Assignment5
Yuqi Zhou
A20423555

Question #1:

a) (10 points). List the parameter estimates (round to four decimal places) in a table. The rows are the Intercept, the predictor x, and the predictor y. The columns are the ring labels.

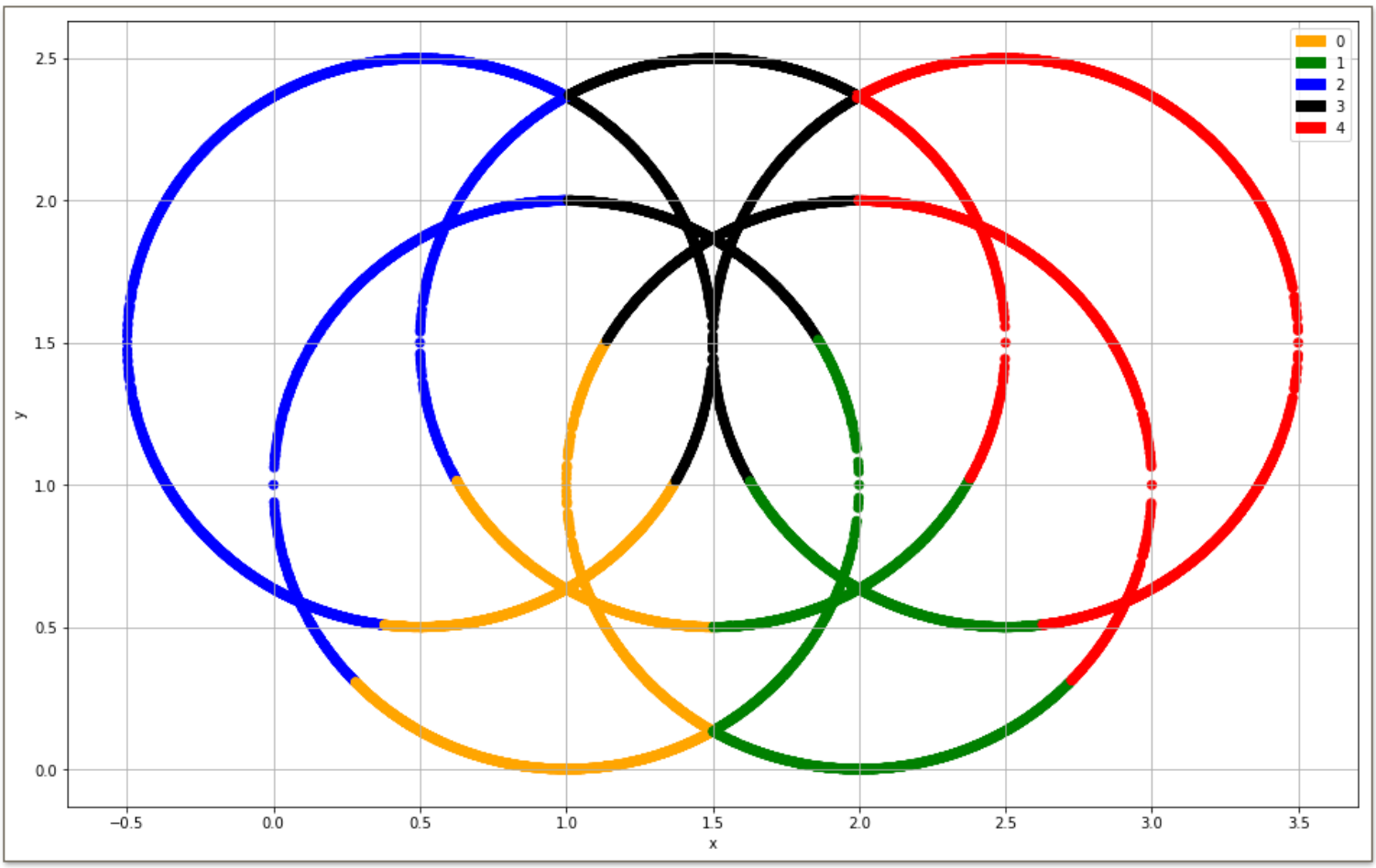
Index	0	1	2	3
const	-4.4355	0.1877	-2.7014	-8.7026
x	2.9533	-1.4679	1.4554	4.4382
y	0.0007	0.7201	0.6947	0.7194

b) (10 points). What is the Misclassification Rate?

The Misclassification rate without bagging technque:
0.5880059970014992

c) (10 points). What is the Root Average Squared Error?
The Root Average Squared Error without bagging technque:
0.55819520715770 44

d) (10 points). Redraw the above picture (i.e., the field y on the vertical axis and the field x on the horizontal axis), however, use the predicted ring label for coloring. The coloring scheme is 0 = orange, 1 = green, 2 = blue, 3 = black and 4 = red.

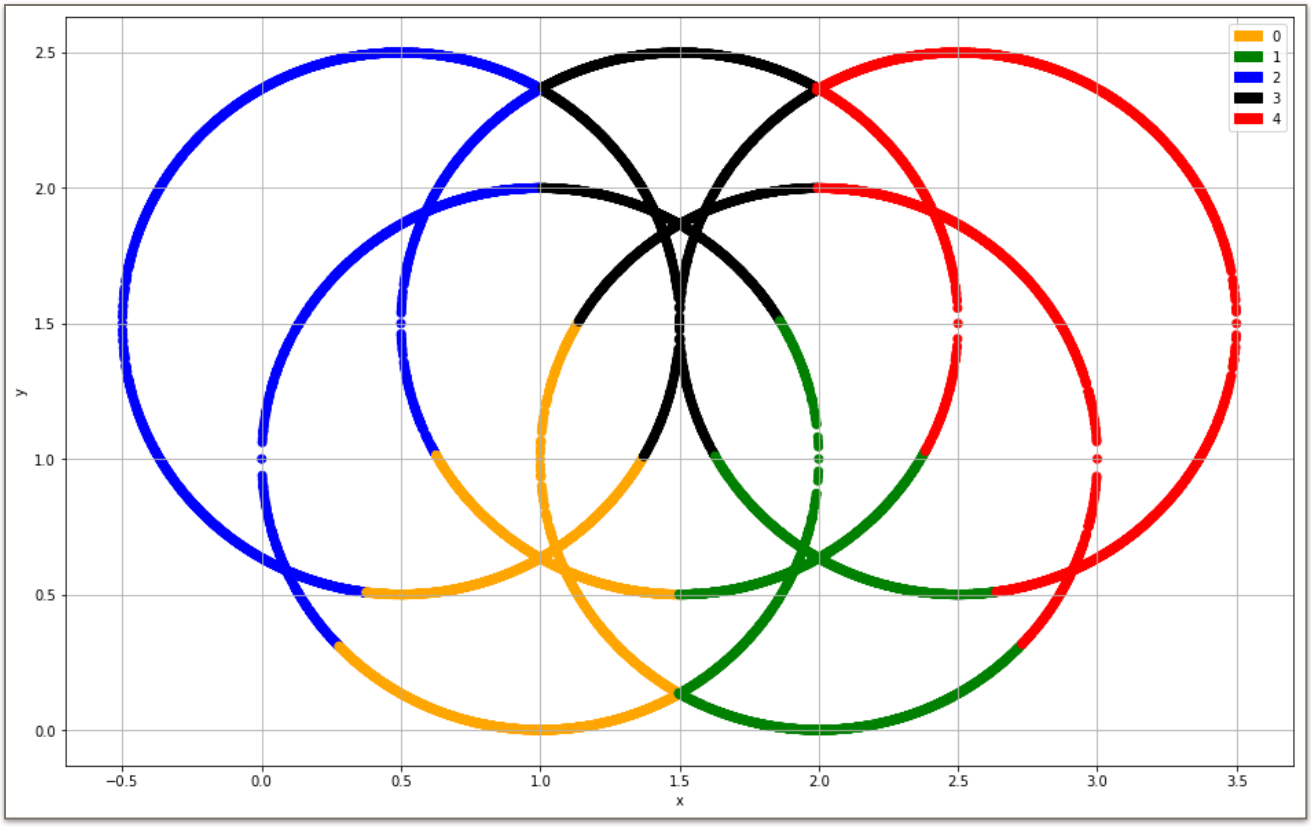


e) (40 points). List the Misclassification Rate and the Root Average Squared Error of the bootstrap results. The columns are the two metrics. The rows are the number of bootstraps. Also, include the no-bootstrap (i.e., zero number of bootstrap) metrics.

