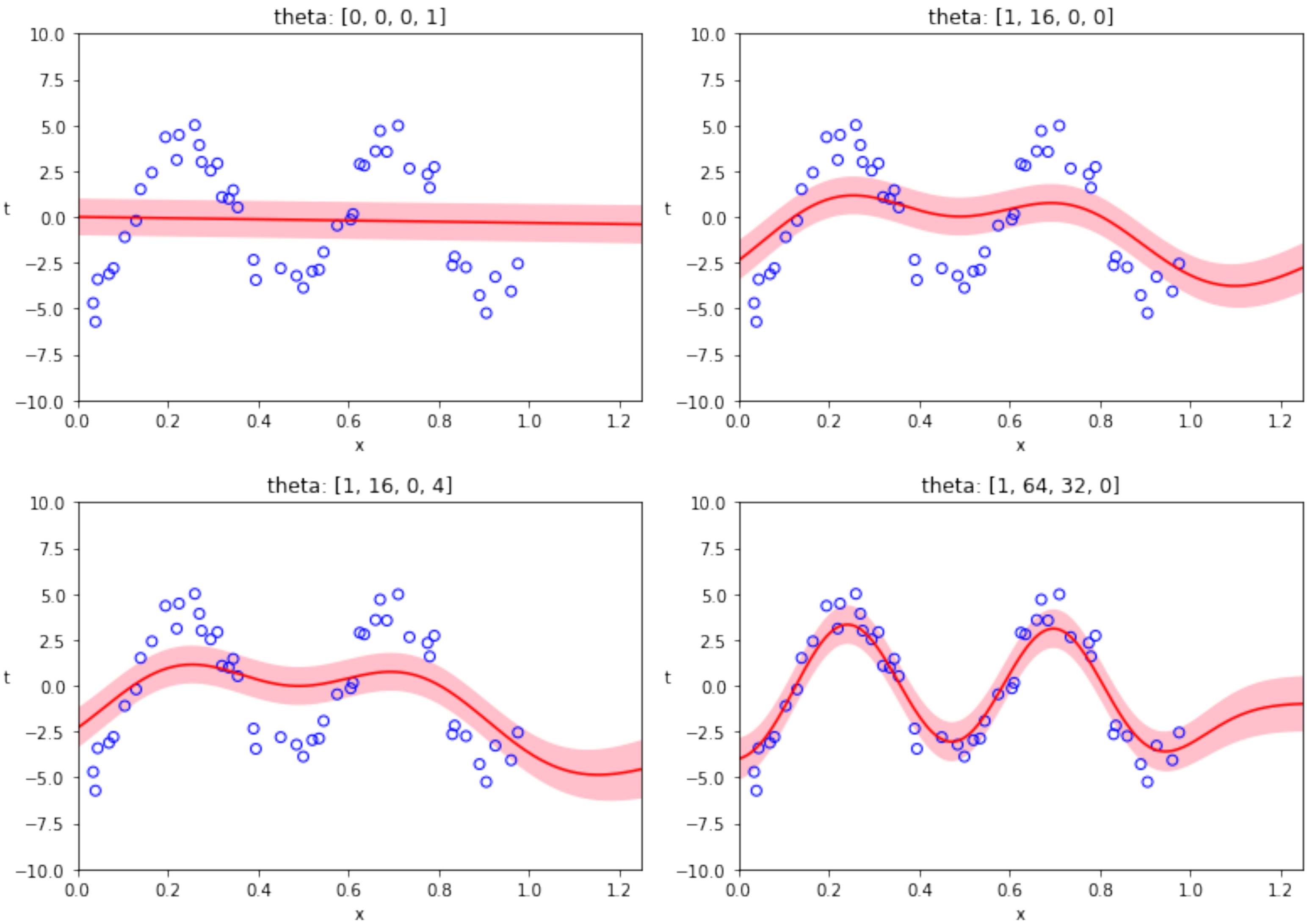
Machine Learning HW3 Report

309505018 郭俊廷

1. Gaussian Process for Regression (40%)

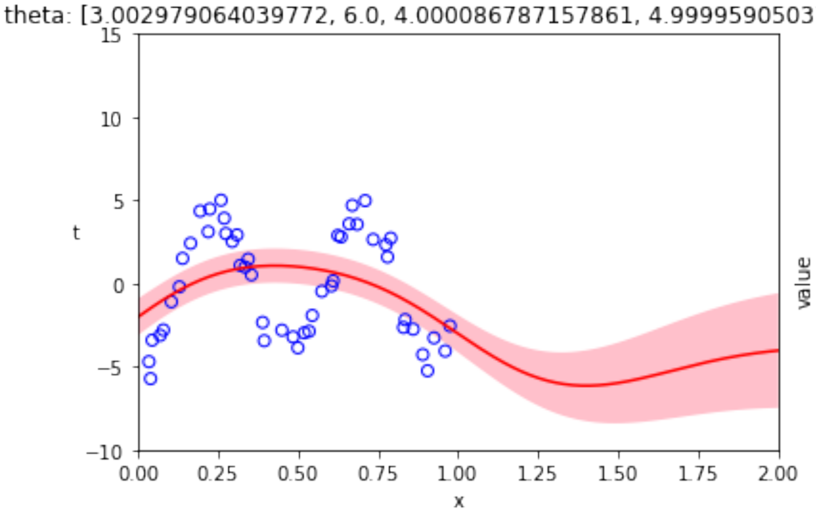
(1). Plot the prediction result for training set but one standard deviation

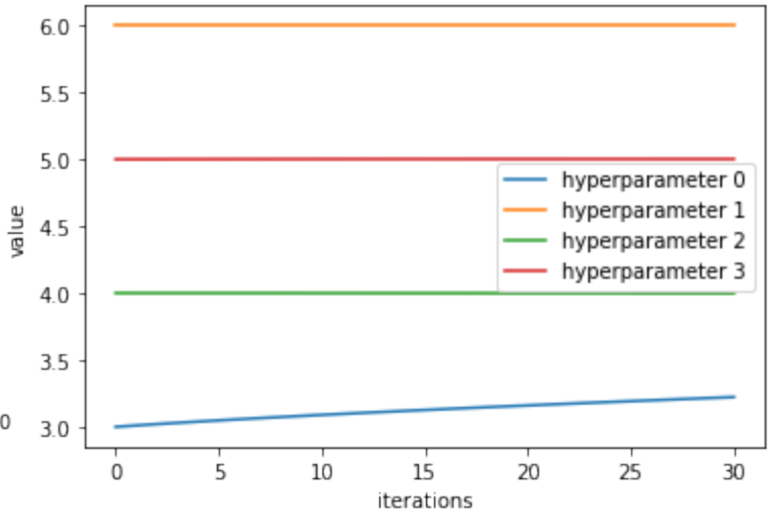


(2). Show the corresponding root-mean-square errors

param	training error	testing error
(0, 0, 0, 1)	3.129201	3.344399
(1, 16, 0, 0)	2.423928	2.668052
(1, 16, 0, 4)	2.410576	2.656998
(1, 64, 32, 0)	1.042886	1.162759

(3). Tuned the hyperparameters used Automatic Relevance Determination (ARD)





param	training error	testing error
(3.002979064039772, 6.0, 4.000086787157861, 4.9999590503778215)	2.804456	3.173625

(4). Explain your findings and do some discussion.

這次是實作不同theta對於 Exponential-Quadratic Kernel function的影響，在theta=[0,0,0,1]的錯誤比其他的都高很多，因為他前三個參數都是零，等於只用了Linear Kernel，所以相對變化少，在 training data和 testing data的表現error都偏高，屬於underfitting.

在theta=[1,16,0,0] 和theta=[1,16,0,4]時，以得出來的結果來看，兩者不管在 training data或testing data的表現error是差不多的，畢竟前三個參數一樣，兩者表現都比第一種 Linear Kernel 的好。

在theta=[1,64,32,0] 時error是最小的，因為他的function是最複雜的，從decision region圖來看，四種只有這一種是幾乎跟training data 分佈最接近的。

