CS 584: Machine Learning
Spring 2020 Assignment 5

Question 1 (40 points)

The SpiralWithCluster.csv contains four variables.

Name	Description	Measurement Level	Role
Id	Case Identifier	Nominal	Identifier
X	x-coordinate	Interval	Feature
Y	y-coordinate	Interval	Feature
SpectralCluster	Cluster Identifier	Binary	Target

You are asked to use the Multi-Layer Perceptron (MLP) algorithm to classify SpectralCluster. You will use the sklearn.neural_network.MLPClassifier function with the following specifications.

1. Each hidden layer will have the same number of neurons
2. The initial learning rate is 0.1
3. The maximum number of iterations is 5000
4. The random seed is 20200408
5. The solver for weight optimization is lbfgs

Please answer the following questions based on your model.

a) (5 points) What percent of the observations have SpectralCluster equals to 1?
   Percent of the observations have SpectralCluster equals to 1 = 50.0 %

b) (15 points) You will search for the neural network that yields the lowest loss value and the lowest misclassification rate. You will use your answer in (a) as the threshold for classifying an observation into SpectralCluster = 1. Your search will be done over a grid that is formed by cross-combining the following attributes: (1) activation function: identity, logistic, relu, and tanh; (2) number of hidden layers: 1, 2, 3, 4, and 5; and (3) number of neurons: 1 to 10 by 1. List your optimal neural network for each activation function in a table. Your table will have four rows, one for each activation function. Your table will have six columns: (1) activation function, (2) number of layers, (3) number of neurons per layer, (4) number of iterations performed, (5) the loss value, and (6) the misclassification rate.

	Activation Function	No. of Layers	Neurons / layers	Iterations	Loss	Misclassification rate
0	Identity	1	1	9	0.666290	0.5
1	Logistic	3	6	543	0.000659	0.0
2	Relu	4	10	55	0.000107	0.0
3	tanh	5	6	84	0.000114	0.0

c) (5 points) What is the activation function for the output layer?
   Activation function for the output layer: Logistic

d) (5 points) Which activation function, number of layers, and number of neurons per layer give the lowest loss and the lowest misclassification rate? What are the loss and the misclassification rate? How many iterations are performed?

   ActivationFunc   relu
   NoOfLayers   4
   Neurons/layers   10
   NoOfIters   55
   Loss   0.000107361
   Misclassification%   0
   Name: 189, dtype: object

e) (5 points) Please plot the y-coordinate against the x-coordinate in a scatterplot. Please color-code the points using the predicted SpectralCluster (0 = Red and 1 = Blue) from the optimal MLP in (d). To obtain the full credits, you should properly label the axes, the legend, and the chart title. Also, grid lines should be added to the axes.

