

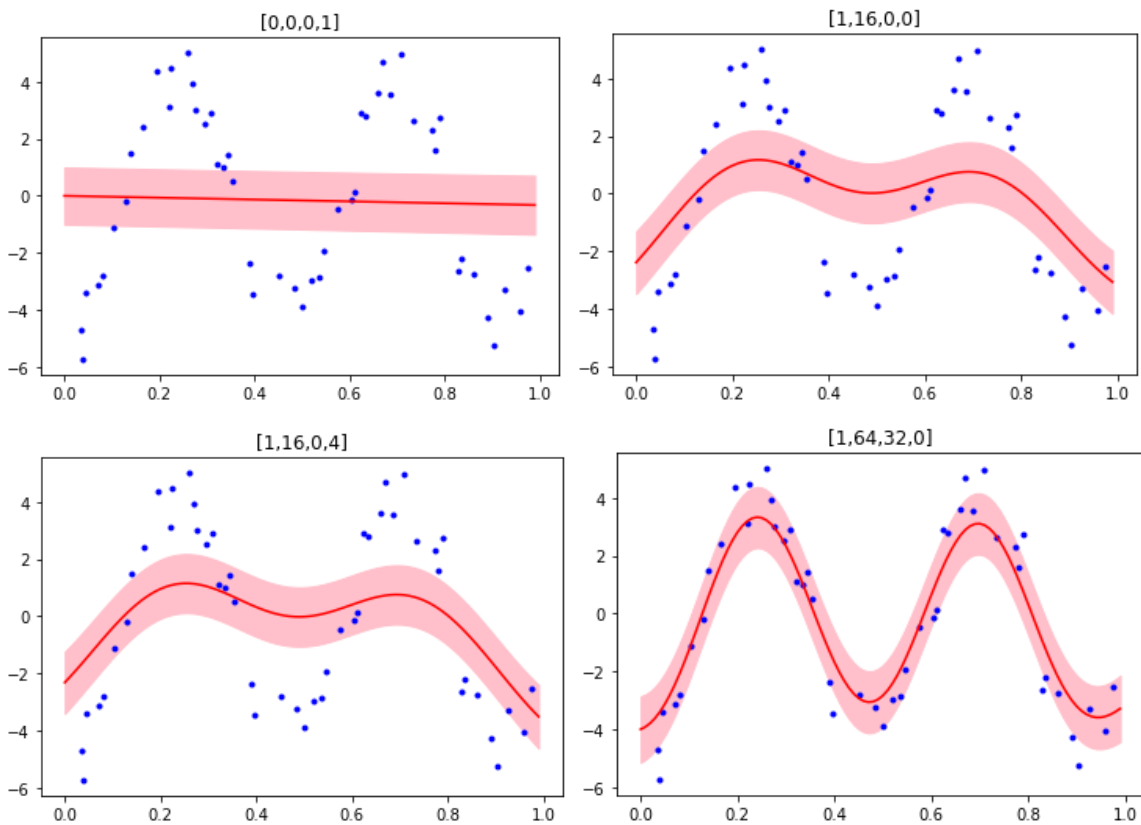
HW3

309555025 羅文笙

補交 **Note:** 更新了第二題全部/第三題補上 log likelihood curve!

1. Gaussian Process for Regression

Predicion Result

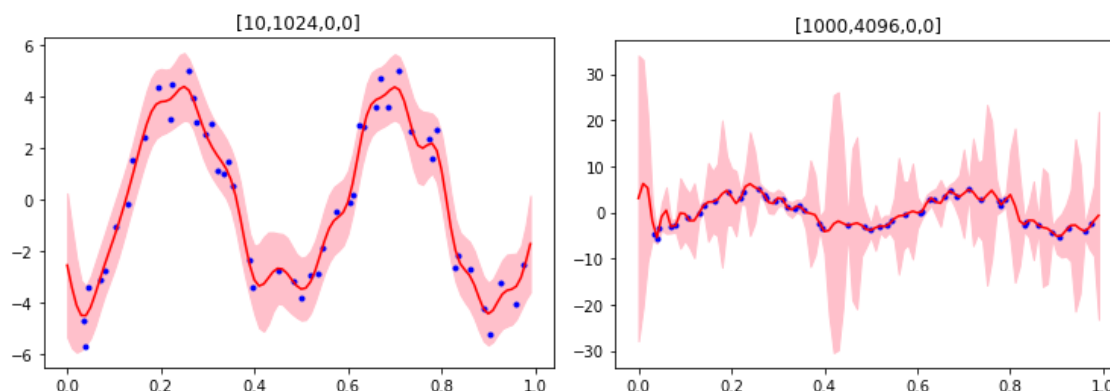


Show RMS

	RMS of training	RMS of test
$[0,0,0,1]$	3.1292	3.3201
$[1,16,0,0]$	2.4239	2.4656
$[1,16,0,4]$	2.4105	2.4557
$[1,64,32,0]$	1.0428	1.0925