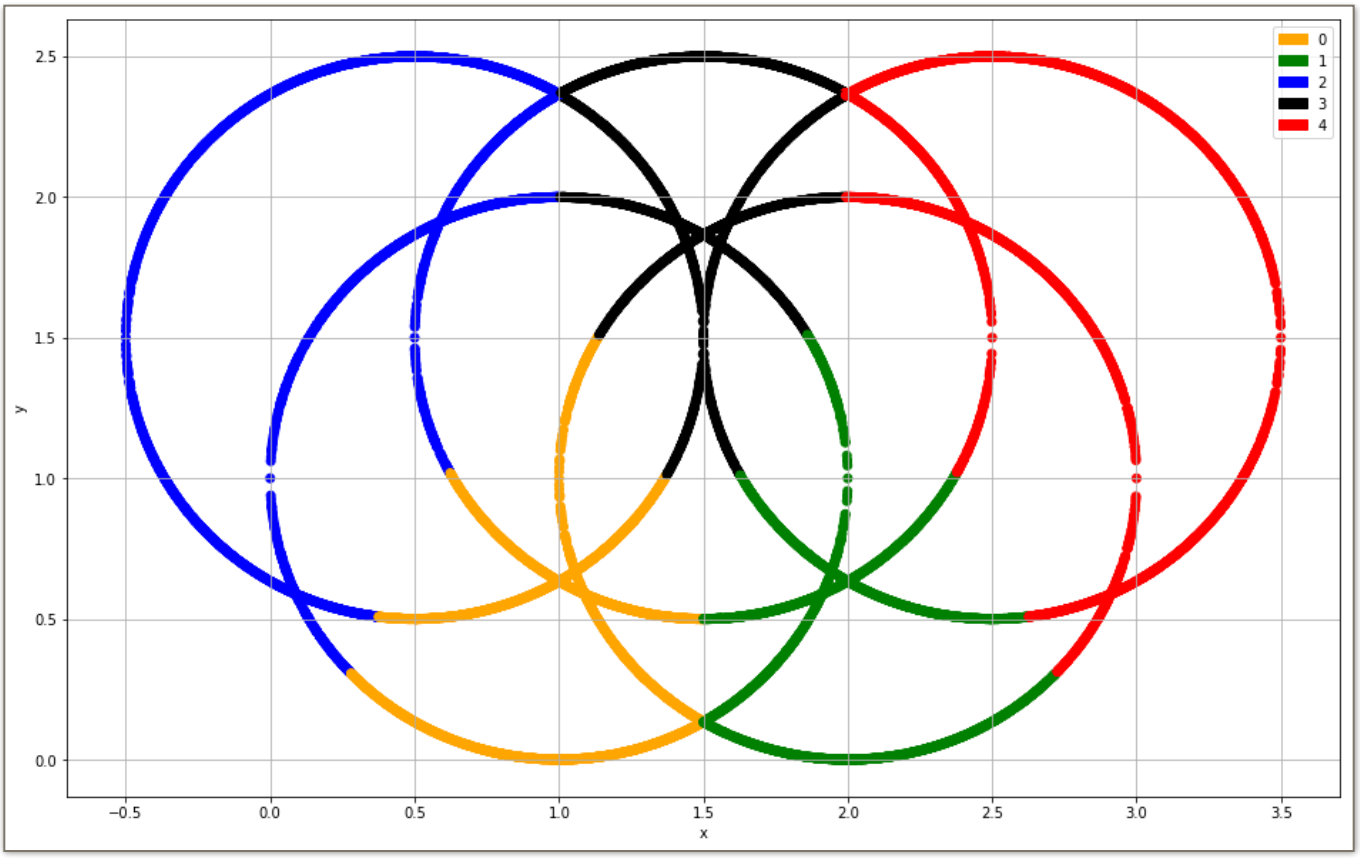
Index	MisClassRate	RASE
0	0.588006	0.558195
10	0.588256	0.558238
20	0.588256	0.558198
30	0.588206	0.558182
40	0.588206	0.558185
50	0.588206	0.558189
60	0.587906	0.558193
70	0.587906	0.558195
80	0.588006	0.558199
90	0.588006	0.558198
100	0.588056	0.558197

f). (10 points). Redraw the above picture (i.e., the field y on the vertical axis and the field x on the horizontal axis), however, use the predicted ring label for coloring. The coloring scheme is 0 = orange, 1 = green, 2 = blue, 3 = black and 4 = red.

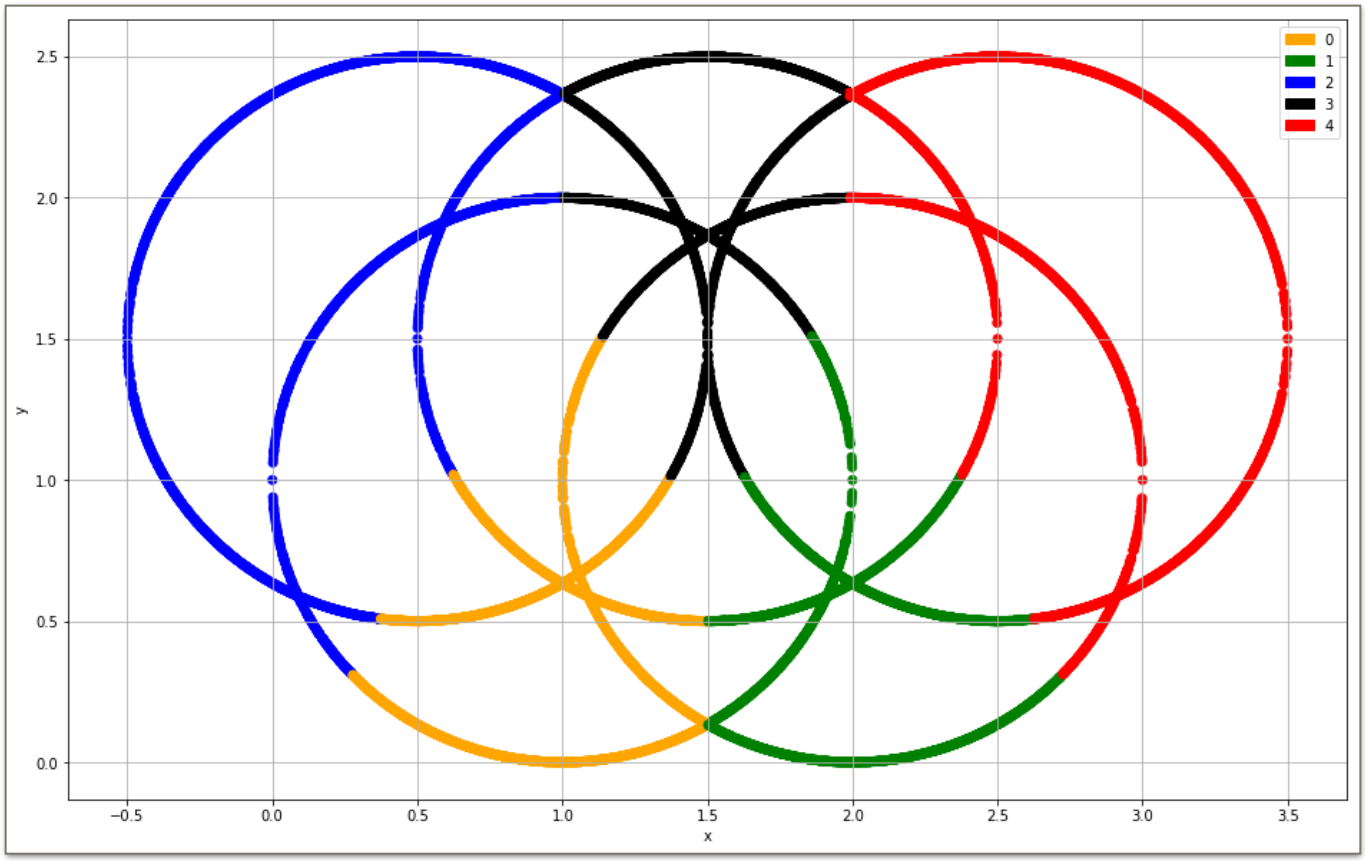
Number of bootstraps equals to 10:



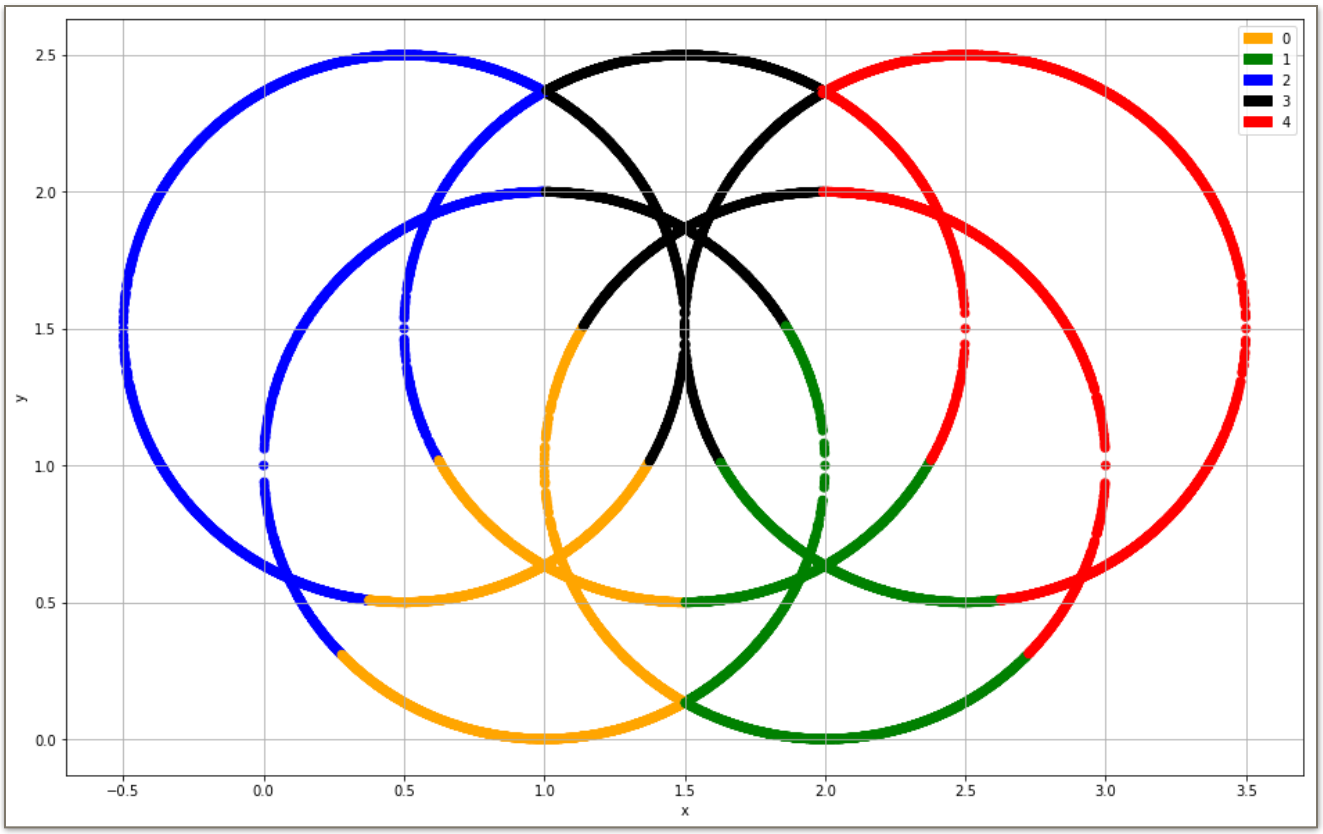
Number of bootstraps equals to 20:



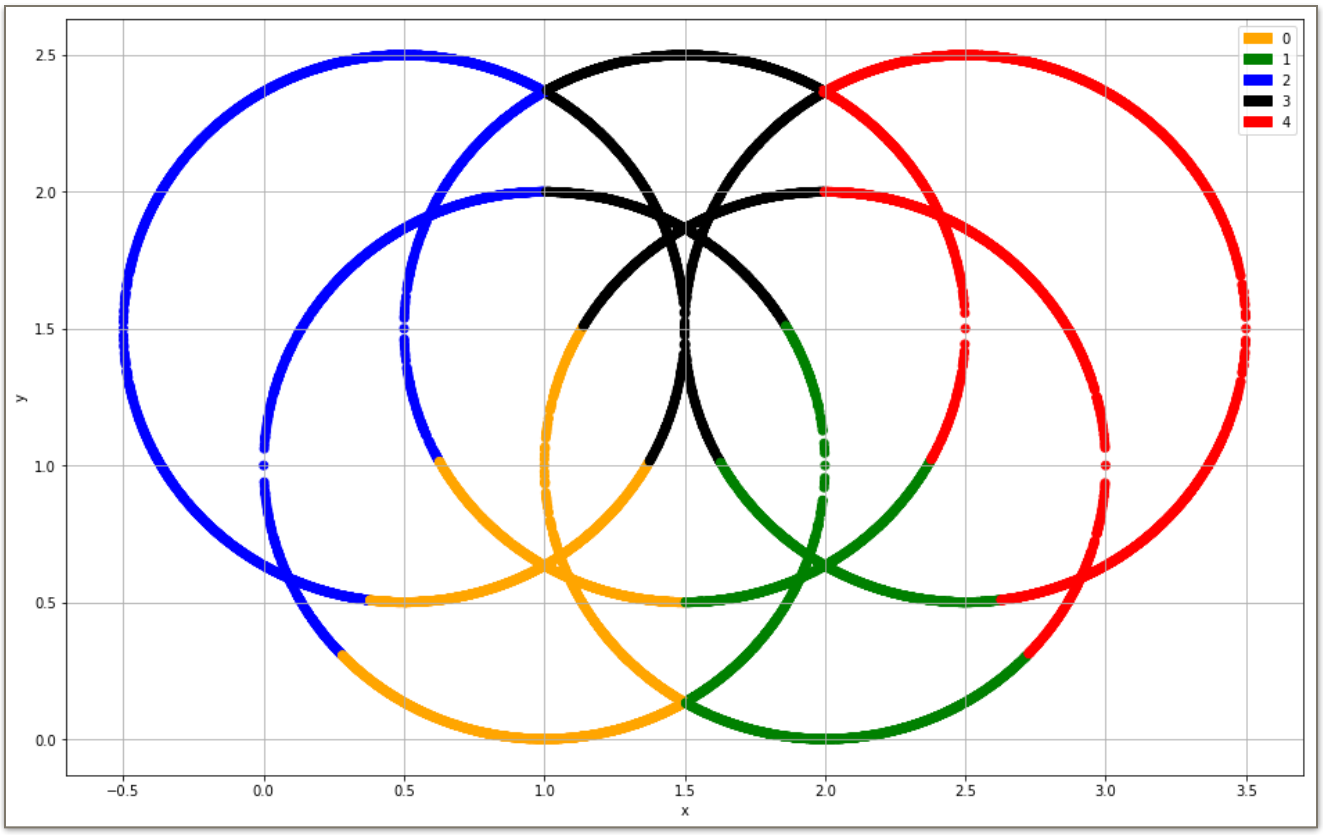
Number of bootstraps equals to 30:



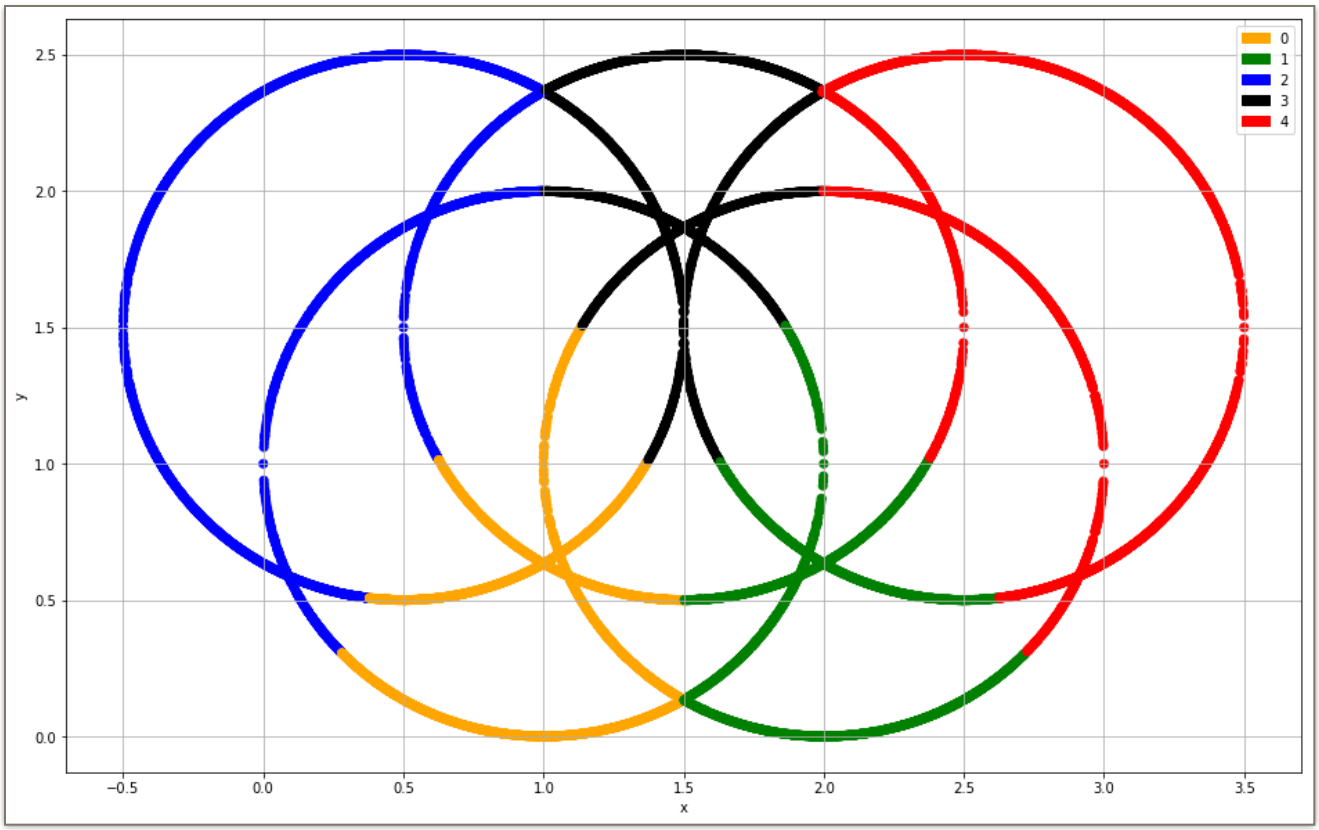
Number of bootstraps equals to 40:



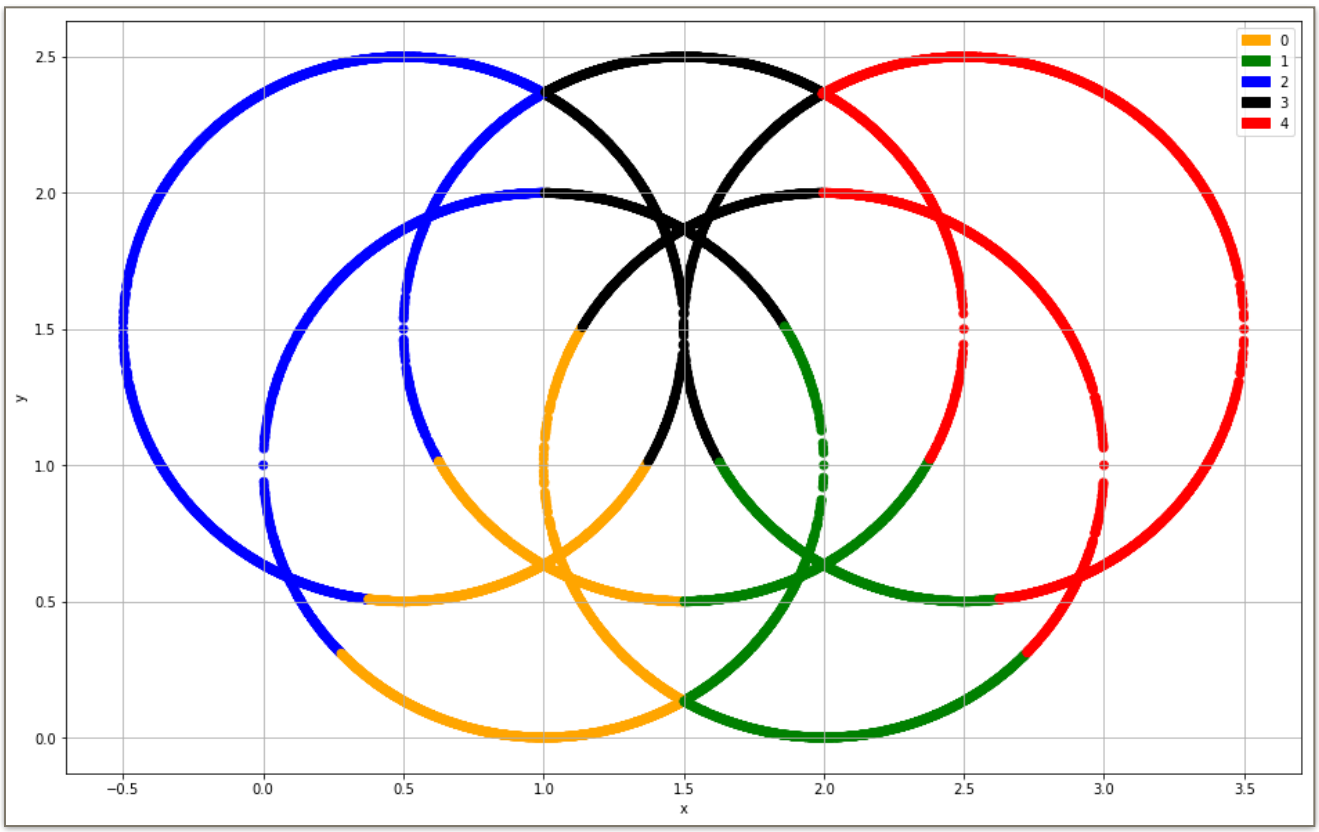
Number of bootstraps equals to 50:



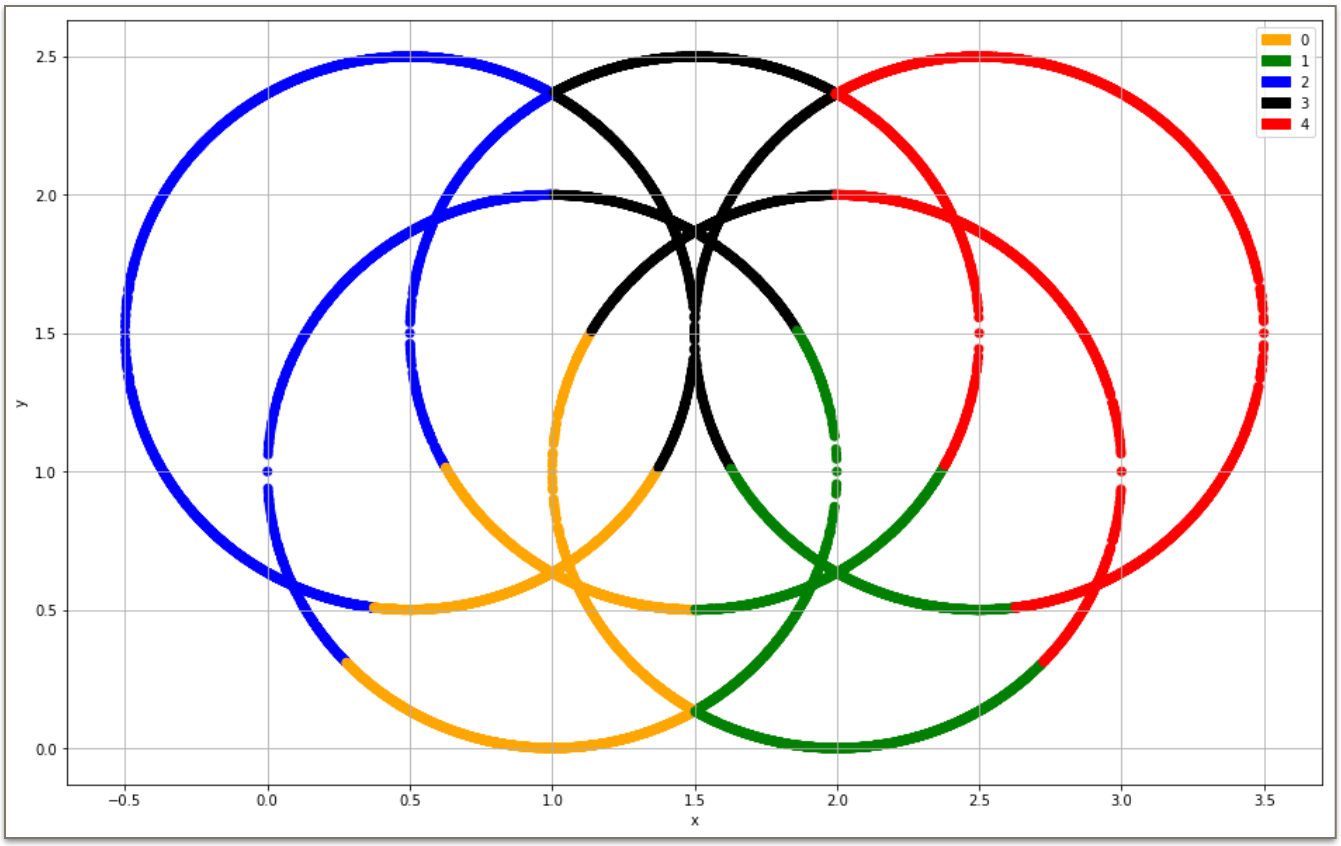
Number of bootstraps equals to 60:



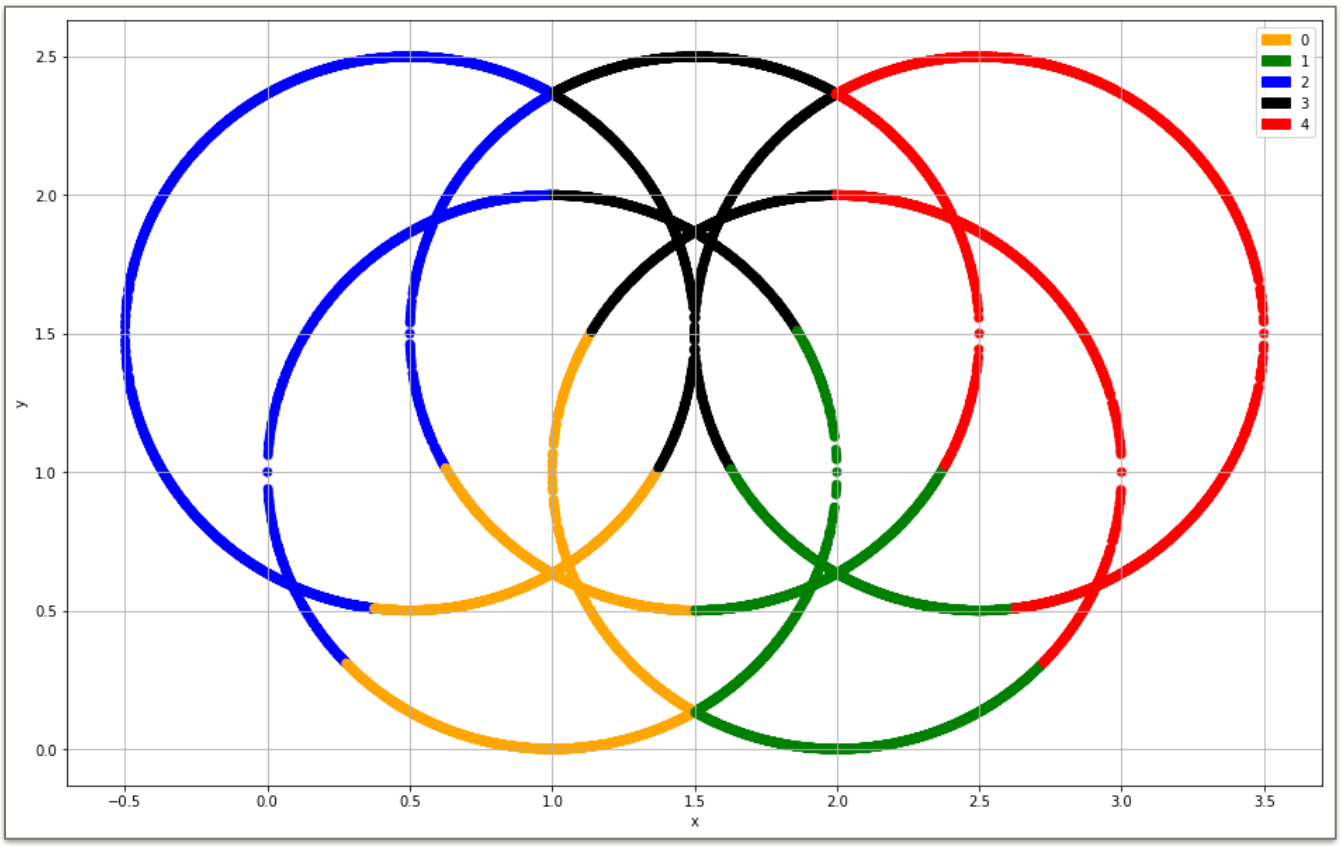
Number of bootstraps equals to 70:



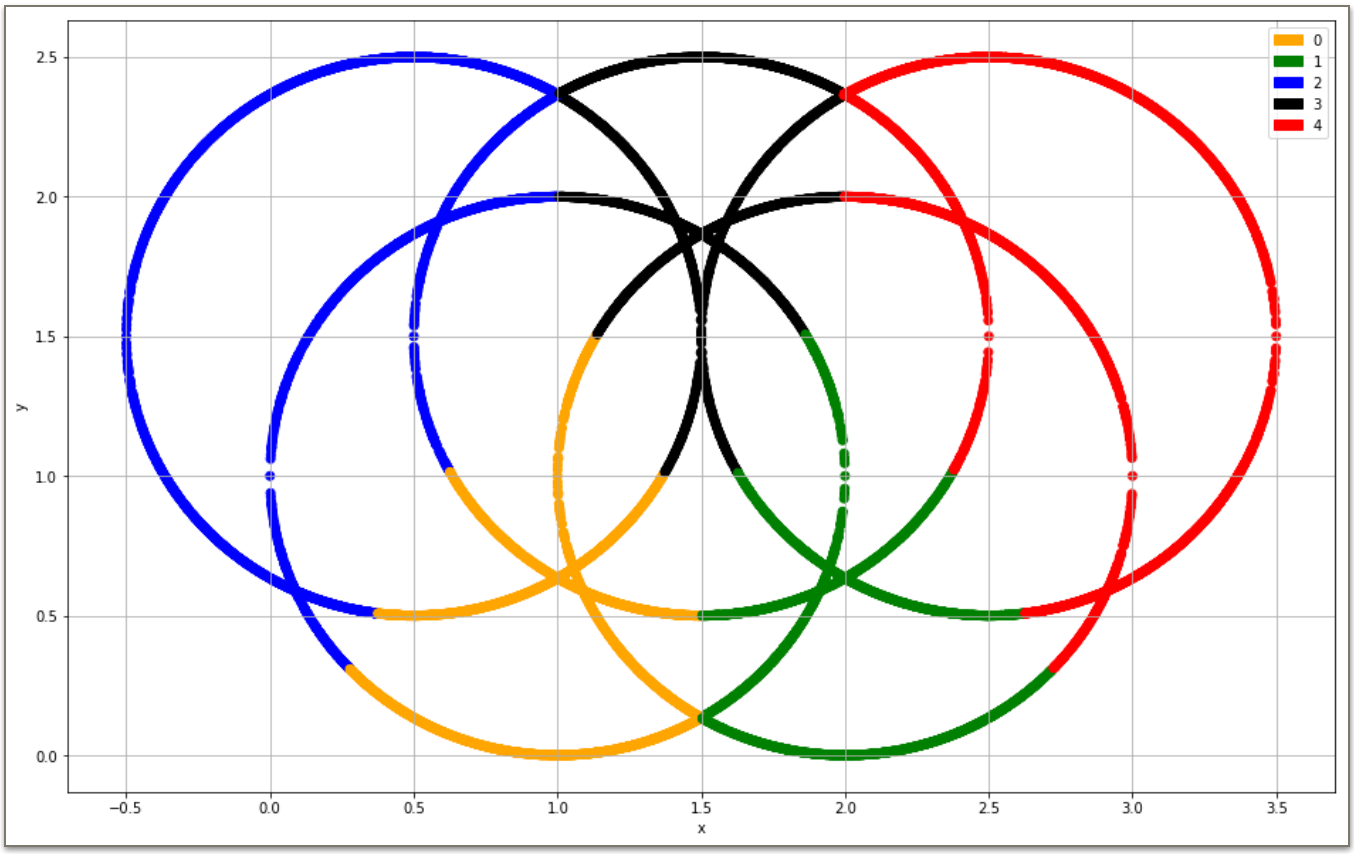
Number of bootstraps equals to 80:



Number of bootstraps equals to 90:



Number of bootstraps equals to 100:



g) (10 points). Compare the results between the bagging results and the non-bagging results. Briefly comment on the comparison.

Based on the comparison between the bagging results and the non bagging results, there is no obvious improvement after performing bagging.
The misclassification rate increases to 0.588256 and keeps the same in 10 and 20 bootstraps and decreases to 0.588206 and keeps the same in 30, 40, 50 bootstraps and then decreases to 0.587906 and keeps the same in 60 and 70 bootstraps and increases to 0.588006 and keeps the same in 80 and 90 bootstraps and increases to 0.588056 in the 100 bootstraps.
The RASE increases to 0.558238 in 10 bootstraps and decreases to 0.558198 in 20 bootstraps and decreases to 0.558182 in 30

bootstraps and keeps increasing from 40 to 80 bootstraps and then keeps decreasing from 90 to 100 bootstraps.

100 bootstraps are not enough to improve the prediction on so large and complicated data set.

Question #2:

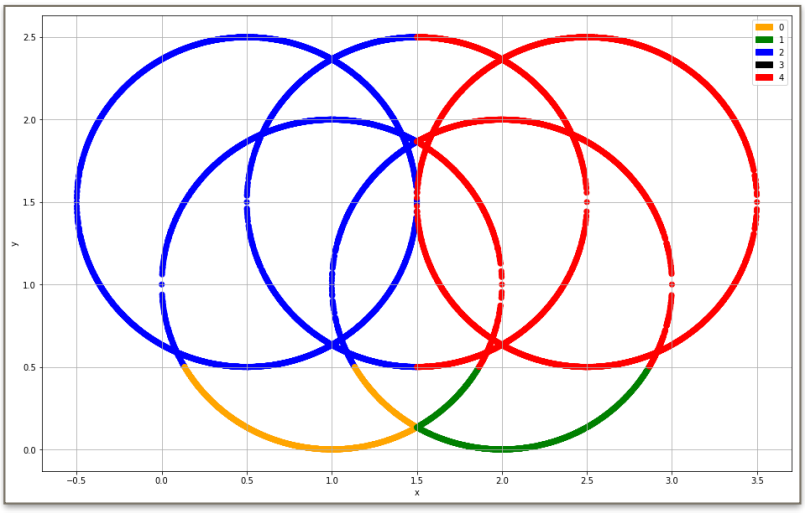
a) (10 points). What is the Misclassification Rate?

The Misclassification rate without boosting: 0.4641179410294853

b) (10 points). What is the Root Average Squared Error?

The Root Average Squared Error without boosting: 0.5507919160756358

c) (10 points). Redraw the above picture (i.e., the field y on the vertical axis and the field x on the horizontal axis), however, use the predicted ring label for coloring. The coloring scheme is 0 = orange, 1 = green, 2 = blue, 3 = black and 4 = red.

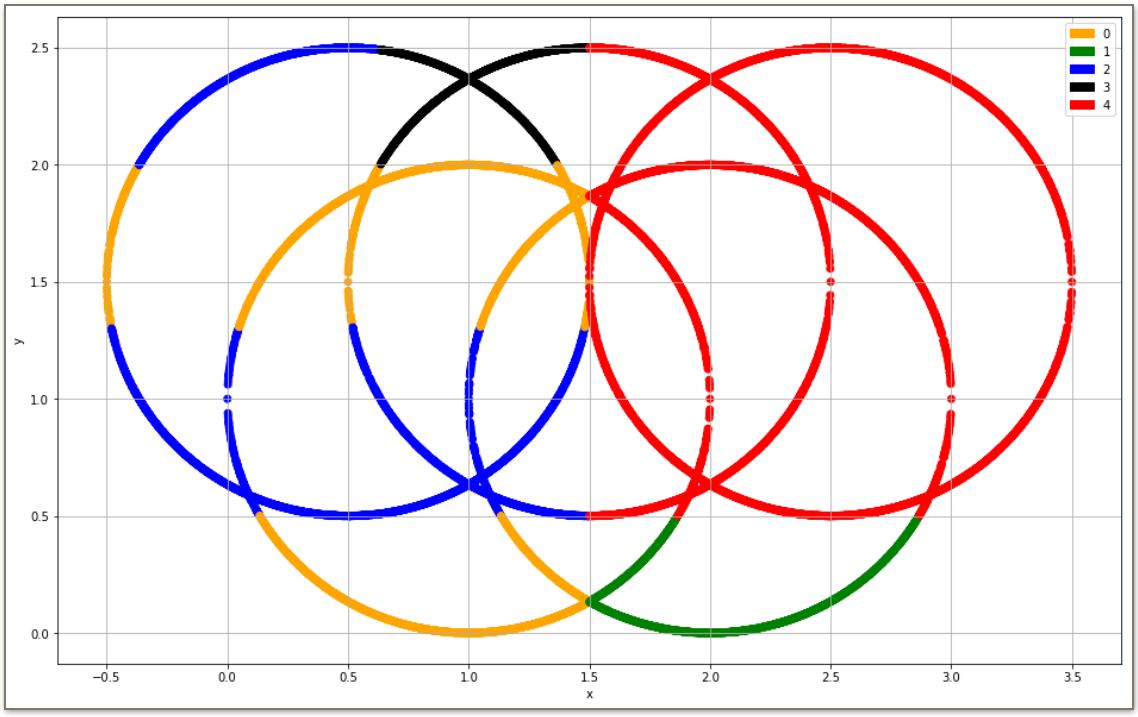


d) (50 points). List the Misclassification Rate and the Root Average Squared Error of the boosting results. The columns are the number of iterations performed and the two metrics. The rows are the maximum number of iterations. Also, include the no-boosting metrics.

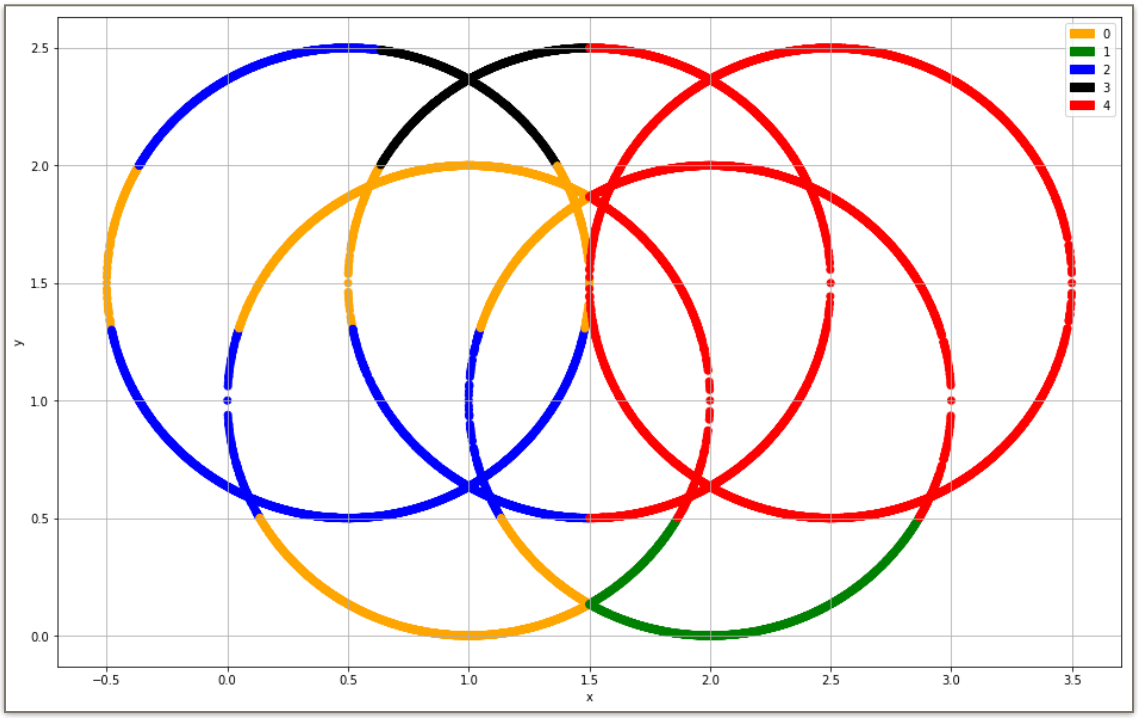
Index	MisClassRate	RASE
0	0.464118	0.550792
100	0.40005	0.534614
200	0.40005	0.534598
300	0.40005	0.534593
400	0.40005	0.534591
500	0.40005	0.534589
600	0.40005	0.534588
700	0.40005	0.534587
800	0.40005	0.534587
900	0.40005	0.534587
1000	0.40005	0.534586

e) (10 points). Redraw the above picture (i.e., the field y on the vertical axis and the field x on the horizontal axis), however, use the predicted ring label for coloring. The coloring scheme is 0 = orange, 1 = green, 2 = blue, 3 = black and 4 = red.

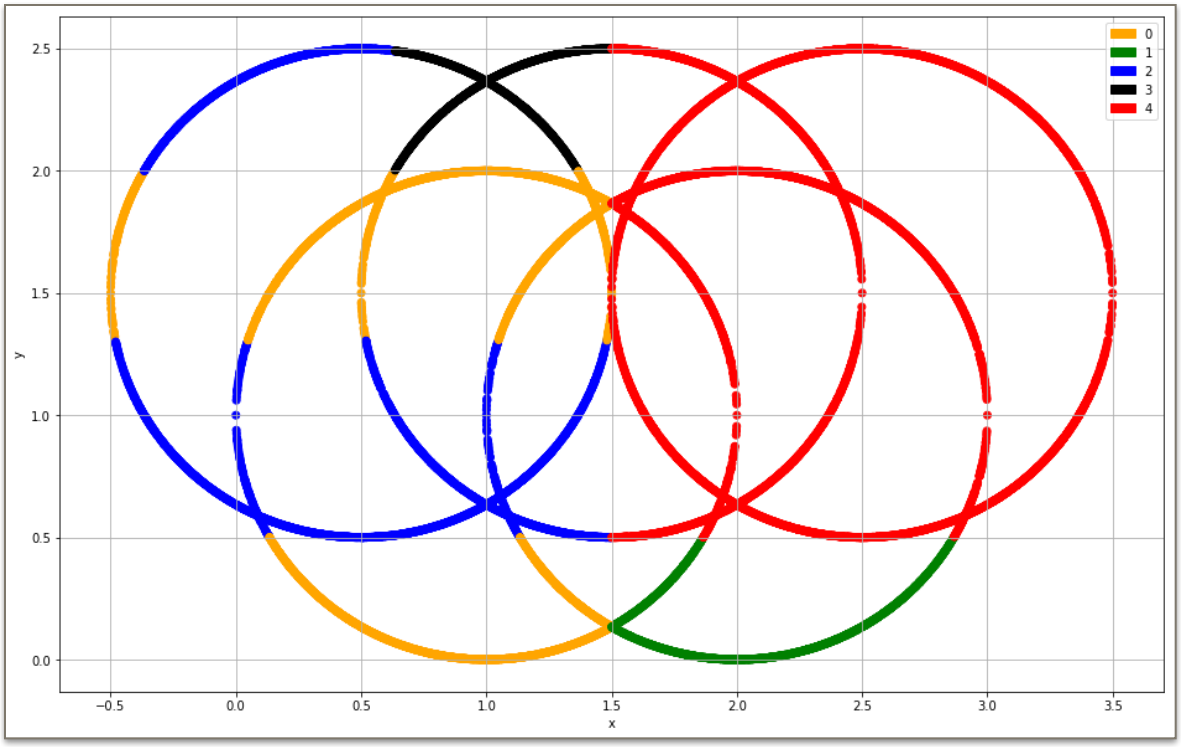
100 iterations:



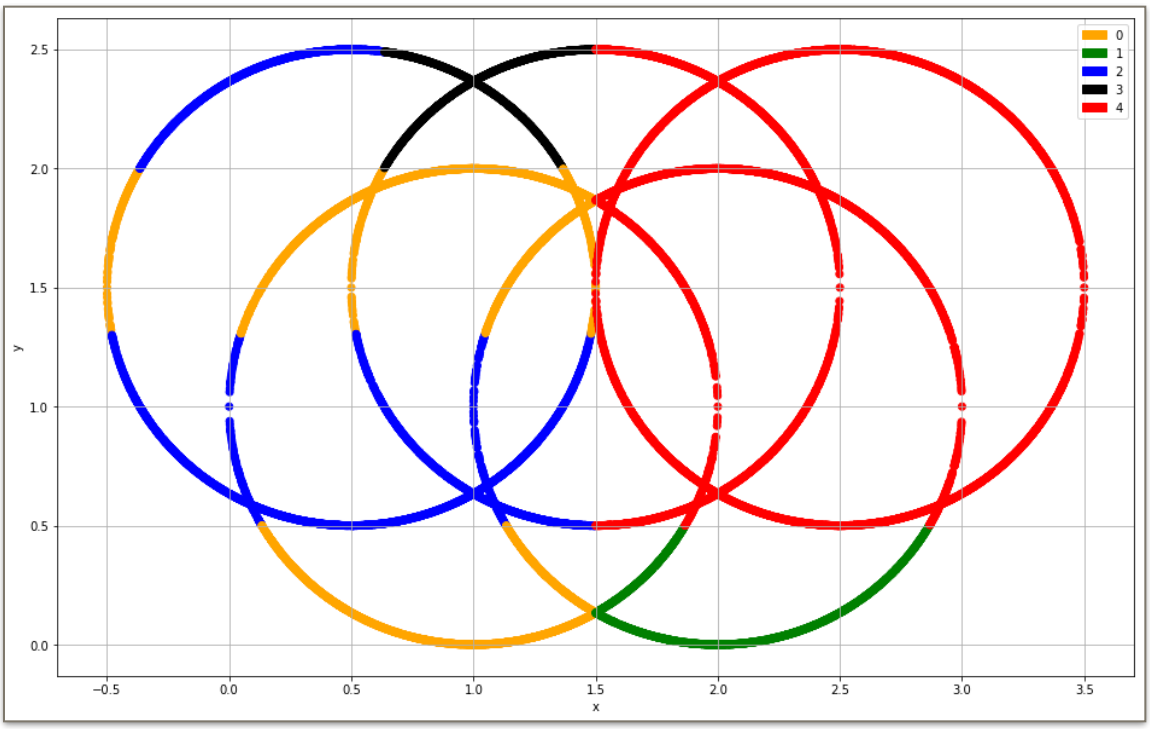
200 iterations:



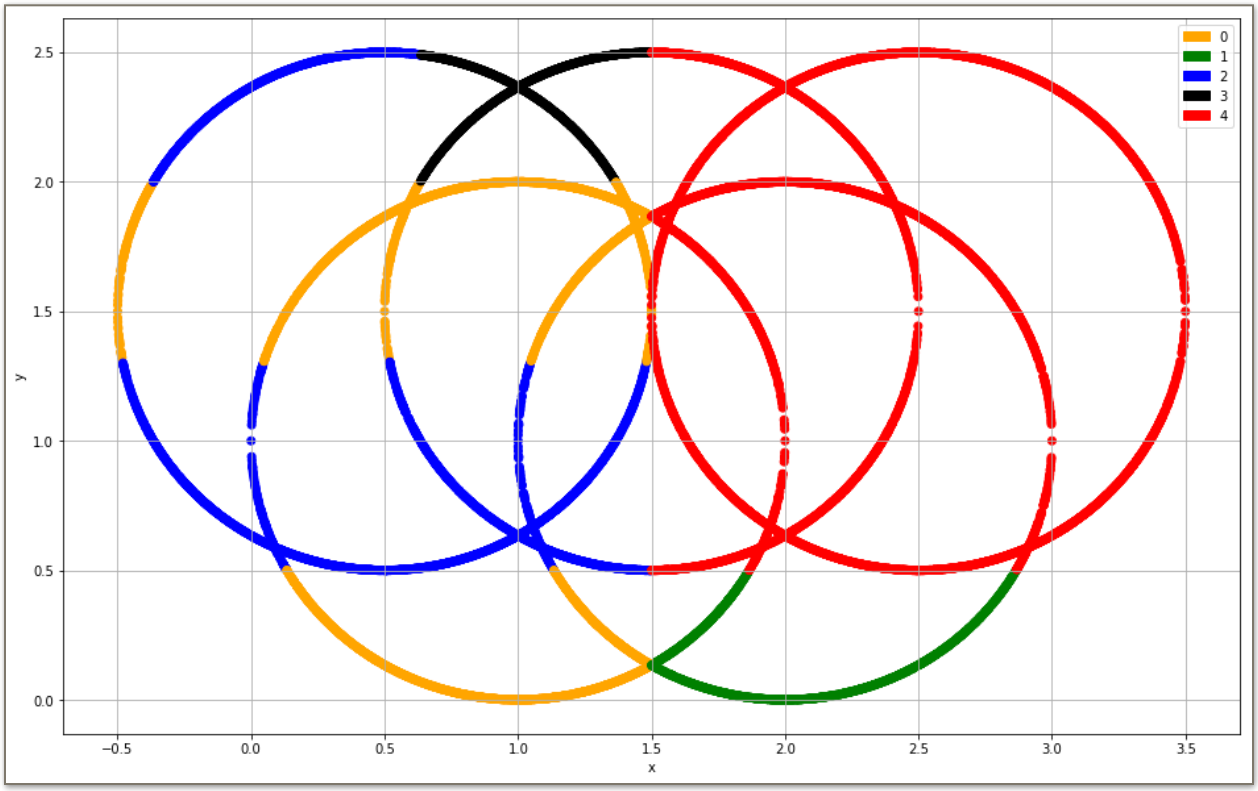
300 iterations:



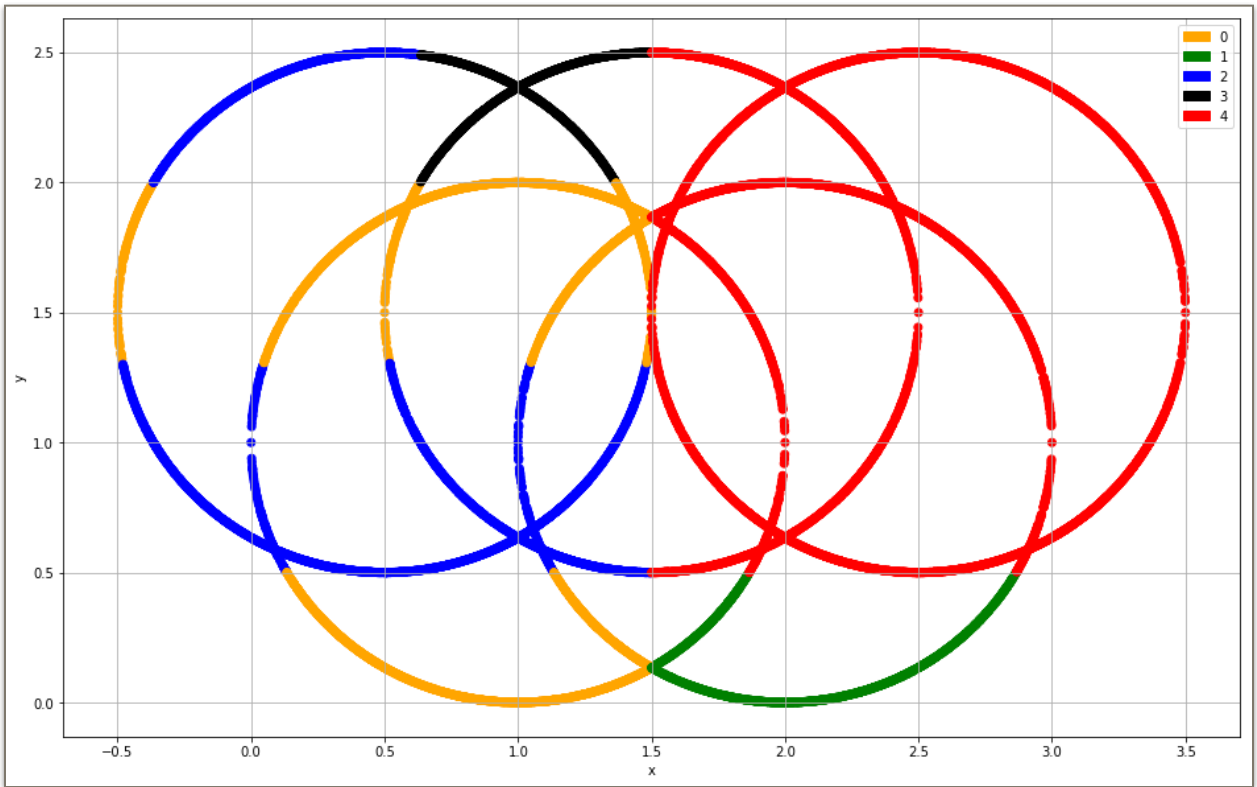
400 iterations:



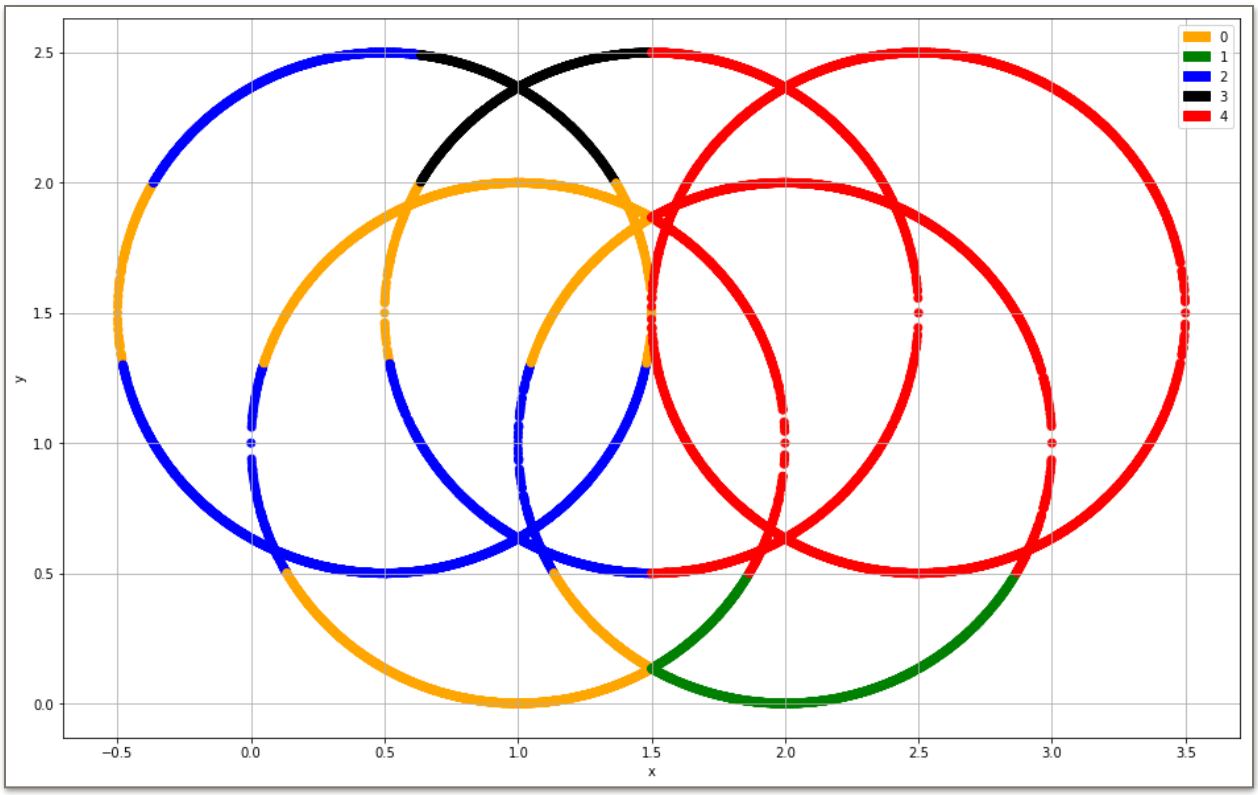
500 iterations:



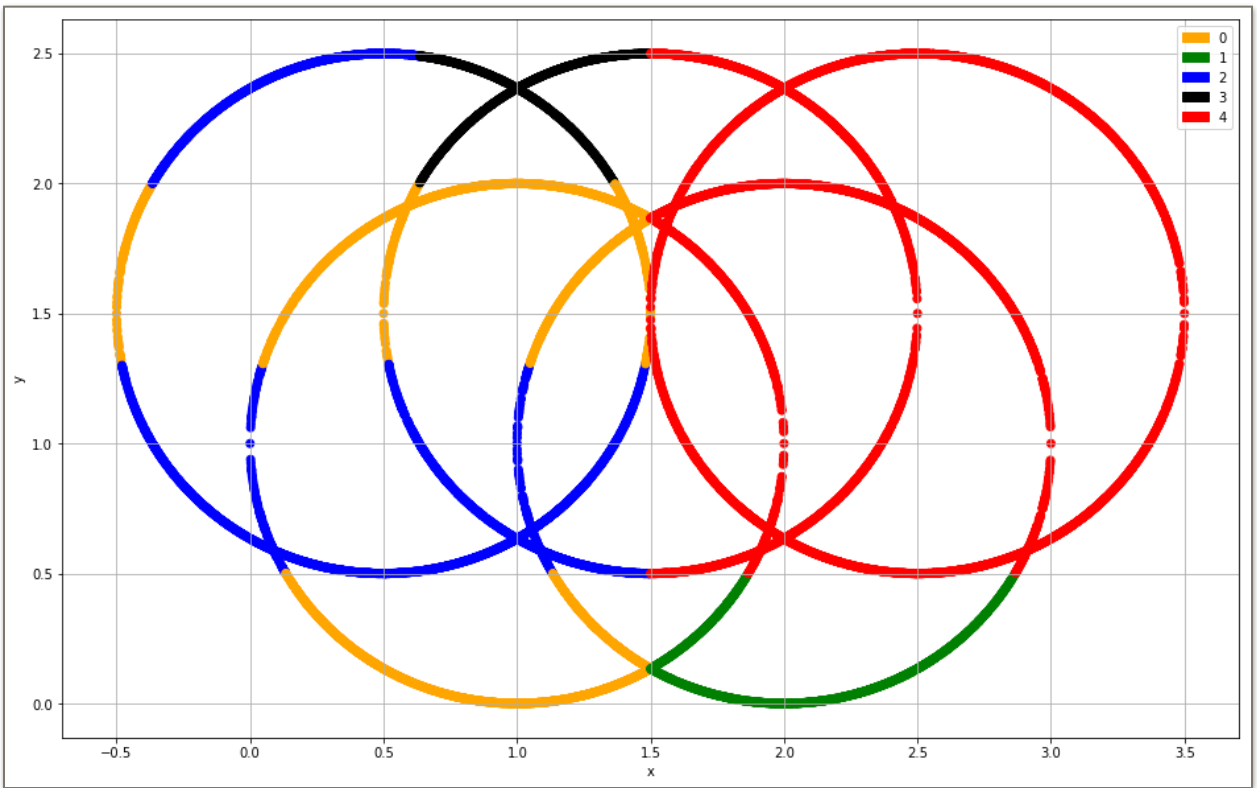
600 iterations:



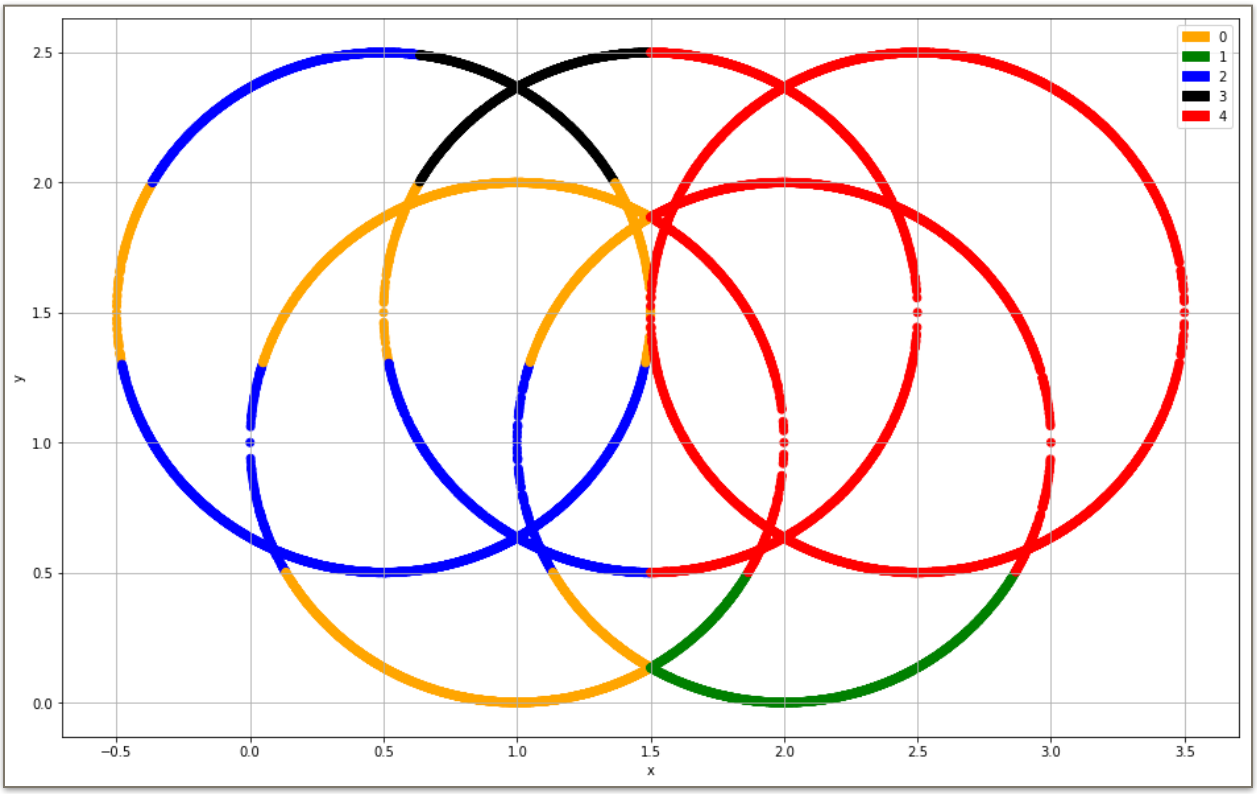
700 iterations:



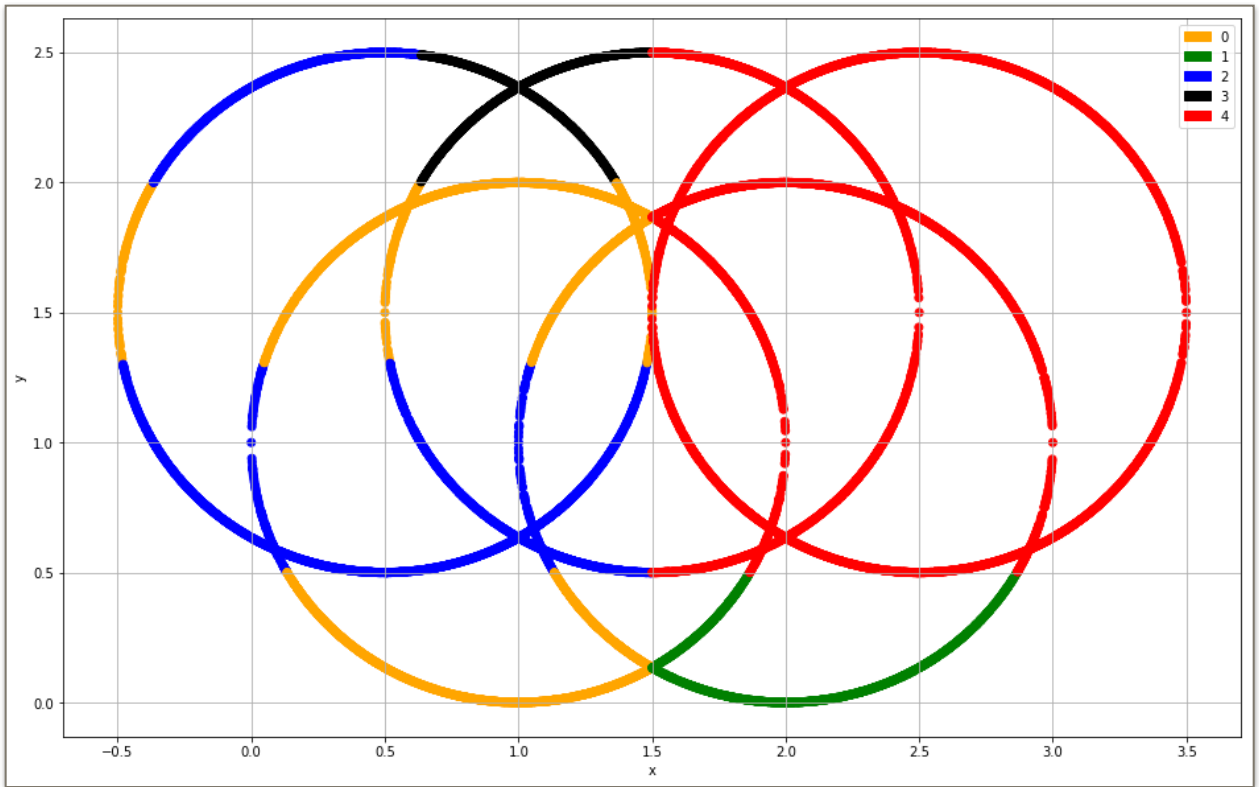
800 iterations:



900 iterations:



1000 iterations:



f) (10 points). Compare the results between the boosting results and the non-boosting results. Briefly comment on the comparison.

Based on the comparison between the boosting results and the non boosting results, there is a slight improvement after performing boosting.
The misclassification rate decreases to 0.40005 and keeps the same. The RASE decreases from 0.534614 to 0.534598 and fluctuates around 0.5345. 1000 iterations are not enough.