Graph for lowest loss
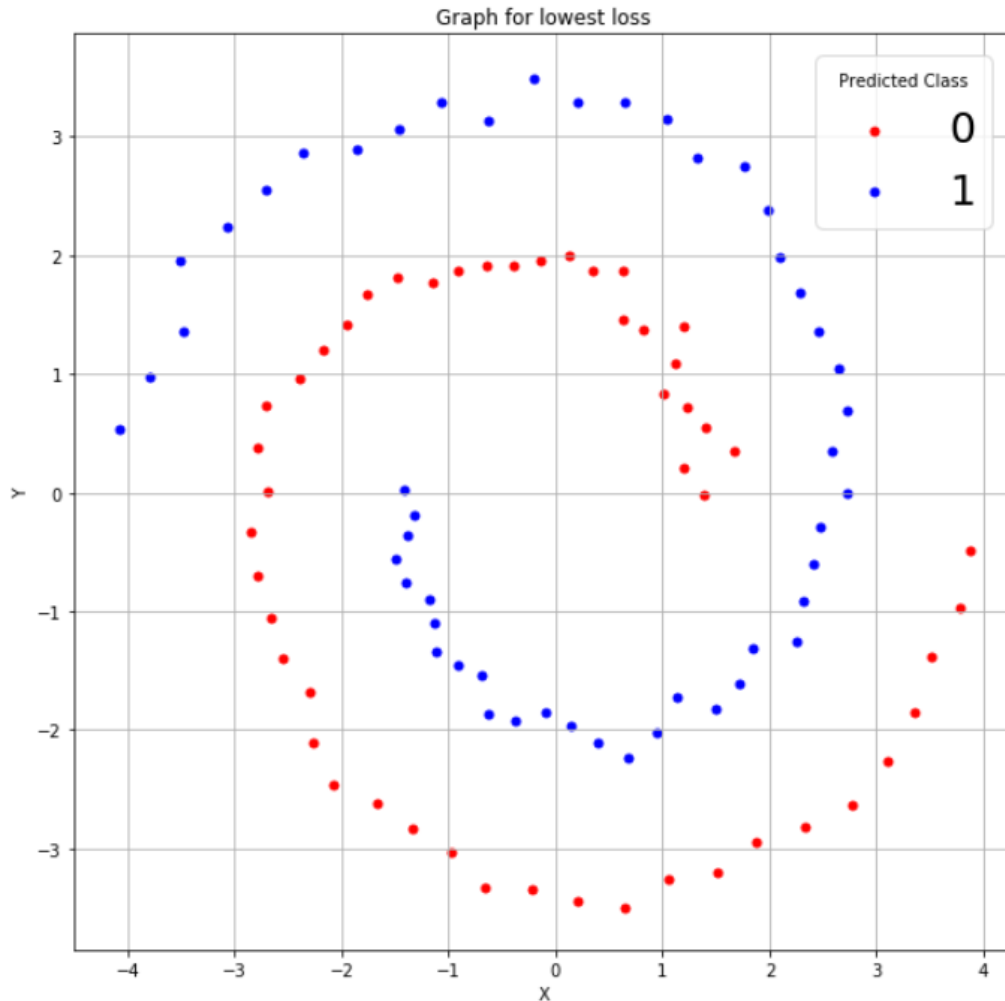
f) (5 points) What is the count, the mean and the standard deviation of the predicted probability Prob(SpectralCluster = 1) from the optimal MLP in (d) by value of the SpectralCluster? Please give your answers up to the 10 decimal places.

Spectral Cluster = 1

Count = 50

Mean = 0.9999980000

Standard Deviation = 0.0000030000

(Code yielded up to 7 decimal places only.)

Question 2 (60 points)

The SpiralWithCluster.csv contains four variables.

Name	Description	Measurement Level	Role
Id	Case Identifier	Nominal	Identifier
X	x-coordinate	Interval	Feature
Y	y-coordinate	Interval	Feature
SpectralCluster	Cluster Identifier	Binary	Target

You are asked to use the Support Vector Machine (SVM) algorithm to classify SpectralCluster. You will use the sklearn.svm.SVC function with the following specifications.

1. The linear kernel
2. The decision function shape is One Over Rest (OVR)
3. No limit on the number of iterations
4. The random seed is 20200408

Please answer the following questions based on your model.

a) (5 points) What is the equation of the separating hyperplane? Please state the coefficients up to seven decimal places.
   The intercept and co – efficient found are:
   Intercept = [0.003345]
   Coefficients = [[0.05333512 0.32868383]]

   Equation of separating hyperplane = 0.05333512 * x + 0.32868383 * y + 0.003345 = 0

b) (5 points) What is the misclassification rate?

   Misclassification rate = 0.5

c) (5 points) Please plot the y-coordinate against the x-coordinate in a scatterplot. Please color-code the points using the predicted SpectralCluster (0 = Red and 1 = Blue). Besides, plot the hyperplane as a dotted line to the graph. To obtain the full credits, you should properly label the axes, the legend, and the chart title. Also, grid lines should be added to the axes.