Discussion

最後我選擇 $\theta = [10, 1024, 0, 0]$ 當作最好的超參數。 θ_2 和 θ_3 似乎不太重要，就沒有去嘗試了，我發現 θ_0 因為是乘上後面一大串 \exp ，所以一定要不等於 0，且一定要 >0 ，不然後面求變異數的時候會產生負數，那 θ_0 和 θ_1 越大的話，training set 和 test set 的 RMS 就會越小，甚至可以小到 0.05，但是相對的，標準差就會變得很大，所以我選擇這個組合，取得 RMS 和標準差之間的平衡。



	RMS of training	RMS of test	平均標準差
[10,1024,0,0]	0.5156	0.6415	1.3189
[1000,4096,0,0]	0.0564	0.1704	7.1930
[1,64,32,0]	1.0428	1.0925	1.0621

2. Support Vector Machine (SVM)

2-1 Analyze & Decide which approaches to use

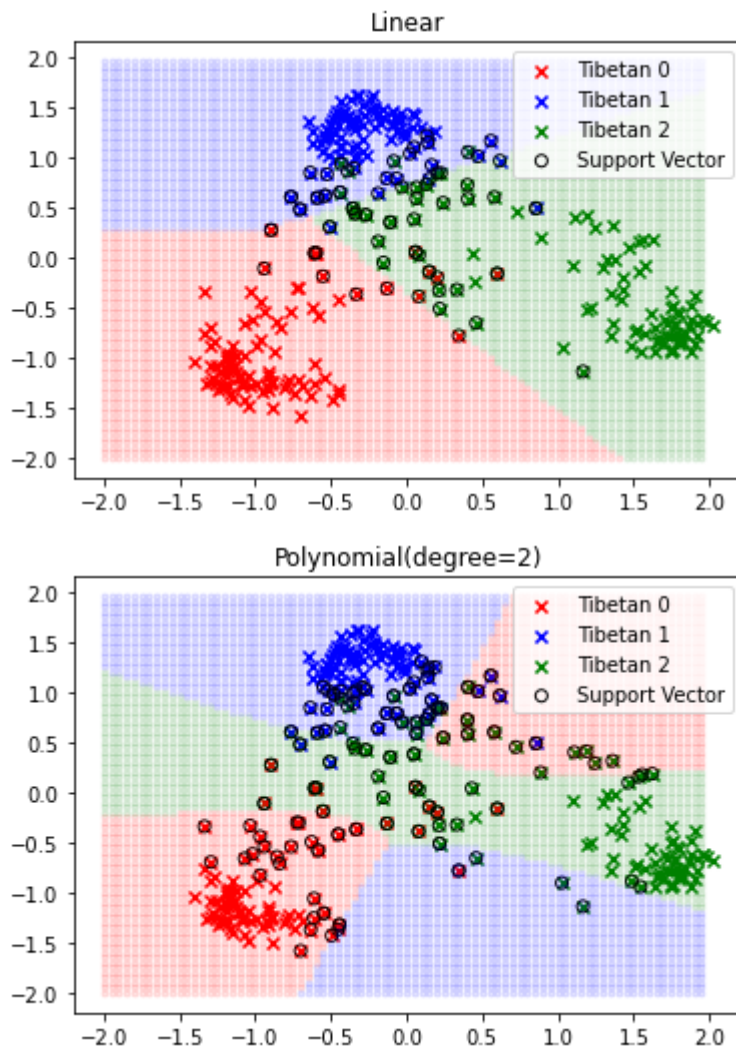
One-VS-Rest，主要會有兩個問題，第一個就是每次的分類都是在訓練不同的 classifier，所以不能夠保證不同 classifier 的 $y_k(x)$ 值會在相同的數量級上，在使用最後的 decision value 來決定最後的分類結果時，可能會有誤差。第二個問題是，如果本來每個 label 的 training data 都佔了相同的數量，但用 One-VS-Rest 的時候，在每次訓練 classifier 時會變成 One 只佔了一點點，Rest 有很大一塊，產生 training sets 的 imbalance。

One-VS-One 的話就比較沒有 One-VS-Rest 的這些問題，但是假設有 N 個類別的話，就必須要訓練出 $C(N,2)$ 個 classifier，整體的時間上會需要比 One-VS-Rest 還要久。

最後，考量到 training data set 每一個 label 的數量都差不多，為了避免用 One-VS-Rest 產生 imbalance 的問題，且只有 3 個 label，One-

VS-One 和 One-VS-Rest 都需要訓練出 3 個 classifier，不會需要比較久的時間，最後我選擇使用 One-VS-One 的方法。

2-2/2-3 Plot the corresponding decision boundary and support vector.



