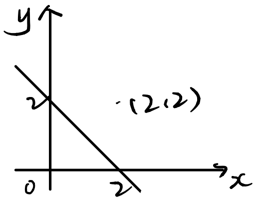
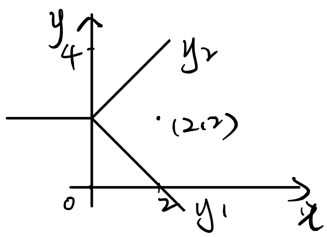
1\. a)



$(1,1)$ $\quad y = -x + b \quad b = 2$

$y = -x + 2$

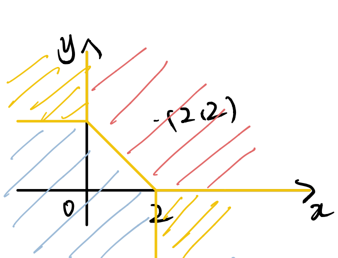
b)



$y_1 = -x + 2$

$y_2 = x + 2$

c)



$|x| + |y|$

$|x-2| + |y-2|$

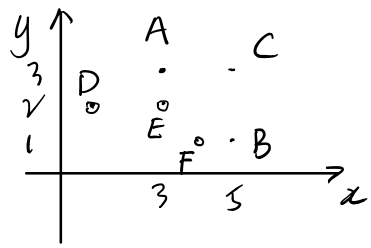
	$x \le 0$	$x \in [0,2)$	$x \ge 2$
$y \le 0$	$-x-y$ $-x+2-y+2$	$x-y$ $\quad x=2$ $-x+2-y+2$	$x-y$ $x-2-y+2$
$y \in [0,2)$	$-x+y$ $-x+2-y+2$ $y=2$	$x+y$ $\quad y=2-x$ $-x+2-y+2$	$x+y$ $\quad y=0$ $x-2-y+2$
$y \ge 2$	$-x+y$ $-x+2+y-2$	$x+y$ $\quad x=0$ $-x+2+y-2$	$x+y$ $x-2+y-2$

for $x \ge 2$, $y \le 0$, all points are equal distance from X_A, X_B.
Same for when $x \le 0$, $y \ge 2$.

2. $k=1$. $E(3,2)$ $A(3,3)$

$B(5,1)$ $F(4,1)$

$CCR = 1 - \frac{2}{3} = \frac{1}{3}$



$k=3$. A B F

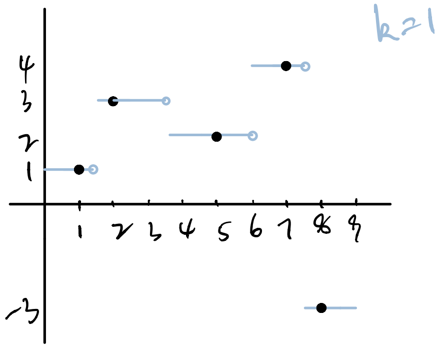
$CCR = 1 - \frac{1}{2} = \frac{1}{2}$

$k=5$ A. B. C. D. E. F

$CCR = 0$

$k=3$ best

3.

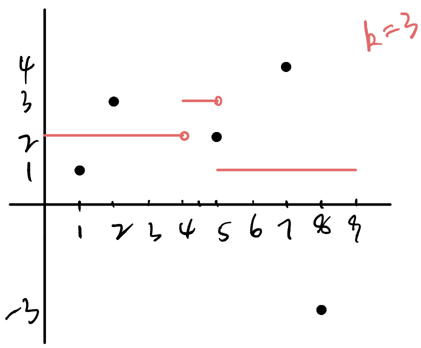


$x_1 = \frac{1+2}{2} = 1.5$

$x_2 = \frac{2+5}{2} = 3.5$

$x_3 = \frac{5+7}{2} = 6$

$x_4 = \frac{7+8}{2} = 7.5$



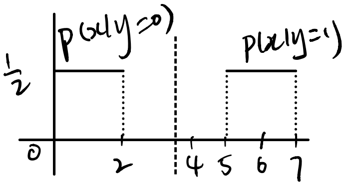
$x_1 = \frac{1+7}{2} = 4$ $\qquad \frac{1+3+2}{3} = 2.$

$x_2 = \frac{2+8}{2} = 5$ $\qquad \frac{3+2+4}{3} = 3.$

$\frac{2+4-3}{3} = 1$

$x = 4.5. \quad \hat{y} = \frac{2+3+4}{3} = 3$

4. a)



for Xtest to fall in one region

for n train samples to fall in another region

$k=1 \quad P(h_{1NN}(X_{test};D) \neq y_{test}) = \frac{1}{2}\left(\frac{1}{2}\right)^n + \frac{1}{2}\left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n$

$k=3. \quad P(h_{3NN}(X_{test};D) \neq y_{test}) = \left(\frac{1}{2}\right)^n + C_n^1\left(\frac{1}{2}\right)^n$

probability of no training point falls in the same region as Xtest add the probability of 1 training point falls in the same region as Xtest.