Support Vector Machine on input data
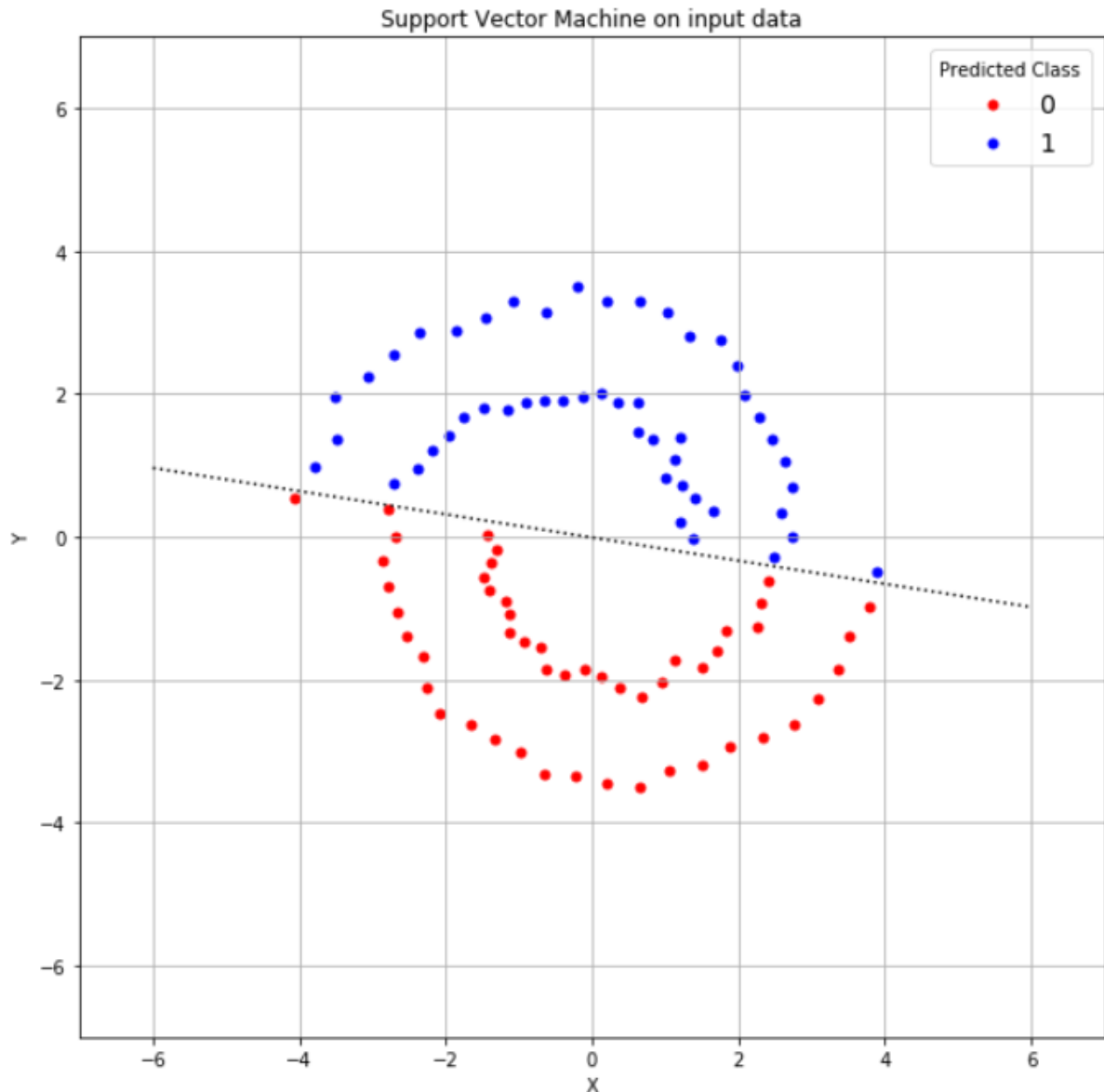
d) (10 points) Please express the data as polar coordinates. Please plot the theta-coordinate against the radius-coordinate in a scatterplot. Please color-code the points using the SpectralCluster variable ($0 =$ Red and $1 =$ Blue). To obtain the full credits, you should properly label the axes, the legend, and the chart title. Also, grid lines should be added to the axes.

