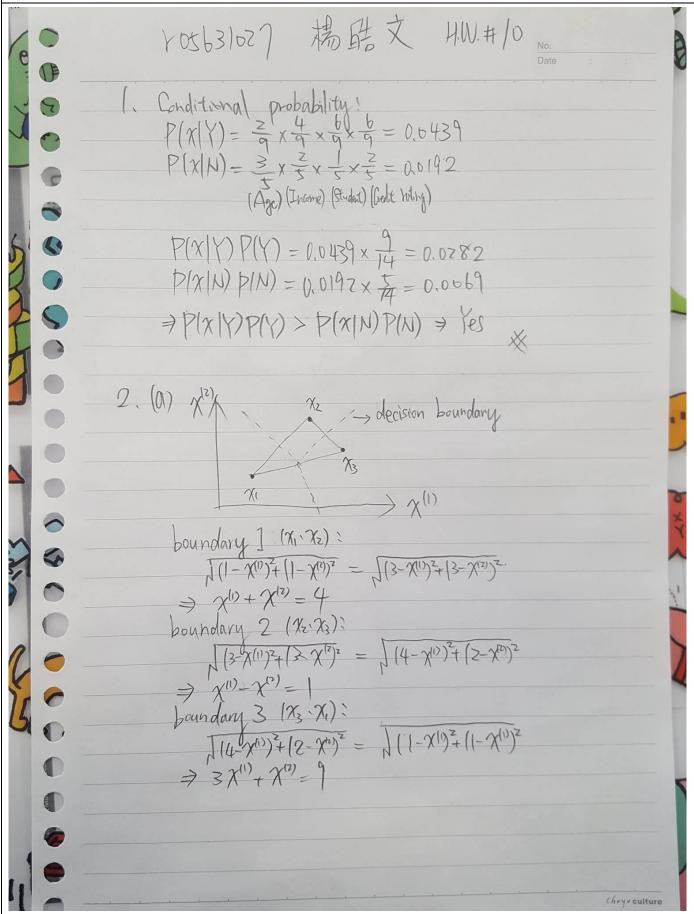
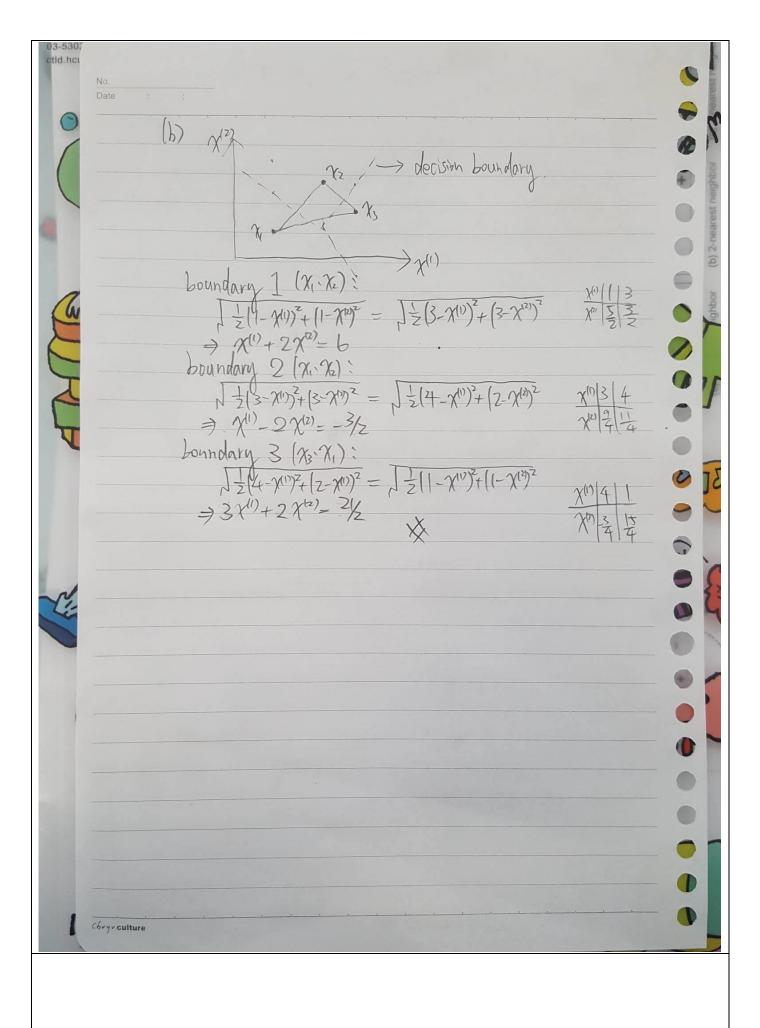
Problem 1 (Naïve Bayes)