$k=5. \quad P(h_{5NN}(X_{test};D) \neq y_{test}) = \left(\frac{1}{2}\right)^n + C_n^1\left(\frac{1}{2}\right)^n + C_n^2\left(\frac{1}{2}\right)^n$

same as for k=3. P(no training point in region) + P(1 point...) + P(2 points...)

⋮

$k: \quad P(h_{kNN}(X_{test};D) \neq y_{test}) = \left(\frac{1}{2}\right)^n \sum_{k=1}^{k} C_n^{\frac{k-1}{2}}$

$\frac{k-1}{2}$ is the largest number of points that can fall into the same region as Xtest.

b) best $k \to \infty$

⋮

$k=5$

$k=3$

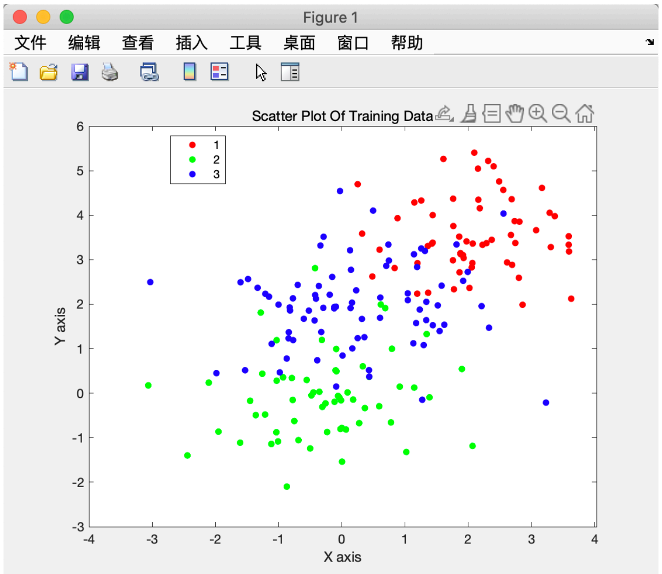
worst $k=1$

since the probability of misclassification gets smaller as k increases. So the CCR gets larger as k gets larger.

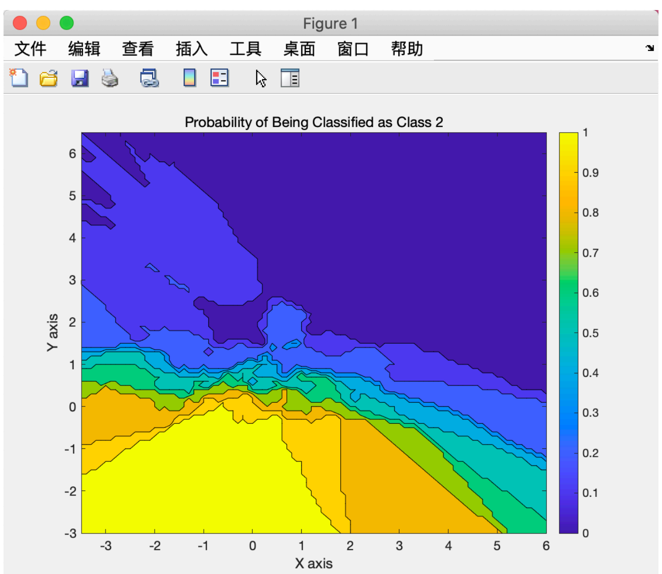
c) for all k, as $n \to \infty$, the misclassification rate would be 0 since the limit as $n \to \infty$ of functions for each k is 0.

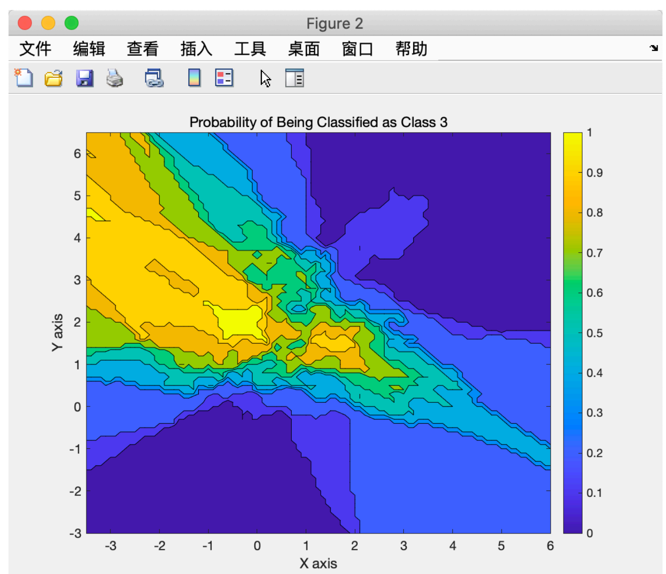
2.5

a) The scatter plot of all the training data:



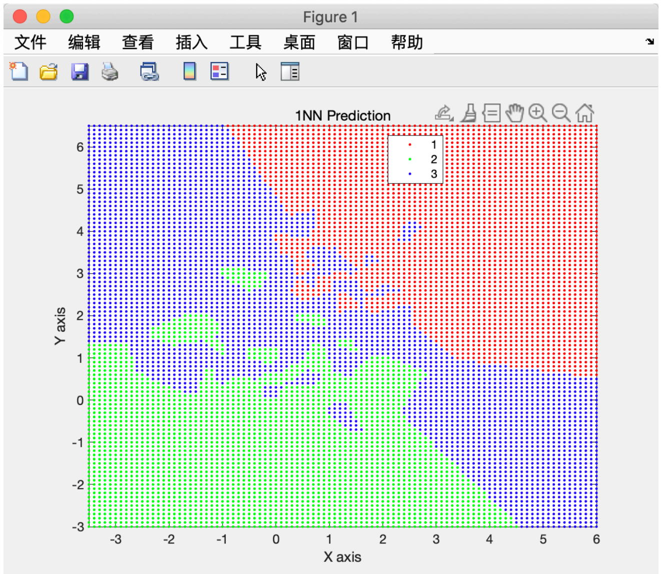
b) Two figures:

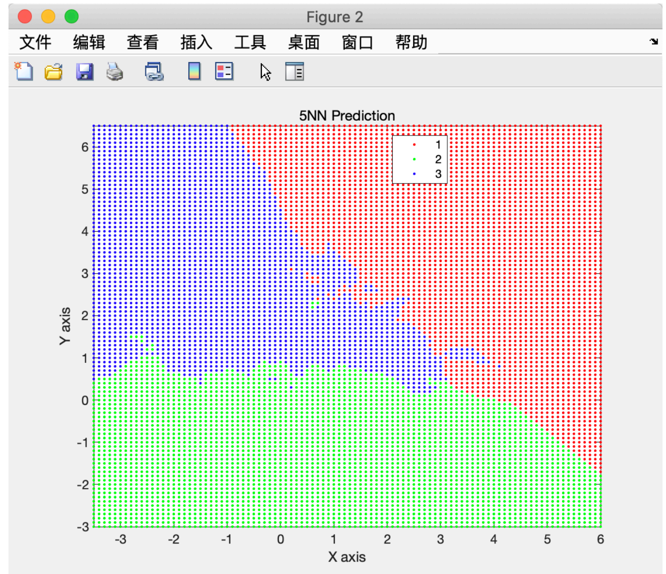




For point to be classified as class 2, the probability gets larger as the point approaches low left corner, this coincides with the scatterplot we got from part a. Same as for point to be classified as class 3, when the point goes near the coordinates of points labeled as class 3 in scatterplot, the probability gets larger. We also know from class that the probability should range from 0 and 1, and the figures resemble the heat maps.

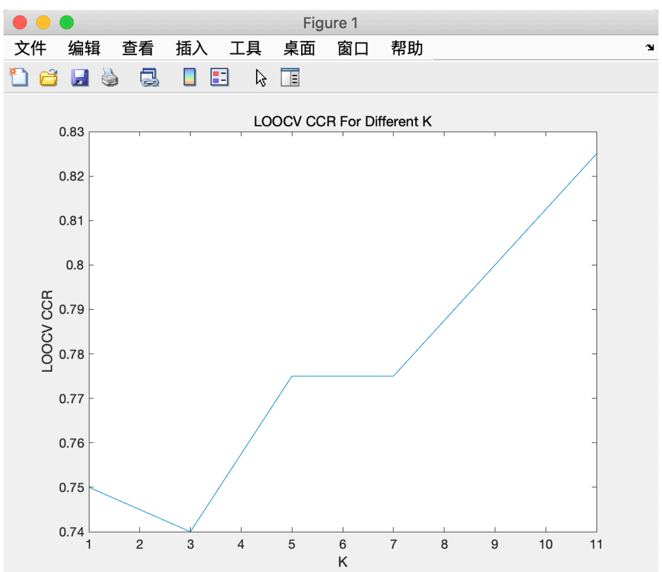
c) Two figures:





We can see for when K = 5, the boundary is smoother and the predict value of some points changed since we are considering more nearest training points. This is the same as we learn from class, as K increases, the boundary will be smoother.

d)



From the picture, we can see when k = 11 the LOOCV CCR is maximize. So we should choose k = 11. From class we see when K gets really large, the LOOCV CCR actually decreases. But for this case, since our training set comparatively larger than k, so the larger the k, the larger CCR is.

e) The confusion matrix:

conf_mat

10x10 double

	1	2	3	4	5	6	7	8	9	10	11
1	973	1	1	0	0	1	3	1	0	0	
2	0	1129	3	0	1	1	1	0	0	0	
3	7	6	992	5	1	0	2	16	3	0	
4	0	1	2	970	1	19	0	7	7	3	
5	0	7	0	0	944	0	3	5	1	22	
6	1	1	0	12	2	860	5	1	6	4	
7	4	2	0	0	3	5	944	0	0	0	
8	0	14	6	2	4	0	0	992	0	10	
9	6	1	3	14	5	13	3	4	920	5	
10	2	5	1	6	10	5	1	11	1	967	
11											

CCR = 0.9691

We can see the CCR rate is already very good with K=1, probably because our training sample is very large.