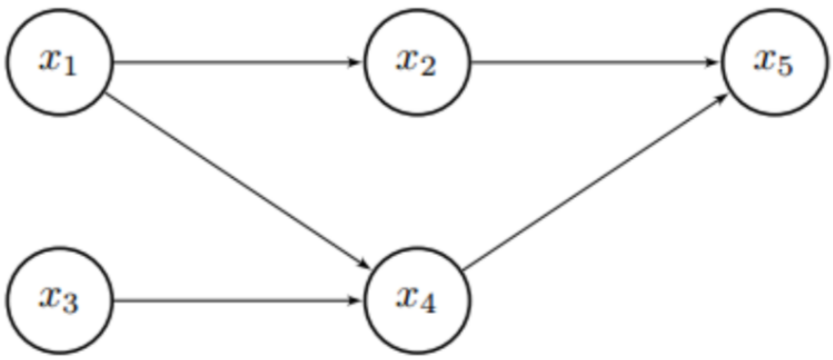
Homework 5

Yunian Pan

December 8, 2018

1 Problem 1: Bayesian Network Conditional Independence



Below is the factorization of the probability distribution $p(x_1, \dots, x_5)$implied by this directed graph:

$$p(x_1, \dots, x_5) = p(x_1)p(x_3)p(x_2|x_1)p(x_4|x_1, x_3)p(x_5|x_2, x_4)$$

Using the Bayes ball algorithm, the statements

1. x_2 and x_4 are independent.

   False, the path going through between x_2 and x_4 is $x_2 \to x_1 \to x_4$.

2. x_2 and x_4 are conditionally independent given x_1, x_3, and x_5.

   False, the path going through between x_2 and x_4 is $x_2 \to x_5 \to x_4$.

3. x_2 and x_4 are conditionally independent given x_1, x_3.

   True, there's no path going through between x_2 and x_4 given x_1, x_3.

4. x_5 and x_3 are conditionally independent given x_4,

   False, the path is $x_5 \to x_2 \to x_1 \to x_4 \to x_3$.

5. x_5 and x_3 are conditionally independent given x_1, x_2, and x_4.

   True, there's no path going through between x_5 and x_3 given x_1, x_2 and x_4.

6. x_1 and x_3 are conditionally independent given x_5.

   False, the path is $x_1 \to x_2 \to x_5 \to x_4 \to x_3$.

7. x_1 and x_3 are conditionally independent given x_2.

   True, there's no path going through between x_1 and x_3 given x_2.

8. x_2 and x_3 are independent.

   True, there's no path going through between x_1 and x_3.

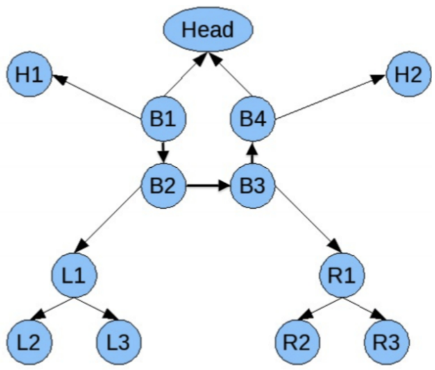
9. x_2 and x_3 are conditionally independent given x_5.

   False, the path is $x_2 \to x_5 \to x_4 \to x_3$.

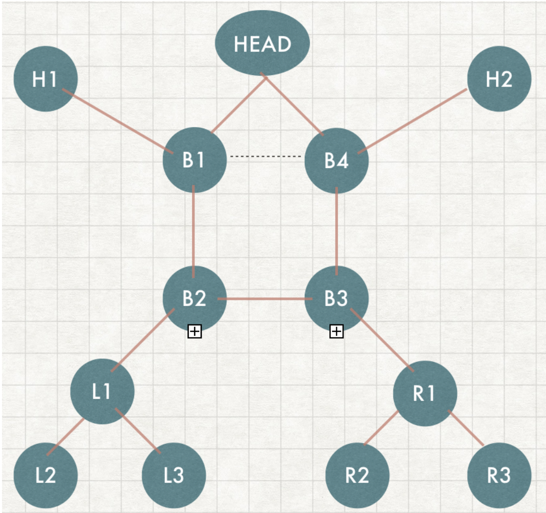
10. x_1 and x_3 are conditionally independent given x_5 and x_4.

