠擬合，因此符合理想上 C 中後段會有個分類率最高的峰值。

2. sigma 的大小的改變與分類率是否有關係？若有，請探討 sigma 的差異與特徵的數值有什麼關聯性？

Ans: 從結果上來看，sigma 對分類率的影響較大，sigma 的影響我在 hw3 也有提到，sigma 越大，標準差越大，RBF kernel 更平緩而不集中，我的特徵數值帶入都分散，導致了泛化能力或許較高，但也可能欠擬合；另一方面若 sigma 越小，標準差越小，RBF kernel 集中於局部區域，我的特徵數值帶入都集中於某處，導致過擬合發生，泛化能力低，在我的結果中可以發現最佳分類率也及終於中後段，也就是 sigma 一樣不能太大或太小導致過擬合及欠擬合發生。

3. 若分析過程不採用 two-fold cross validation，則分類率是否會更高？

Ans: 以下是我分別用 fold1 和 fold2 當 training data 的各自分類率

- i. Fold 1 當 training，CR1:

	1	5	10	50	100	500	1000
1.05^-100	62.67	62.67	62.67	62.67	62.67	62.67	62.67
1.05^-95	82.67	82.67	82.67	82.67	82.67	82.67	82.67
1.05^-90	62.67	62.67	62.67	62.67	62.67	62.67	62.67
1.05^-85	66.67	66.67	66.67	66.67	66.67	66.67	66.67
1.05^-80	66.67	66.67	66.67	66.67	66.67	66.67	66.67
1.05^-75	66.67	66.67	66.67	66.67	66.67	66.67	66.67
1.05^-70	46.67	46.67	46.67	46.67	46.67	46.67	46.67
1.05^-65	50.67	58.67	58.67	58.67	58.67	58.67	58.67
1.05^-60	60.00	60.00	60.00	60.00	60.00	60.00	60.00
1.05^-55	68.00	68.00	68.00	68.00	68.00	68.00	68.00
1.05^-50	70.67	72.00	72.00	72.00	72.00	72.00	72.00
1.05^-45	78.67	80.00	80.00	80.00	80.00	80.00	80.00
1.05^-40	88.00	89.33	89.33	89.33	89.33	89.33	89.33
1.05^-35	92.00	90.67	90.67	90.67	90.67	90.67	90.67
1.05^-30	94.67	93.33	93.33	93.33	93.33	93.33	93.33
1.05^-25	93.33	94.67	94.67	94.67	94.67	94.67	94.67
1.05^-20	94.67	94.67	94.67	94.67	94.67	94.67	94.67
1.05^-15	94.67	96.00	96.00	96.00	96.00	96.00	96.00
1.05^-10	93.33	94.67	96.00	96.00	96.00	96.00	96.00
1.05^-5	96.00	94.67	94.67	94.67	94.67	94.67	94.67
1.05^0	94.67	94.67	93.33	96.00	96.00	96.00	96.00
1.05^5	94.67	93.33	97.33	96.00	96.00	96.00	96.00
1.05^10	93.33	94.67	94.67	93.33	96.00	94.67	94.67
1.05^15	93.33	94.67	93.33	92.00	94.67	93.33	93.33
1.05^20	94.67	94.67	94.67	98.67	93.33	93.33	93.33
1.05^25	93.33	94.67	94.67	98.67	98.67	94.67	93.33
1.05^30	94.67	96.00	94.67	94.67	98.67	94.67	93.33
1.05^35	92.00	94.67	98.67	94.67	98.67	96.00	94.67
1.05^40	NaN	94.67	94.67	94.67	94.67	98.67	94.67
1.05^45	NaN	94.67	96.00	94.67	94.67	98.67	97.33
1.05^50	NaN	93.33	93.33	94.67	94.67	98.67	98.67
1.05^55	NaN	90.67	93.33	96.00	94.67	94.67	98.67
1.05^60	NaN	90.67	90.67	94.67	97.33	94.67	98.67
1.05^65	NaN	NaN	NaN	93.33	93.33	96.00	94.67
1.05^70	NaN	NaN	93.33	92.00	94.67	93.33	94.67
1.05^75	NaN	NaN	NaN	92.00	94.67	98.67	94.67
1.05^80	NaN	NaN	NaN	90.67	93.33	94.67	94.67
1.05^85	NaN	NaN	NaN	94.67	90.67	96.00	97.33
1.05^90	NaN	NaN	NaN	NaN	90.67	93.33	94.67
1.05^95	NaN	NaN	NaN	NaN	90.67	94.67	94.67
1.05^100	NaN	NaN	NaN	NaN	NaN	90.67	94.67
Optimal parameters and CR are :	[C = 50, sigma = 1.05^20 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 50, sigma = 1.05^25 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 100, sigma = 1.05^25 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 100, sigma = 1.05^30 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 10, sigma = 1.05^35 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 100, sigma = 1.05^35 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 500, sigma = 1.05^40 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 500, sigma = 1.05^45 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 500, sigma = 1.05^50 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 1000, sigma = 1.05^50 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 1000, sigma = 1.05^55 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 1000, sigma = 1.05^60 , best CR = 98.67]						
Optimal parameters and CR are :	[C = 500, sigma = 1.05^75 , best CR = 98.67]						

ii. Fold 2 當 training , CR2:

	1	5	10	50	100	500	1000
1.05^-100	45.33	45.33	45.33	45.33	45.33	45.33	45.33
1.05^-95	50.67	50.67	50.67	50.67	50.67	50.67	50.67
1.05^-90	36.00	36.00	36.00	36.00	36.00	36.00	36.00
1.05^-85	36.00	36.00	36.00	36.00	36.00	36.00	36.00
1.05^-80	36.00	36.00	36.00	36.00	36.00	36.00	36.00
1.05^-75	36.00	36.00	36.00	36.00	36.00	36.00	36.00
1.05^-70	36.00	36.00	36.00	36.00	36.00	36.00	36.00
1.05^-65	40.00	40.00	40.00	40.00	40.00	40.00	40.00
1.05^-60	48.00	52.00	52.00	52.00	52.00	52.00	52.00
1.05^-55	58.67	58.67	58.67	58.67	58.67	58.67	58.67
1.05^-50	60.00	60.00	60.00	60.00	60.00	60.00	60.00
1.05^-45	69.33	69.33	69.33	69.33	69.33	69.33	69.33
1.05^-40	81.33	81.33	81.33	81.33	81.33	81.33	81.33
1.05^-35	89.33	89.33	89.33	89.33	89.33	89.33	89.33
1.05^-30	93.33	92.00	92.00	92.00	92.00	92.00	92.00
1.05^-25	97.33	96.00	96.00	96.00	96.00	96.00	96.00
1.05^-20	96.00	94.67	96.00	96.00	96.00	96.00	96.00
1.05^-15	96.00	94.67	94.67	97.33	97.33	97.33	97.33
1.05^-10	96.00	96.00	94.67	96.00	97.33	97.33	97.33
1.05^-5	96.00	96.00	96.00	94.67	94.67	96.00	96.00
1.05^0	96.00	96.00	96.00	94.67	94.67	96.00	96.00
1.05^5	97.33	97.33	96.00	96.00	97.33	94.67	96.00
1.05^10	97.33	97.33	97.33	96.00	96.00	96.00	94.67
1.05^15	97.33	97.33	97.33	96.00	96.00	97.33	96.00
1.05^20	96.00	97.33	97.33	96.00	96.00	97.33	97.33
1.05^25	96.00	96.00	97.33	97.33	96.00	96.00	97.33
1.05^30	94.67	97.33	96.00	97.33	97.33	97.33	96.00
1.05^35	94.67	96.00	97.33	98.67	98.67	97.33	97.33
1.05^40	NaN	96.00	97.33	98.67	98.67	97.33	97.33
1.05^45	NaN	94.67	94.67	97.33	98.67	98.67	97.33
1.05^50	NaN	93.33	96.00	96.00	100.00	98.67	97.33
1.05^55	NaN	90.67	93.33	97.33	97.33	98.67	98.67
1.05^60	NaN	90.67	94.67	96.00	97.33	98.67	98.67
1.05^65	NaN	NaN	NaN	96.00	94.67	98.67	98.67
1.05^70	NaN	NaN	92.00	94.67	94.67	97.33	98.67
1.05^75	NaN	NaN	NaN	93.33	96.00	98.67	98.67
1.05^80	NaN	NaN	NaN	93.33	93.33	98.67	96.00
1.05^85	NaN	NaN	NaN	93.33	90.67	94.67	97.33
1.05^90	NaN	NaN	NaN	NaN	93.33	96.00	96.00
1.05^95	NaN	NaN	NaN	NaN	89.33	94.67	96.00
1.05^100	NaN	NaN	NaN	NaN	NaN	93.33	94.67
Optimal parameters and CR are : [C = 100, sigma = 1.05^50 , best CR = 100.00]							

由各自的分類率可看出，整體來說 fold 1 當 training data 的分類率 CR1 比 fold 2 當 training data 的分類率 CR2 來的高，但 CR2 最高分類率有到 100%，所以不做 two-fold cross validation，結果取決於擬採用哪個 fold 當 training，像用 fold 2 當 training data 可能會有最高的分類率，但難以代表模型好壞，因為 data 使用上太單一沒有跟其他做比較，只是剛好此 fold 能得到好分類率，所以才需要做多折驗證來證明在各種組合下模型都有一定的分類水準，不然就只是剛好結果好看而已了。