

CS 584-04: Machine Learning

Autumn 2019 Assignment 5

Question 1 (40 points)

- a) (5 points) What percent of the observations have SpectralCluster equals to 1?

Answer: Percent of the observations that have SpectralCluster equals to 1 is 50 percent.

- 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.

Answer:

```
stats_df.sort_values(by=['Loss value','Misclassification rate']).head()
```

	Activation function	Number of layers	Neurons per layer	RSquare	Iterations performed	Loss value	Misclassification rate
137	relu	4	8	1.0	40	0.000071	0.0
189	tanh	4	10	1.0	71	0.000112	0.0
199	tanh	5	10	1.0	74	0.000125	0.0
184	tanh	4	5	1.0	107	0.000134	0.0
183	tanh	4	4	1.0	178	0.000134	0.0

As it can be seen from above the first row retrieves the neural network with lowest Loss value and misclassification rate.

Activation Function	No. of layers	Neurons per layer	Loss Value	Misclassification rate
Relu	4	8	0.000071	0.0

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

Answer: As referred in the Neural network ppts provided by professor: Customize your output activation function by assigning the function name to the out_activation_ of the Neural Network object. Hence, name of the output activation function: 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?

Answer:

Below is the neural network that gives lowest loss: 0.000071, lowest misclassification rate: 0.0, Iterations: 40, Activation function: Relu, No. of layers: 4, No. of neurons per layer: 8

	Activation function	Number of layers	Neurons per layer	RSquare	Iterations performed	Loss value	Misclassification rate
137	relu	4	8	1.0	40	0.000071	0.0

Above result is achieved by taking into consideration all the activation functions and then choosing the that one activation function which has lowest loss and the lowest misclassification rate as shown below:

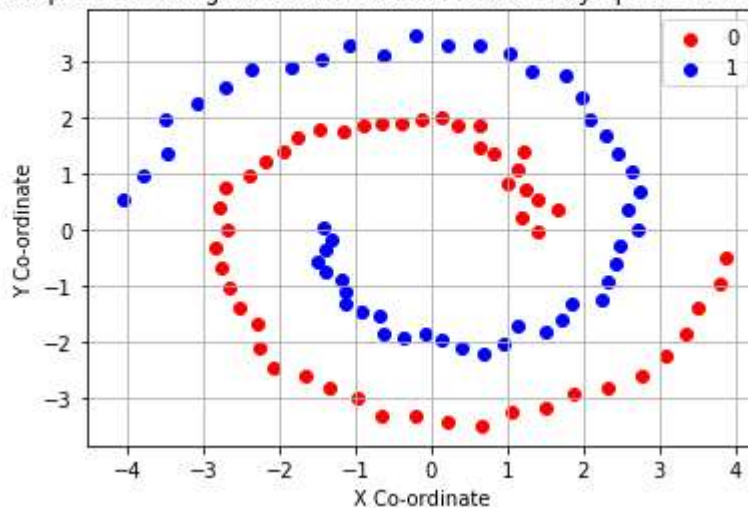
	Activation function	Number of layers	Neurons per layer	RSquare	Iterations performed	Loss value	Misclassification rate
137	relu	4	8	1.0	40	0.000071	0.0
0	identity	1	1	-1.0	10	0.666290	0.5
75	logistic	3	6	1.0	266	0.000525	0.0
189	tanh	4	10	1.0	71	0.000112	0.0

- 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.

Answer:

```
<matplotlib.legend.Legend at 0x11b276a5278>
```

Scatterplot according to Cluster Values Predicted by optimal neural network



- f) (5 points) What is the count, the mean and the standard deviation of the predicted probability $\text{Prob}(\text{SpectralCluster} = 1)$ from the optimal MLP in (d) by value of the SpectralCluster? Please give your answers up to the 10 decimal places.

Answer:

```
Count of predicted probability Prob(SpectralCluster = 1): 50
Mean of predicted probability Prob(SpectralCluster = 1): 0.9999960564
Standard Deviation of predicted probability Prob(SpectralCluster = 1): 1.32191e-05
```

Question 2 (60 points)

- a) (5 points) What is the equation of the separating hyperplane? Please state the coefficients up to seven decimal places.