    False, the path is $x_2 \to x_1 \to x_4 \to x_3$.

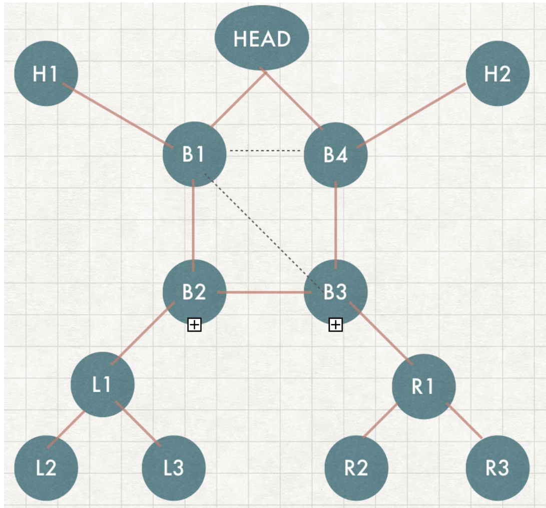
2 Problem 2: Junction Tree



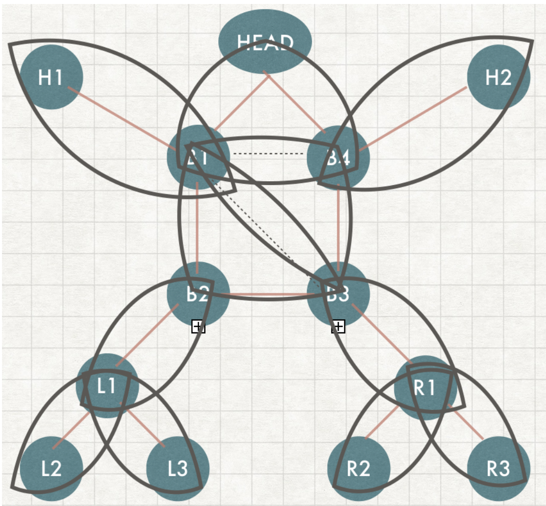
After moralization and triangulation:
Build a junction tree:



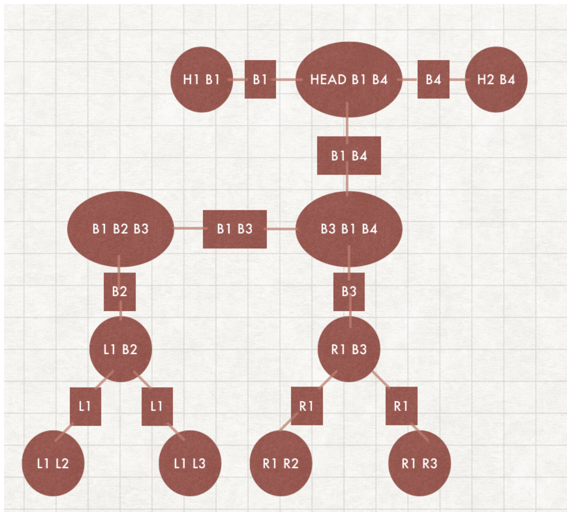
(a) moralization



(b) triangulation



(c) find cycle



(d) junction tree

3 Problem 3: Neural Networks

For dataset 1, I consider it to be mapped to a linearly separable set whose features are the norms of each data point $X_1^2 + X_2^2$, so the simplest way is to design a single neuron, it can be a SVM or a perceptron or whatever, as shown in 3.1, I use the ReLu function with no regularization, feed in the X_1^2 and X_2^2 properties, it can be done with test loss and training loss both being 0.001. But with a hidden layer