Problem 2 (k-nearest neighbor)





Problem 3 (k-nearest neighbor)

(a) For k = 1, 3, 5, 11, 16, and 21, build kNN classifiers from the training data.

Training error:

1	2	3	4	5	6	
0	6.7000	8.4000	11.6000	13.7000	15.4000	

Validation error:

1	2	3	4	5	6
12.6667	14.3333	13	17.3333	19.6667	20.3333

由 validation error 來選擇 model 的話,可以觀察到當 k = 1 時,有最小的 error = 12.6667,因此選擇 k = 1 作為最佳 model。

Test error:

1	2	3	4	5	6	
10.6667	9.3333	8.3333	12.6667	14.0000	15.3333	

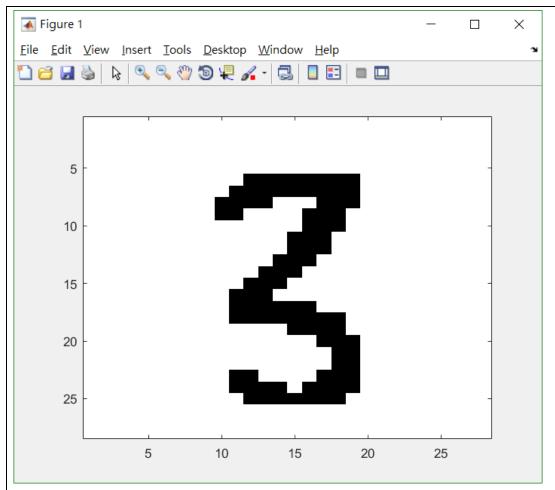
(b)Compute the confusion matrix of the classifier based on the data in "10HW3_test.txt".

	1	2	3	4	5	6	7	8	9	10
1	0.8929	0	0	0	0	0	0	0	0	0
2	0	1	0.0526	0.1333	0	0	0.0541	0	0	0
3	0	0	0.9211	0.0333	0	0	0	0	0	0.0370
4	0	0	0	0.7667	0	0.0769	0	0	0.0357	0
5	0	0	0	0	0.9286	0	0	0	0	0.0370
6	0	0	0	0	0	0.8462	0.0270	0	0	0
7	0.1071	0	0	0	0	0	0.9189	0	0.0357	0
8	0	0	0	0	0	0.0385	0	0.9333	0	0
9	0	0	0.0263	0.0333	0	0	0	0	0.9286	0
10	0	0	0	0.0333	0.0714	0.0385	0	0.0667	0	0.9259

由上面產生出的 confusion matrix 可以觀察到,第二行,也就是代表數字 $\mathbf{1}$,的準確率最高,可能因為筆劃簡單,不易和其他數字混淆。

另一方面,第四行,也就是代表數字3,的準確率最低,容易和1、2、8、9 這幾個數字搞混。

(c) Identify one falsely classified vector.



這個數字 3 被誤分類為數字 8,因為在幾何圖形上兩者的確很像(如下圖,為真正的數字 8)

