Carl Cortez
CIS 735
May 17th 2023

CIS 735 – Assignment 2

This assignment will be due by midnight on the day of live session 7

1. (50 pts) Using the two classes of data: dataSet1a.csv (Class a) and dataSet1b.csv (Class b) from assignment 1, classify the point (3,3) as being in either Class a or Class b using nearest means classifier (same one from the last assignment), and a develop a k-nearest neighbor classifier selecting k=3. You can use any Minkowski distance function you want.

Work in google sheet found here.

Summarize:		
Manhattan for (3,3)	Euclidean for (3,3)	Mahalanobis for (3,3)
class b	class b	class a

a. How does this change the results of your answer?
Class B again.

b. How does changing k={1, 5, 7, 11} affect the performance of the classifier?
Class B the whole way through.

c. What are the advantages/disadvantages of increasing the k hyperparameter for k-nearest neighbor?
Outliers can be overlooked depending on the hyperparameter used. I imagine if we used larger k numbers we may get pushed into a different class respectively.

Carl Cortez
CIS 735
May 17th 2023

2. (50 pts) Go to the website https://playground.tensorflow.org/. Note: if you want to use figures or screen shots to explain your answers you are more than welcome to.

a. Selecting the circle data set on the far right and using the default parameters, how epochs did it take for the neural network to make a "correct" model for the data?
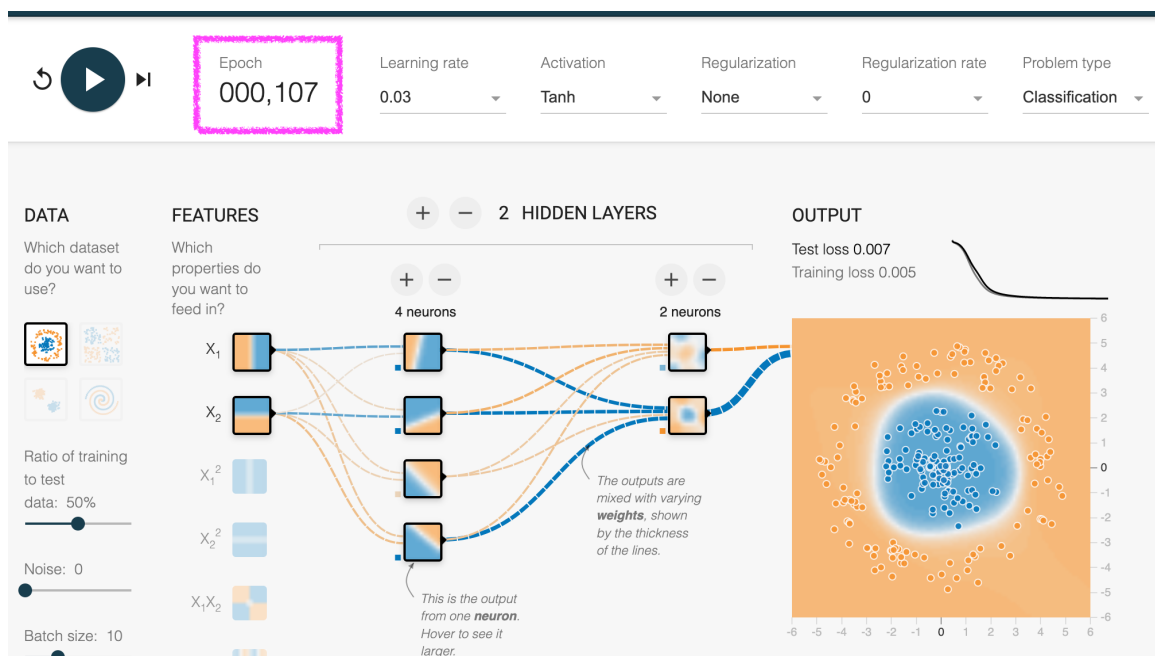
# of iterations