Polar Coordinates Plot
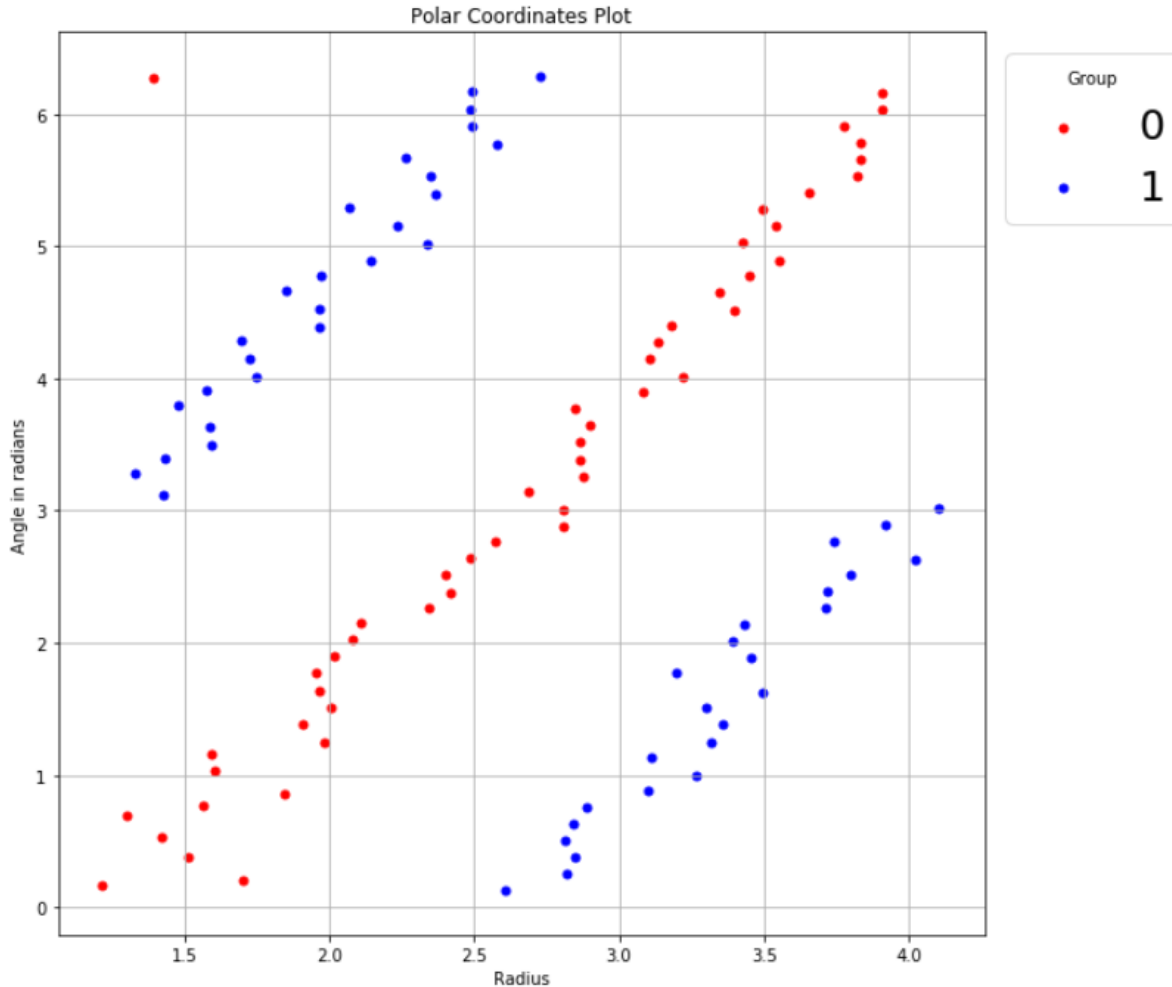
e) (10 points) You should expect to see three distinct strips of points and a lone point. Since the SpectralCluster variable has two values, you will create another variable, named Group, and use it as the new target variable. The Group variable will have four values. Value 0 for the lone point on the upper left corner of the chart in (d), values 1, 2,and 3 for the next three strips of points.
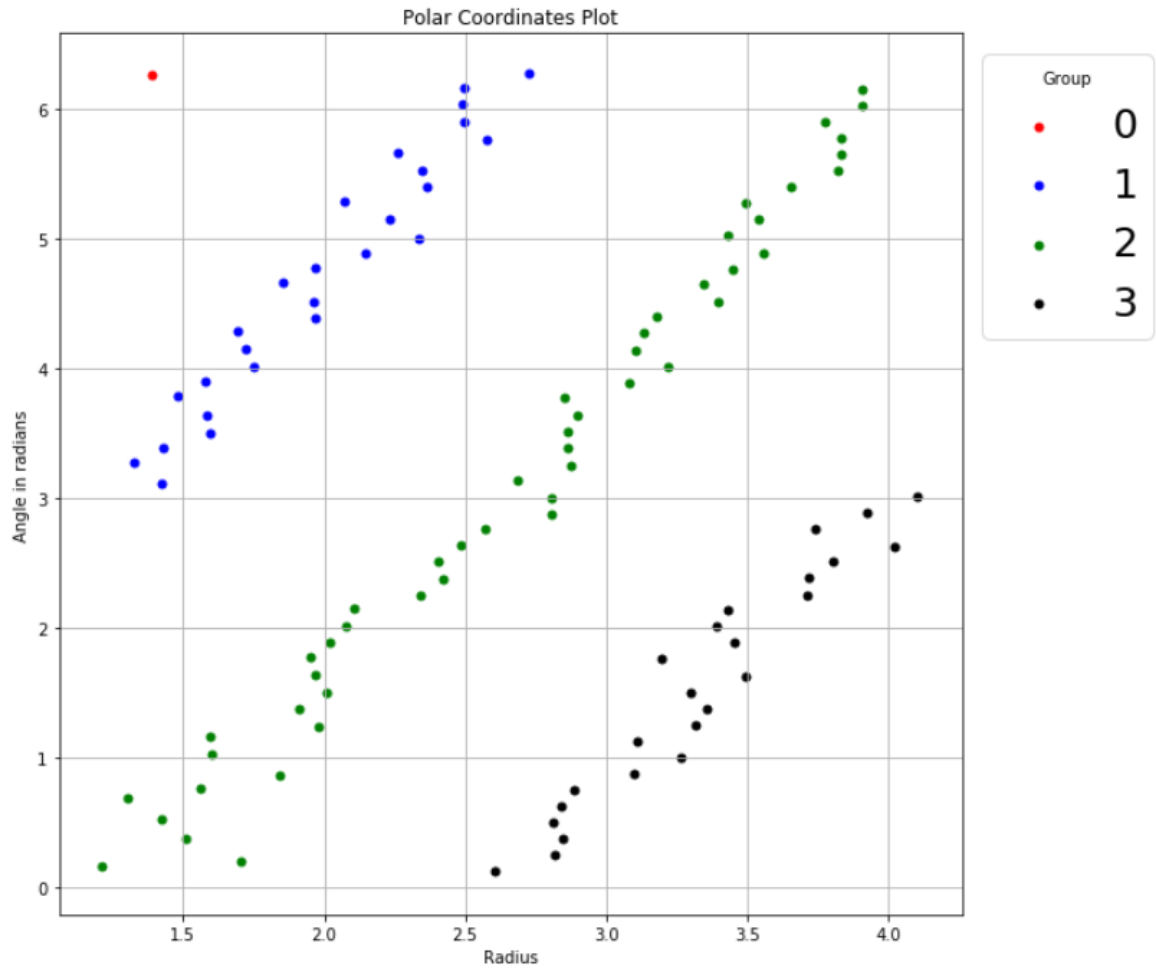
Polar Coordinates Plot

Please plot the theta-coordinate against the radius-coordinate in a scatterplot. Please color-code the points using the new Group target variable (0 = Red, 1 = Blue, 2 = Green, 3 = Black). To obtain the full credits, you should properly label the axes, the legend, and the chart title. Also, grid lines should be added to the axes.

f) (10 points) Since the graph in (e) has four clearly separable and neighboring segments, we will apply the Support Vector Machine algorithm in a different way. Instead of applying SVM once on a multi-class target variable, you will SVM three times, each on a binary target variable.

SVM	0:	Group	0	versus	Group	1
SVM	1:	Group	1	versus	Group	2
SVM	2:	Group	2	versus	Group	3

Please give the equations of the three hyperplanes.

SVM 0

 Intercept [1.46912508]
 Coefficients = [[0.93378415 -0.45380249]]
The separating hyperplane is $0.93378415 *x + (-0.45380249)*y + 1.46912508 = 0$

SVM 1

Intercept [-0.87689426]

Coefficients = [[1.89209533   -0.89613249]]

The separating hyperplane is $1.89209533 *x + (-0.89613249)*y + (-0.87689426) = 0$

SVM 2

Intercept [-4.13284488]

Coefficients = [[2.01258355   -0.83756164]]

The separating hyperplane is $2.01258355 *x + (-0.83756164)*y + (-4.13284488) = 0$

g) (5 points) Please plot the theta-coordinate against the radius-coordinate in a scatterplot. Please color-code the points using the new Group target variable (0 = Red, 1 = Blue, 2 = Green, 3 = Black). Please add the hyperplanes to the graph. To obtain the full credits, you should properly label the axes, the legend, and the chart title. Also, grid lines should be added to the axes.

SVM on data for New Target Variable
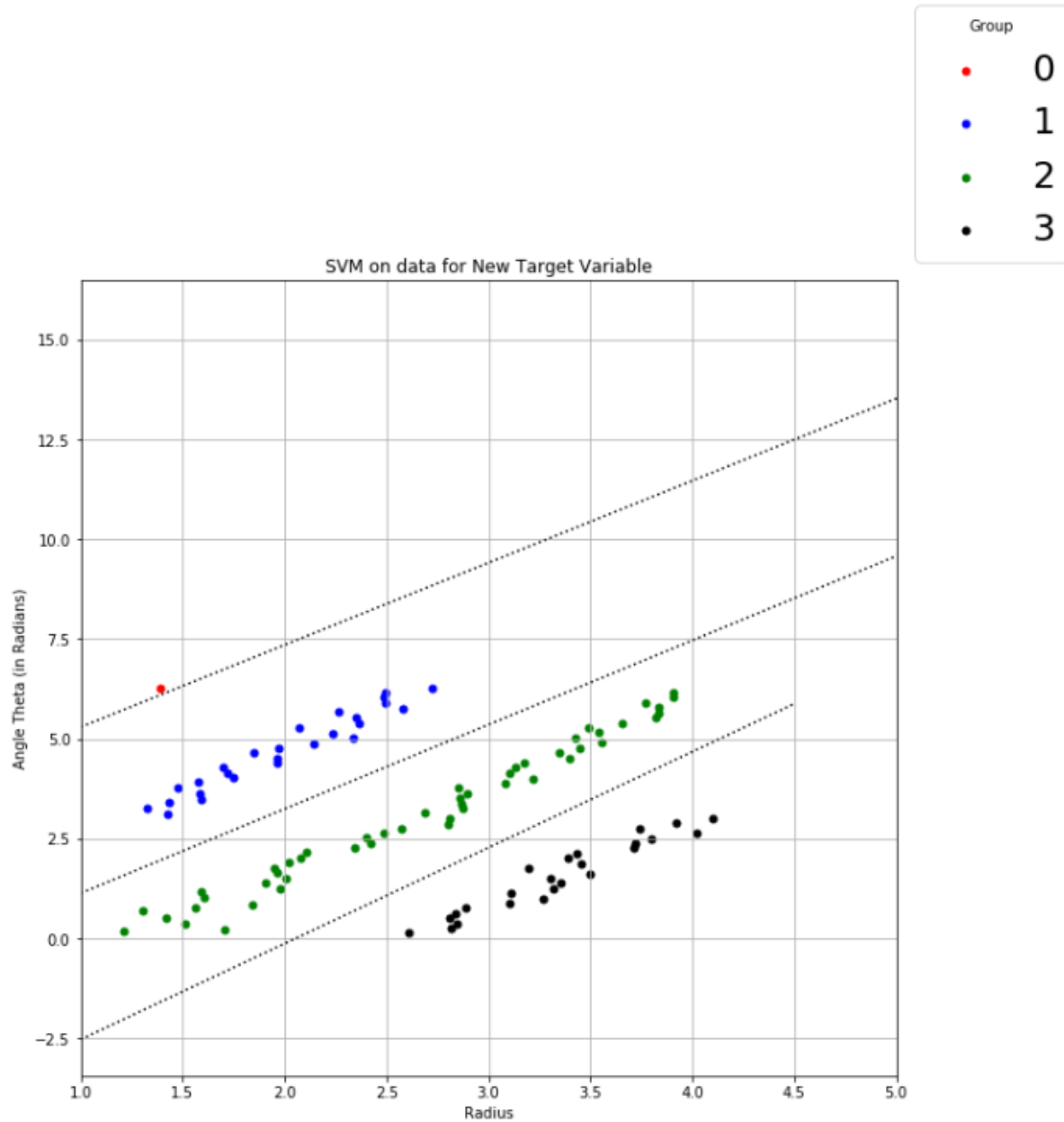
h) (10 points) Convert the observations along with the hyperplanes from the polar coordinates back to the Cartesian coordinates. Please plot the y-coordinate against the x-coordinate in a scatterplot. Please color-code the points using the SpectralCluster (0 = Red and 1 = Blue). Besides, plot the hyper-curves as dotted lines to the graph. To obtain the full credits, you should properly label the axes, the legend, and the chart title. Also, grid lines should be added to the axes.
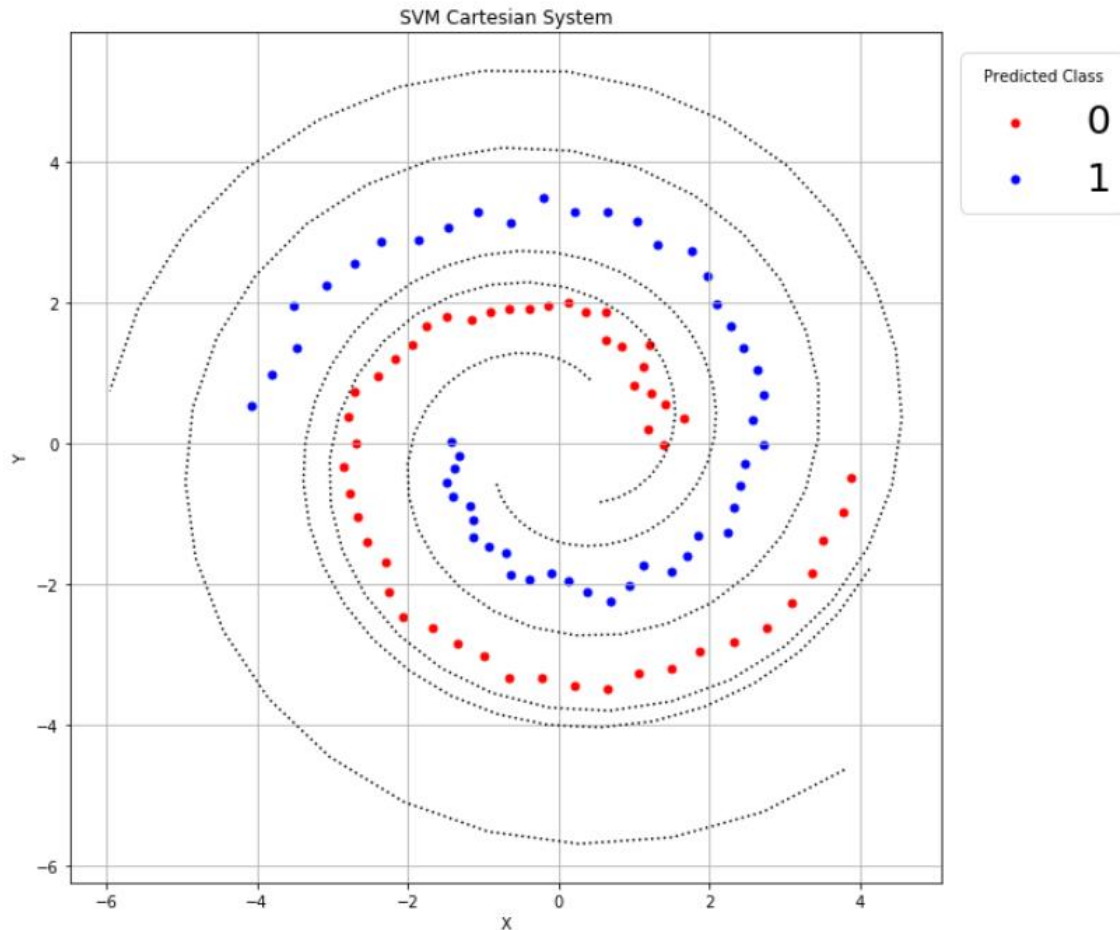
Based on your graph, which hypercurve do you think is not needed?

The curve associated with SVM 0 that maps Group 0 vs Group 1 is not needed as it is overlapping the dots and there are chances for it to misclassify the given data.

REFERENCE
1. Professor's code given during lectures.