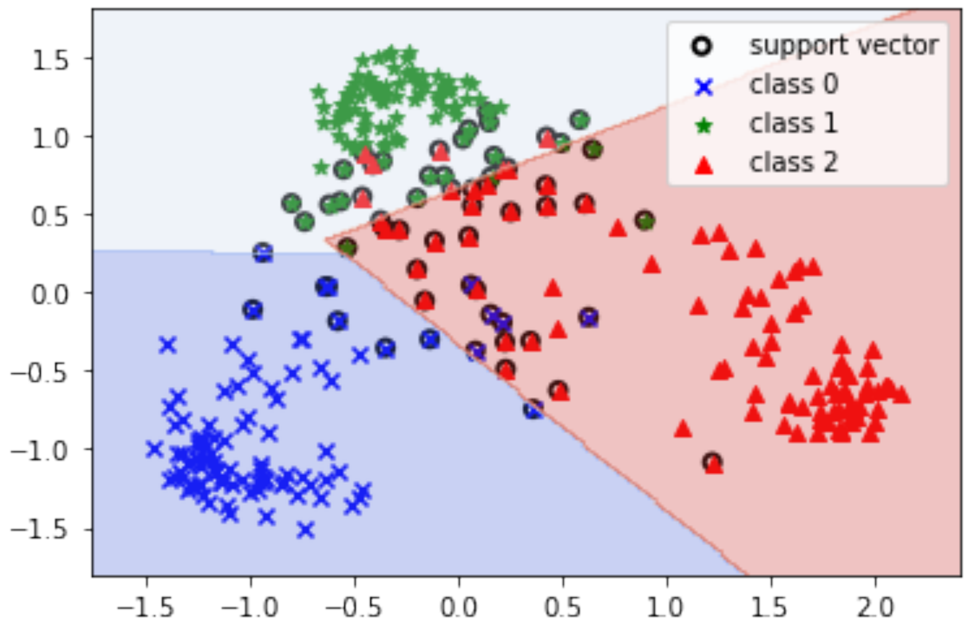
2. Support Vector Machine (SVM) (40%)

(1). Analyze the difference between one-versus-the-rest and one-versus-one approaches

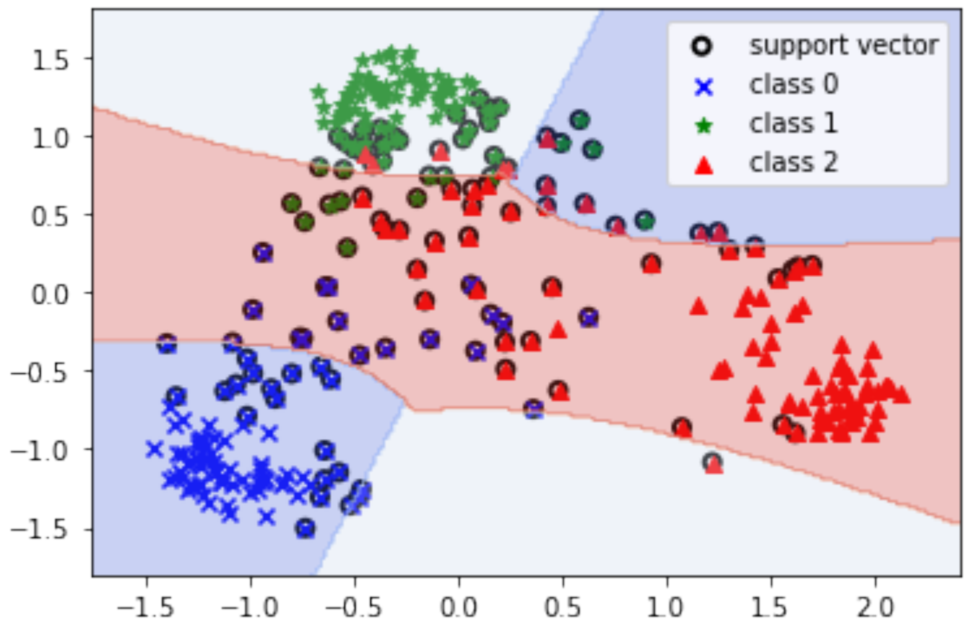
兩者差別在於
One v.s The rest 是某個類別的樣子歸為一類,其他剩餘的樣本歸為另一類，這樣k個類別的樣本就構造出了k個SVM
One v.s One 任意兩個樣本之間設計一個SVM，因此k個類別的樣本就需要設計k(k-1)/2個SVM。當對一個未知樣本進行分類時，最後得票最多的類別即為該為之樣本的類別。
Libsvm中的很多分類都是用One v.s One實現的。

以使用Sklearn API來說，如果要使用One v.s The rest 則用ovr，One v.s One則用ovo
One v.s The rest : SVC(kernel='linear', C=1, decision_function_shape='ovr')
One v.s One : SVC(kernel='linear', C=1, decision_function_shape='ovo')

(2). Plot the decision boundary and support vector (SVM, Linear kernel)



(3). Plot the decision boundary and support vector (SVM, Polynomial kernel degree=2)



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

我的作法是，先把資料做PCA降維降到2維，再對資料做標準化，SVM採用One v.s One做multiclass 的分類，利用sklearn裡的 fit() 取得coefficient，算出每個分類的 weight和 bias，最後再投票分類結果。

Linear kernel SVM 的class大致上分為三群，從圖片上來看class 1和class 2之間的分類效果沒有非常好，可以看到少數的class 1跑到class 2，少數的class 2跑到class 1，不過整體來說，大致上結果還不錯。

Polynomial kernel SVM是採用degree = 2, 從圖片上來看overlap的結果比較嚴重一點點，中間有一整塊是同時有class1 class2 class3的，不過大致上還行。

3. Gaussian Mixture Model (30%)

(1). K-means model show the table of estimated $\{\mu_k\}_{k=1}$

======= K = 3 (K_means) =======

K_means	R	G	B
0	178	185	169
1	71	54	47
2	154	133	104

======= K = 5 (K_means) =======

K_means	R	G	B
0	93	82	76
1	68	54	43
2	132	132	80
3	128	125	90
4	111	93	89

======= K = 7 (K_means) =======

K_means	R	G	B
0	140	129	97
1	176	179	100
2	148	126	105
3	78	67	63
4	54	35	29
5	97	99	50
6	216	213	178

======= K = 10 (K_means) =======

K_means	R	G	B
0	157	158	163
1	151	120	117
2	184	196	186
3	71	59	61
4	152	162	127
5	210	206	205
6	213	207	147
7	151	123	120
8	252	248	247
9	185	161	133

(2). K-means model show the resulting images for K = 3, 5, 7, and 10.

K = 3

K_means

K = 5

K_means

K_means

K = 7.

K_means

K = 10

(3). Gaussian Mixture Model show the table of estimated $\{\mu_k\}_{k=1}$

======= K = 3 (GMM) =======

GMM	R	G	B
0	77	65	56
1	158	163	130
2	121	115	100

======= K = 5 (GMM) =======

GMM	R	G	B
0	120	124	73
1	149	156	156
2	75	62	55
3	152	140	114
4	97	85	65

======= K = 7 (GMM) =======

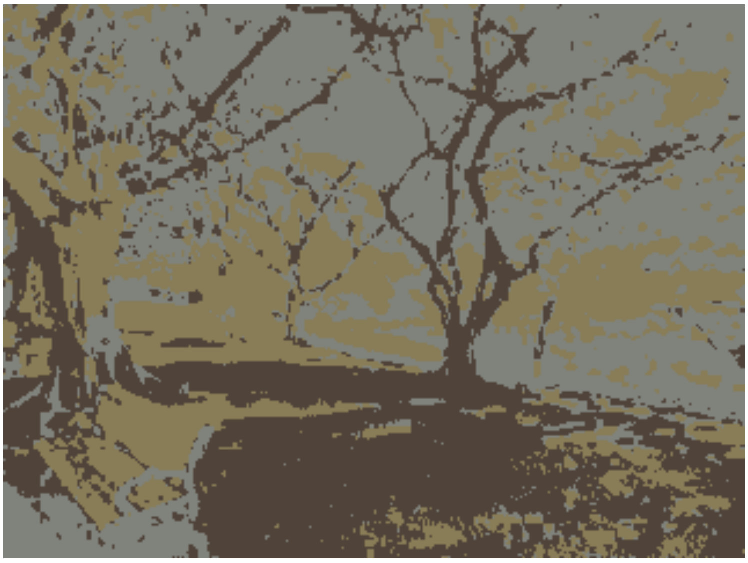
GMM	R	G	B
0	176	182	184
1	75	69	41
2	129	115	95
3	78	64	57
4	81	79	71
5	165	171	161
6	138	142	86

======= K = 10 (GMM) =======

GMM	R	G	B
0	135	117	89
1	138	144	147
2	135	139	88
3	159	168	165
4	73	61	54
5	134	122	105
6	64	58	34
7	92	84	70
8	94	78	70
9	223	223	222

(4). Gaussian Mixture Model Show the resulting images for K = 3, 5, 7, and 10.