Steep of step

Activation function~ determines on or off

Reg~ fit the line. L1 is linear l2 is quadratic.
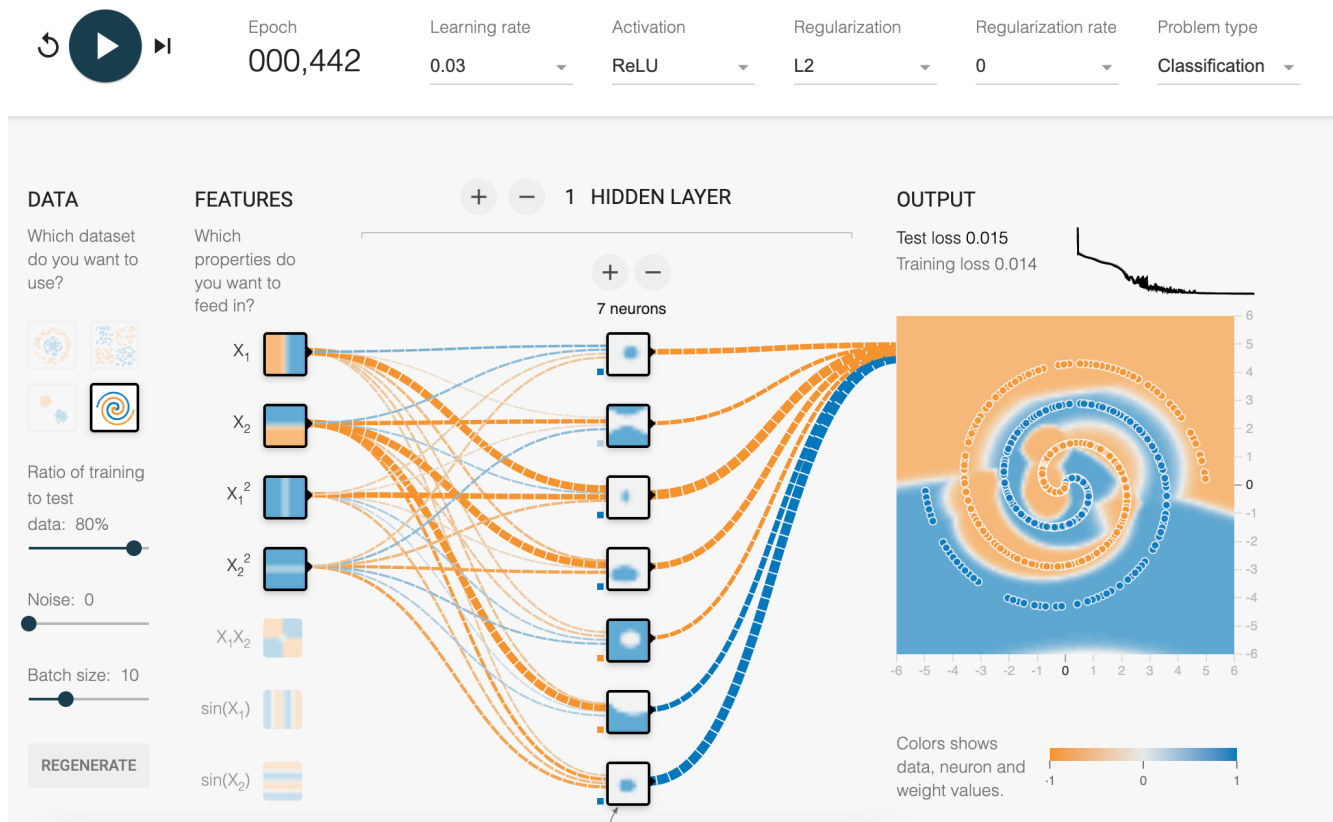
More complexion but takes longer.

Carl Cortez
CIS 735
May 17th 2023

b. Could this data set easily be classified using a nearest neighbor classifier? Why or why not? Compare and contrast with what would be better.

It would be a bit of a challenge since it may pick up on being split between orange in blue. In particular, the middle of the swirls could interfere and make it go in one classification or another.

c. Could this data set easily be classified using a support vector machine? Why or why not? Compare and contrast with what would be better.

I think the support vector machine could easily draw a circle around the middle blue points; sectioning it away from the orange. This could work well in this instance since the points are bunched together already in their respective areas.

d. In https://playground.tensorflow.org/ select the spiral class under Data label (bottom right). The default parameters will not be able to come up a with a good model for the data, try to come up with a model that can correctly classify the data set or at least minimize the loss. What hyperparameters did you end up selecting? Why do you think they helped improve the model?

Carl Cortez
CIS 735
May 17th 2023

Epoch
000,442

Learning rate
0.03

Activation
ReLU

Regularization
L2

Regularization rate
0

Problem type
Classification

DATA

Which dataset do you want to use?

Ratio of training to test data: 80%

Noise: 0

Batch size: 10

REGENERATE

FEATURES

Which properties do you want to feed in?

X_1

X_2

X_1^2

X_2^2

X_1X_2

$sin(X_1)$

$sin(X_2)$

+ − 1 HIDDEN LAYER

+ −

7 neurons

OUTPUT

Test loss 0.015
Training loss 0.014

Colors shows data, neuron and weight values.

-1   0   1

The 4 features helped take a vertical and horizontal shape; this was pivotal. L2 for regulation makes more of a quadratic or curve. The goal is to make the orange and blue curve into itself.

f. What are the benefits of adding more nodes to a neural network model? What are the costs? What are the benefits of adding additional hidden layers to a neural network? What are the costs? *$$$ = nodes. Away from the cloud and onto edge devices.*

Adding more nodes would cost more money.
Hidden layers may help get outputs we desire without revealing too much as to how it's done.

Carl Cortez
CIS 735
May 17th 2023

Per Professor Wagner, it's possible that we could move away from cloud technologies and into edge devices. I see this is a less expensive option that could work better in time. Assuming all of the devices are up to date and have a secure interface, it could work.