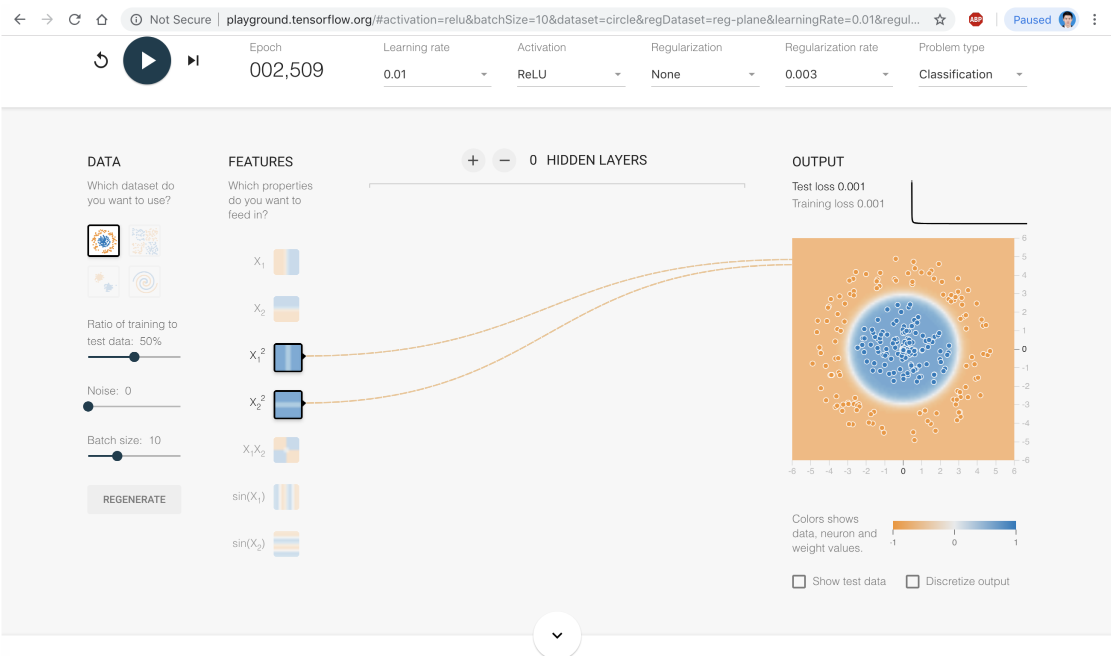


Figure 3.1: dataset 1

and linear activate function as shown in 3.2 the test loss and training loss both reached 0 after 1300 epochs, which means no overfitting.

For dataset 2, we can use a sign classifier to separate $X_1X_2 > 0$ and $X_1X_2 < 0$, therefore I use the ReLu function again as a single neuron to establish the classifier, as shown in 3.3.

For dataset 3, the data points are similar to 2 different 2D-Gaussian distributions with 2 different centers, so I add a hidden layer and a L_2 regularization to model the properties of the data set, as shown in 3.4.

For dataset 4, it's not possible to design an extremely simple network to classify such a double helix data set, 2 hidden layers are used, along with L_2 regularization in case of overfitting and tahn activate function, as shown in 3.5