Answer: Equation of the separating hyperplane is:

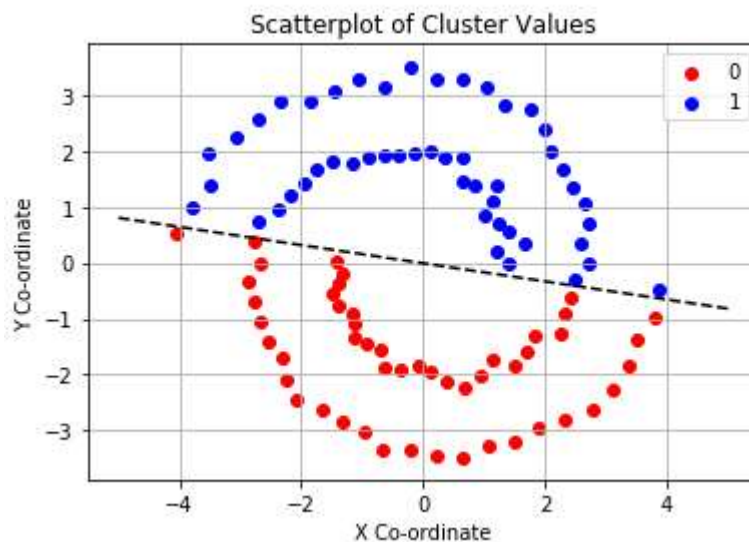
$$0.003345 + (0.0533351) X + (0.3286838) Y = 0$$

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

Answer: The misclassification rate is: 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.

Answer:



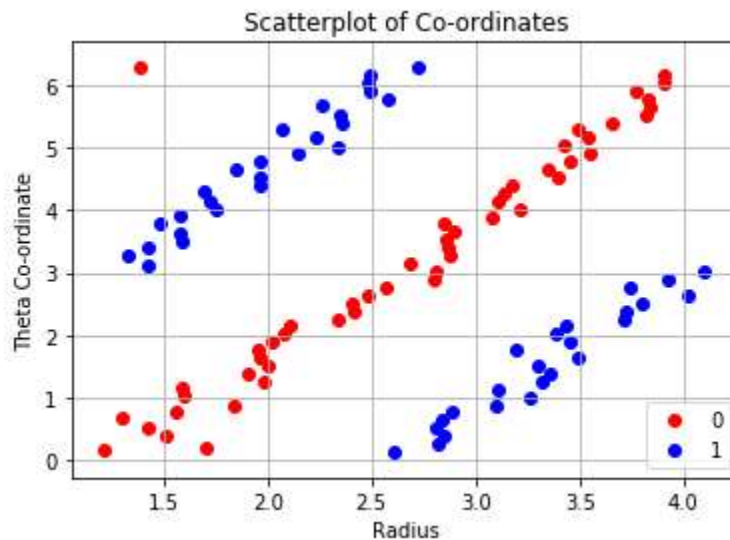
- d) (10 points) Please express the data as polar coordinates. Please plot the theta-coordinate against the radius-coordinate in a scatterplot. Please color-code the points using the SpectralCluster variable (0 = Red and 1 = Blue). To obtain the full credits, you should properly label the axes, the legend,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

Answer:

	radius	theta	class
0	1.390459	6.270135	0
1	1.577231	3.907981	1
2	1.211776	0.170155	0
3	1.702453	0.206176	0
4	1.967712	4.781499	1

- 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.

Answer:



- 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.

Answer:

Equation of the hypercurve for SVM 0 is

$$1.4691250777389275 + (0.933784147085173) X + (-0.45380248720594835) Y = 0$$

Equation of the hypercurve for SVM 1 is

$$-0.8768942577875997 + (0.933784147085173) X + (-0.45380248720594835) Y = 0$$

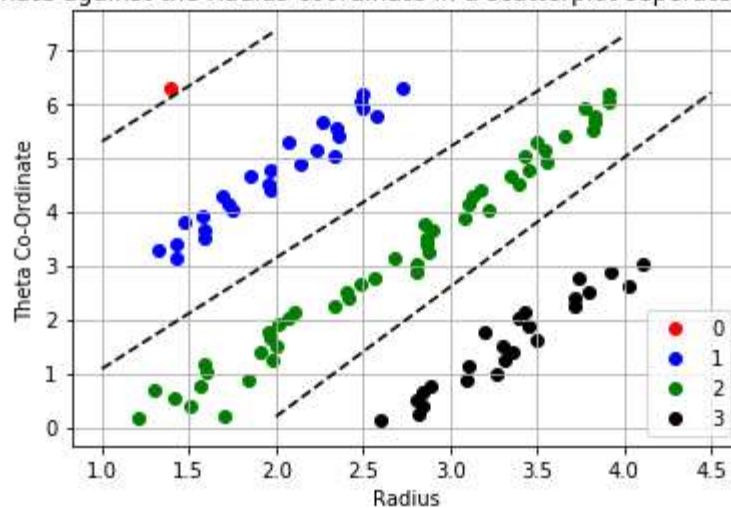
Equation of the hypercurve for SVM 2 is

$$-4.132844878075682 + (1.8920953263166829) X + (-0.8961324867551237) Y = 0$$

- 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.

Answer:

Theta-coordinate against the Radius-coordinate in a scatterplot seperated by 3 hyperplanes



- 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?

Answer:

