CS 584-04: Machine Learning

Autumn 2019 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 20191108
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?

Ans:

col_0	Counts	Percentage
SpectralCluster		
0	50	50.0
1	50	50.0
All	100	100.0

The percentage of SpectralClusters=1 is **50%**

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 five 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.

Ans:

Optimal neural network for each activation function of neurons is as follows, as per question a I have used threshold of 0.50 as the percentage of SpectralClusters=1 is **50%.**

	Number of Layers	Number of Neurons	Loss	Misclassification Rate	Activation Function	Number of Iterations
0	4	8	0.000071	0.0	relu	40
0	1	1	0.666290	0.5	identity	10
0	3	6	0.000525	0.0	logistic	266
0	4	10	0.000112	0.0	tanh	71

From above the neural network having lowest loss and misclassification value is highlighted above.

Activation_function	number of layers	number of neurons per layer	loss	misclassification rate	Iterations
relu	4	8	0.000071	0.0	40

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

Ans:

The activation function for the output layer is logistic. Using out_activation_ gives string Name of the output activation function. When I used with classifier which I modelled it gave me output as logostic layer.

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?

Ans:

As per attached screenshot table in question b we can see that, activation function with lowest loss and lowest misclassication rate is relu.

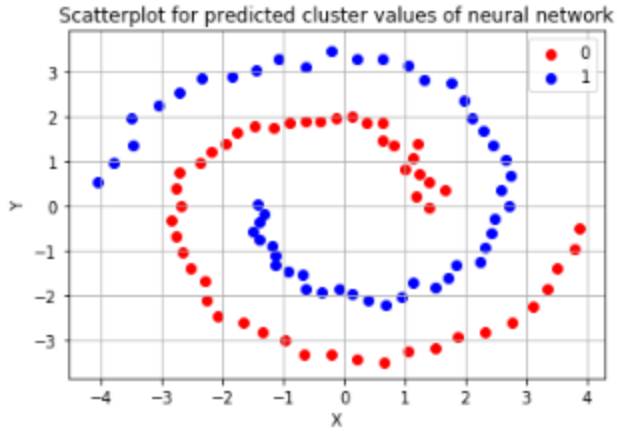
Activation_function	number of layers	number of neurons per layer	loss	misclassification rate	Iterations
relu	4	8	0.000071	0.0	40

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). 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.

Ans:

Following scatterplot depicts classification using predictor value 0 and 1

```
<matplotlib.legend.Legend at 0x17775aa26a0>
```



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.

Ans:

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 is as follows,

Count	50.00000000
Mean	0.9999960564
Standard deviation	0.0000132191

	0
count	50.000000
mean	0.999996
std	0.000013
min	0.999937
25%	1.000000
50%	1.000000
75%	1.000000
max	1.000000

As per given requirement here we calculated predicted probabilities only of spectralCluster=1 from optimal MLP. The values of count, standard deviation and mean are given as above.

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 20191108

Please answer the following questions based on your model.

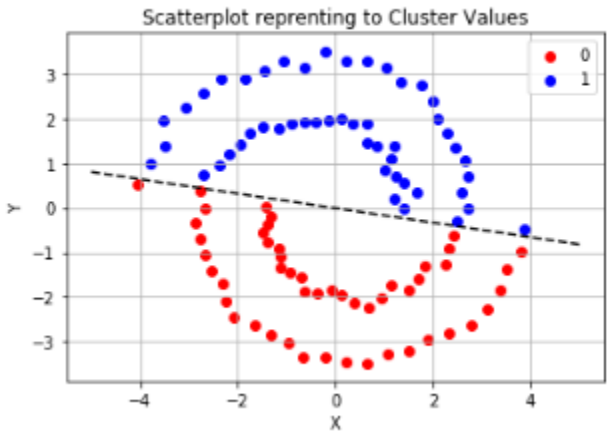
a) (5 points) What is the equation of the separating hyperplane? Please state the coefficients up to seven decimal places.
Ans:
0.0033449990212439536 + (0.05333511993003026) X +(0.32868383163967374) Y = 0

Intercept: 0.0033449990212439536
X: 0.05333511993003026
Y: 0.32868383163967374

b) (5 points) What is the misclassification rate?
Ans:
Misclassification rate is : **0.5 i.e. 50%**

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.
Ans:

```
<matplotlib.legend.Legend at 0x13d53baff60>
```

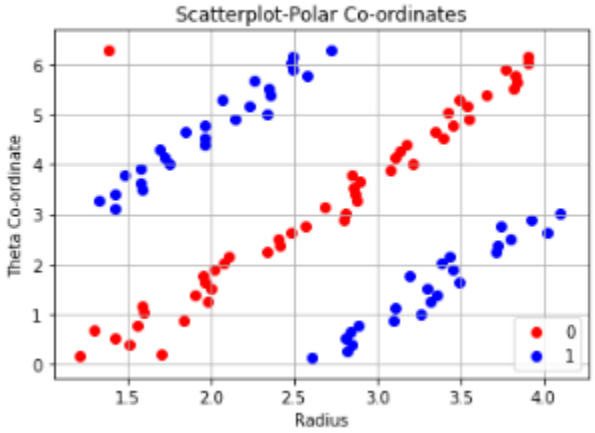


d) (10 points) Please express the data as polar coordinates. Please plot the theta-coordinate against the radius-coordinate in a scatterplot. Please color-code the points using the SpectralCluster

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.

Ans:

```
<matplotlib.legend.Legend at 0x2b8641fdbe0>
```

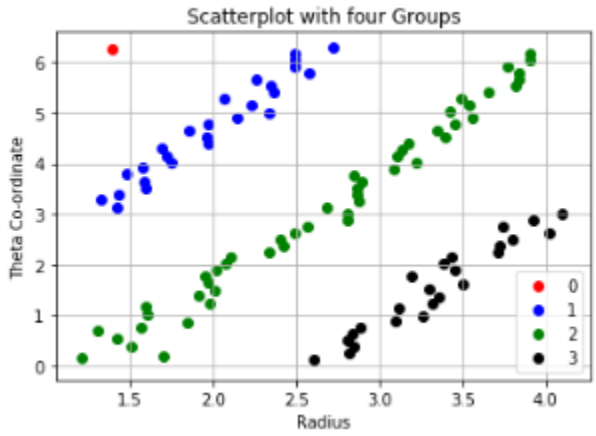


e) (10 points) You should expect to see three distinct strips of points and a lone point. Since the SpectralCluster variable has two values, you will create another variable, named Group, and use it as the new target variable. The Group variable will have four values. Value 0 for the lone point on the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

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.

Ans:

```
<matplotlib.legend.Legend at 0x13d53a07f28>
```



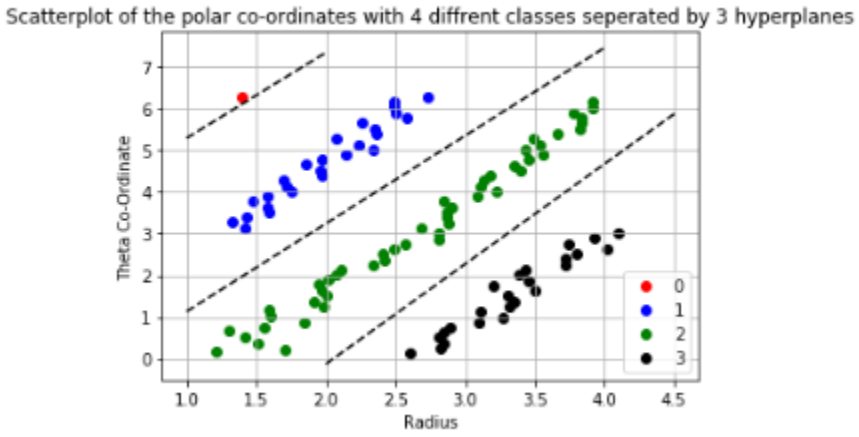
f) (10 points) Since the graph in (e) has four clearly separable and neighboring segments, we will apply the Support Vector Machine algorithm in a different way. Instead of applying SVM once on a multi-class target variable, you will SVM three times, each on a binary target variable.
SVM 0: Group 0 versus Group 1
SVM 1: Group 1 versus Group 2
SVM 2: Group 2 versus Group 3
Please give the equations of the three hyperplanes.
Ans:
o/p sceenshot:

```
THe equation of the hypercurve for SVM 0 is
1.4691250777389275  + ( 0.933784147085173 ) X +( -0.45380248720594835 ) Y = 0
THe equation of the hypercurwtve for SVM 1 is
-0.8768942577875997  + ( 1.8920953263166829 ) X +( -0.8961324867551237 ) Y = 0
THe equation of the hypercurve for SVM 2 is
-4.132844878075682  + ( 2.012583547086879 ) X +( -0.8375616435865432 ) Y = 0
```

SVM 0: Group 0 versus Group 1	1.4691250777389275 + (0.933784147085173) X + (-0.45380248720594835) Y = 0
SVM 1: Group 1 versus Group 2	-0.8768942577875997 + (1.8920953263166829) X +(-0.8961324867551237) Y = 0
SVM 2: Group 2 versus Group 3	-4.132844878075682 + (2.012583547086879) X +(-0.8375616435865432) Y = 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.
Ans:

```
<matplotlib.legend.Legend at 0x2b863b0e8d0>
```

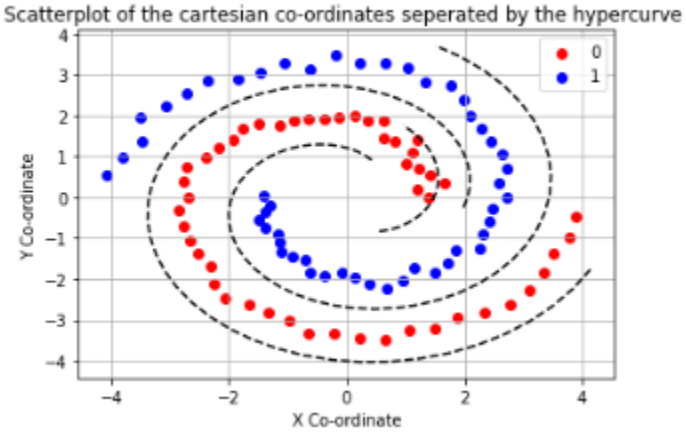


h) (10 points) Convert the observations along with the hyperplanes from the polar coordinates back to the Cartesian coordinates. Please plot the y-coordinate against the x-coordinate in a scatterplot. Please color-code the points using the SpectralCluster (0 = Red and 1 = Blue). Besides, plot the hyper-curves 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?
Ans:

```
<matplotlib.legend.Legend at 0x2b863fc8eb8>
```



When we convert the observations along with the hyperplanes from the polar coordinates back to the Cartesian coordinates. We get above scatterplot in which there are 3 hyper curves shown as dotted lines separating SpectralCluster as 0 and 1. As per analysis of graph, the innermost hyper-curve is not useful we can remove that. When we remove it following graph is obtained in which we are still able to differentiate between SpectralCluster (0/1).

```
<matplotlib.legend.Legend at 0x2b863cb6c88>
```