K = 3

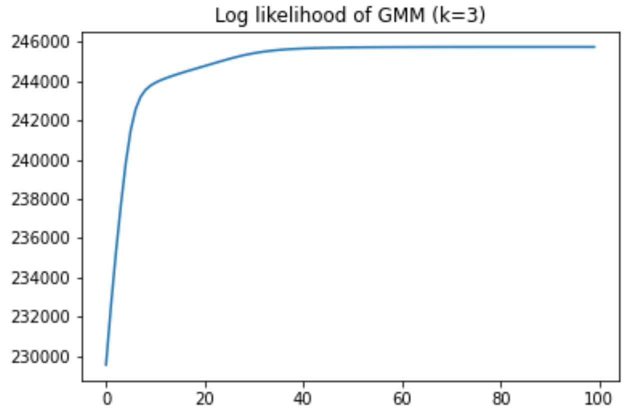


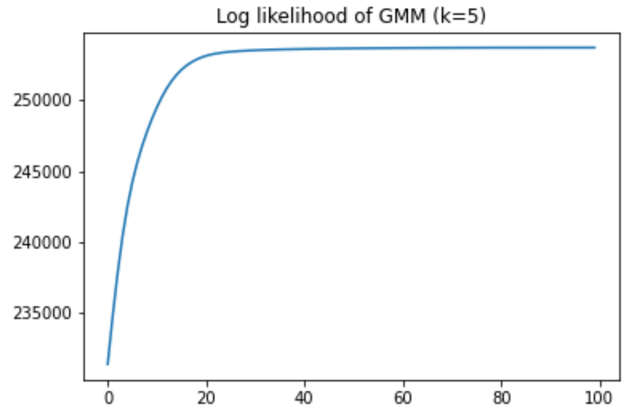
K = 5

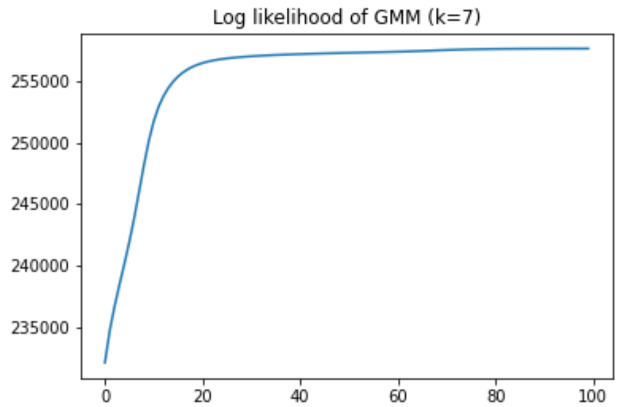
K = 7

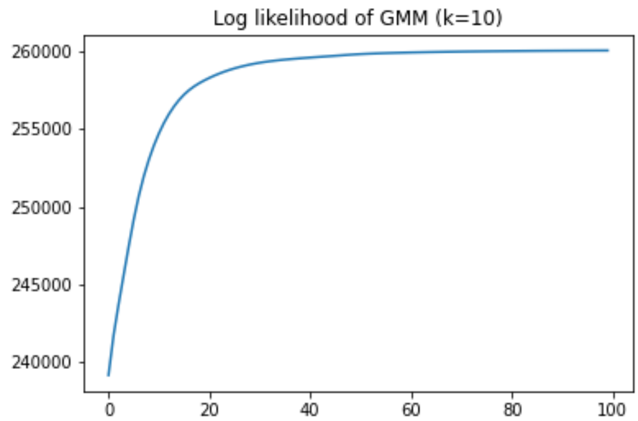
K = 10

(3). Show the graph of Log likelihood at different iterations for K = 3, 5, 7, 10 respectively. Example are shown below.









(4). You can do some discussions about what is crucial factor to influence the output image and explain the reason?

從上面的圖片可以看出，K_Means所產生的圖片有些地方比較黑一點，尤其是 K=3 幾乎黑一大塊，而且顏色比較單調，沒那麼好看，GMM的結果雖然偏墨綠跟咖啡色，但在一些樹的輪廓上表現得比K_Means好，我覺得原因是GMM把K_Means的結果當作事前機率去計算，有了事前機率的基礎，所以GMM在一些樹的細節輪廓，及風景的對比度上表現的比K_means好。