2-4 Please discuss the difference between (2), (3)

Kernel function 主要的功用就是希望當不同類別的資料在原始空間中無法被線性 classifier 區隔開來時，利用非線性的投影使得資料能在更高維度的空間中可以更區隔開。

所以當 linear kernel 無法有效區隔資料點的時候就可以考慮使用 polynomial kernel 來將資料投影到更高維度來試試看。但是在本次作業中，其實 linear 的 kernel 就已經分得不錯了，反而是使用 polynomial kernel 之後，還分類的比較不好。因此 kernel function 的選擇，主要還是要看 input data 和實際結果之後來做決定會比較好。

3. Gaussian Mixture Model

3-1 Table of μ_k

*** Table of k = 2 ***

R	G	B

60	75	82
143	161	164

*** Table of k = 3 ***

R	G	B

167	183	183
91	111	119
47	59	67

*** Table of k = 5 ***

R	G	B

50	79	88
113	135	142
185	197	195
78	91	98
40	51	58

*** Table of k = 7 ***

R	G	B

45	57	65
203	213	210
97	118	126
140	160	162
22	35	41
71	76	78
63	87	98

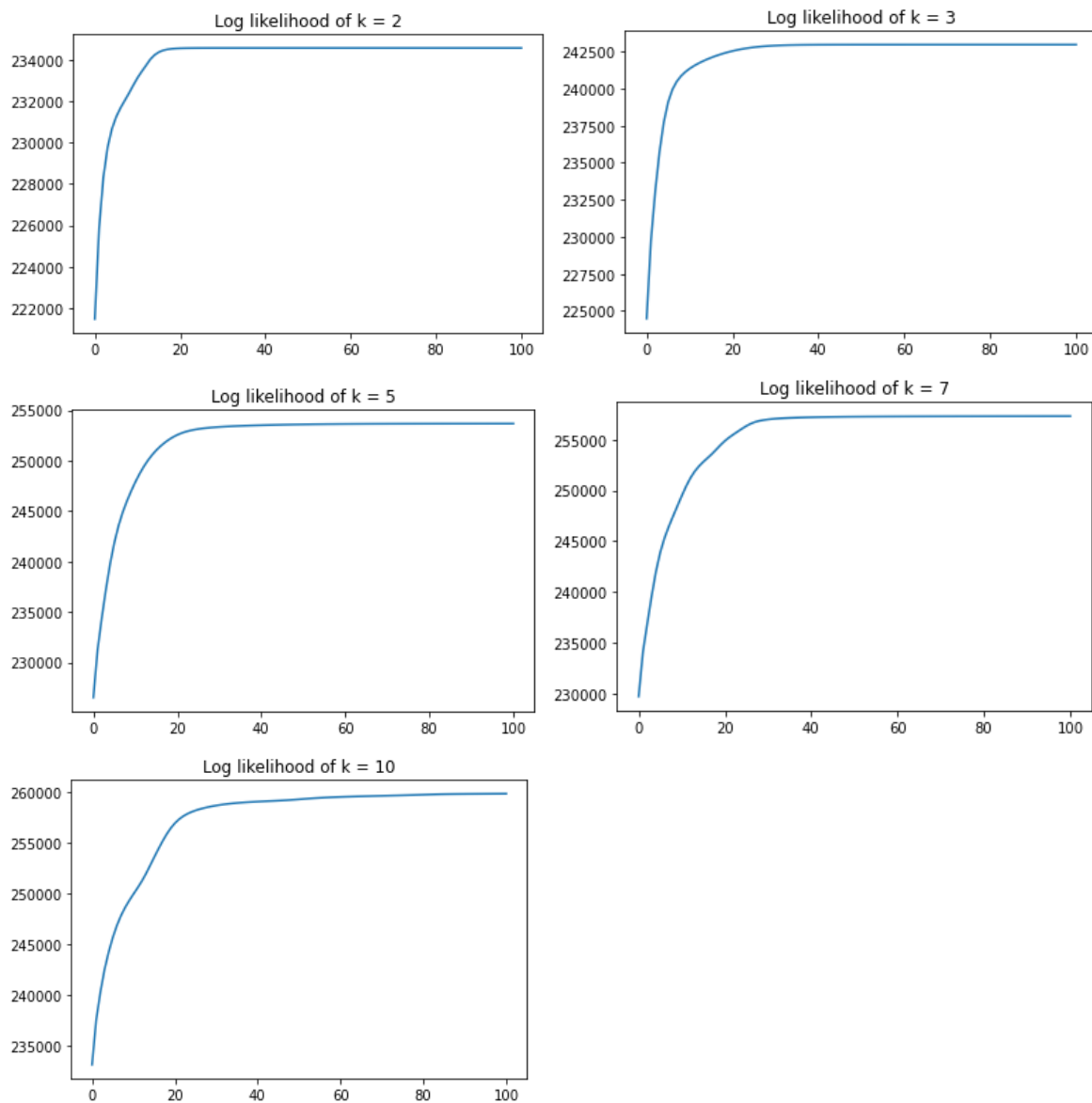
*** Table of k = 10 ***

R	G	B

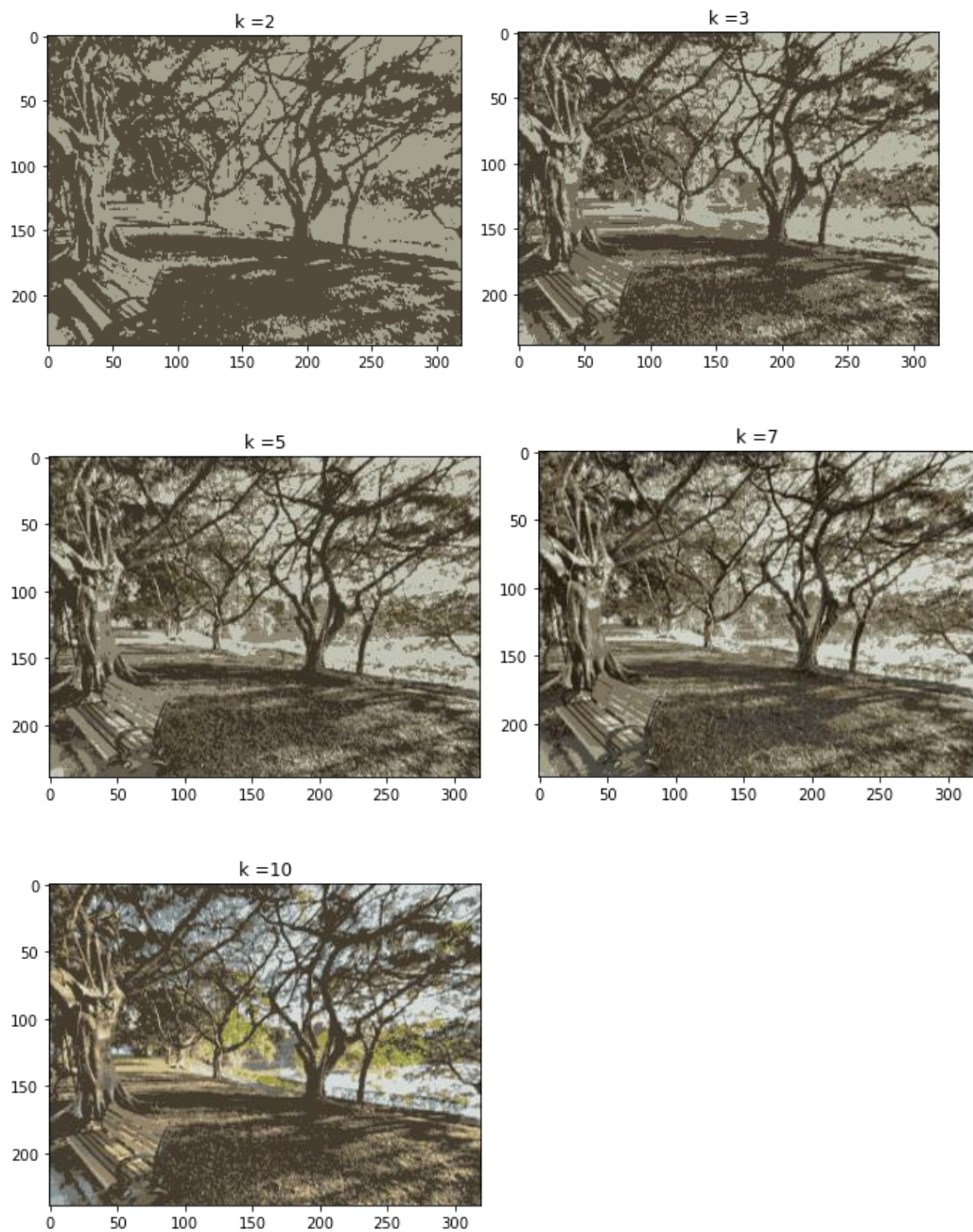
119	121	115
79	99	109

121	154	175
47	60	67
174	176	168
140	194	208
218	221	216
108	132	147
73	159	156
153	147	136

3-2/ 3-4 Log likelihood curve



3-3 Image



3-5 Discussion

隨著 k 越來越大，圖片也越來越清晰， k 在這邊代表的意思可以想成是有幾種顏色的畫筆，當 $k=2$ 時，圖片中的每個 **pixel** 只有灰色或黑色的選擇， $k=10$ 的時候，就變得非常接近原來的圖片了。