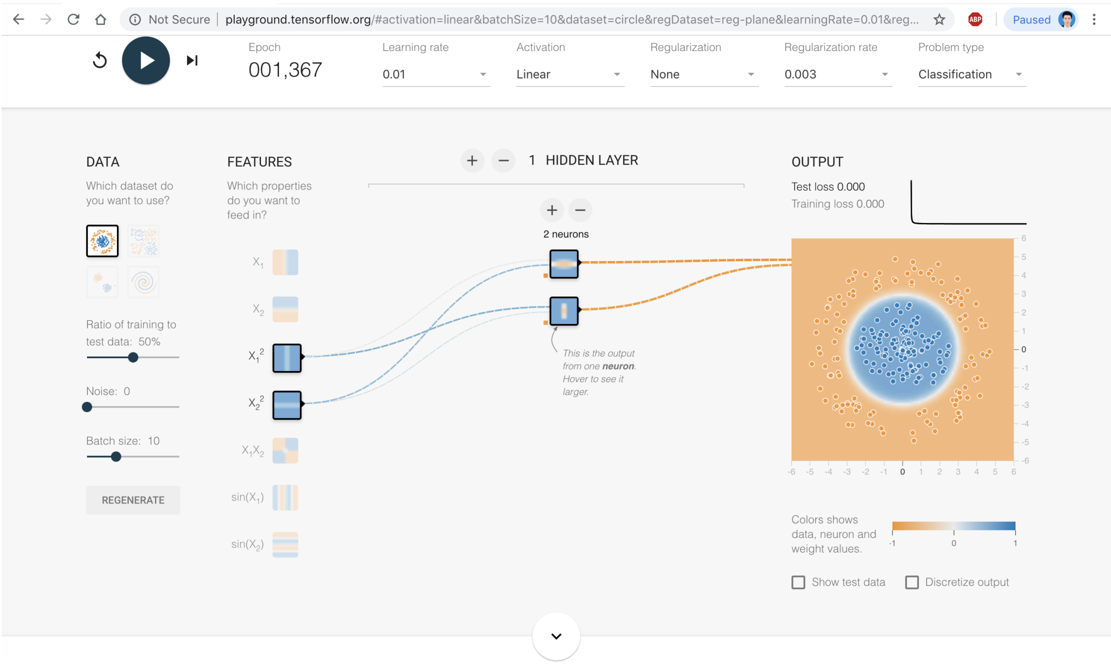


Figure 3.2: dataset 1(2)

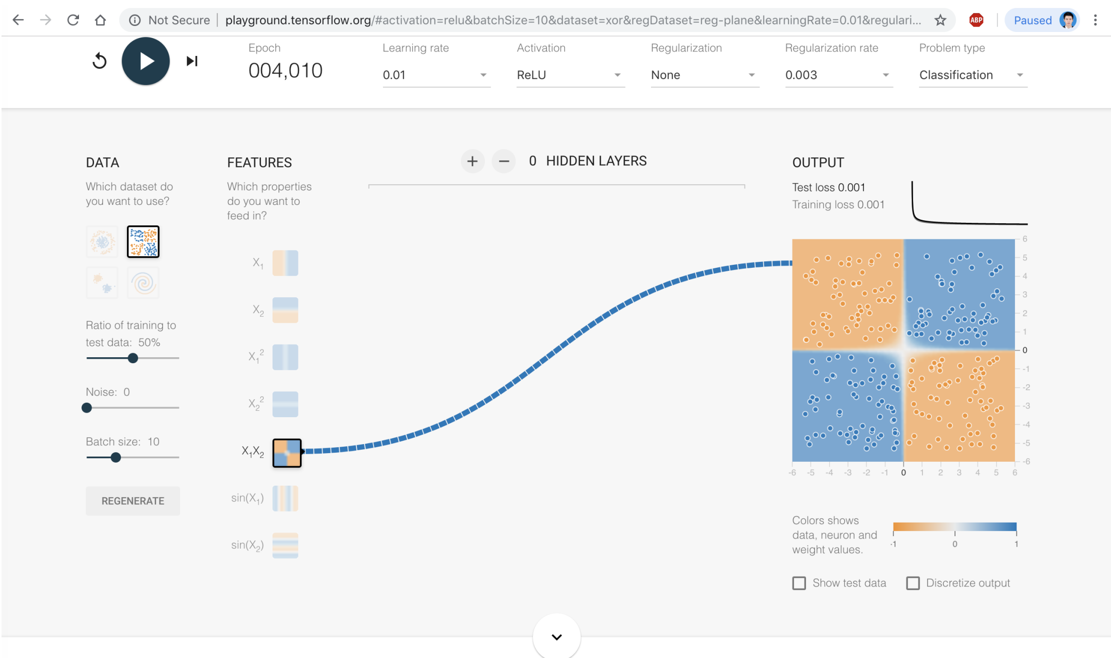


Figure 3.3: dataset 2

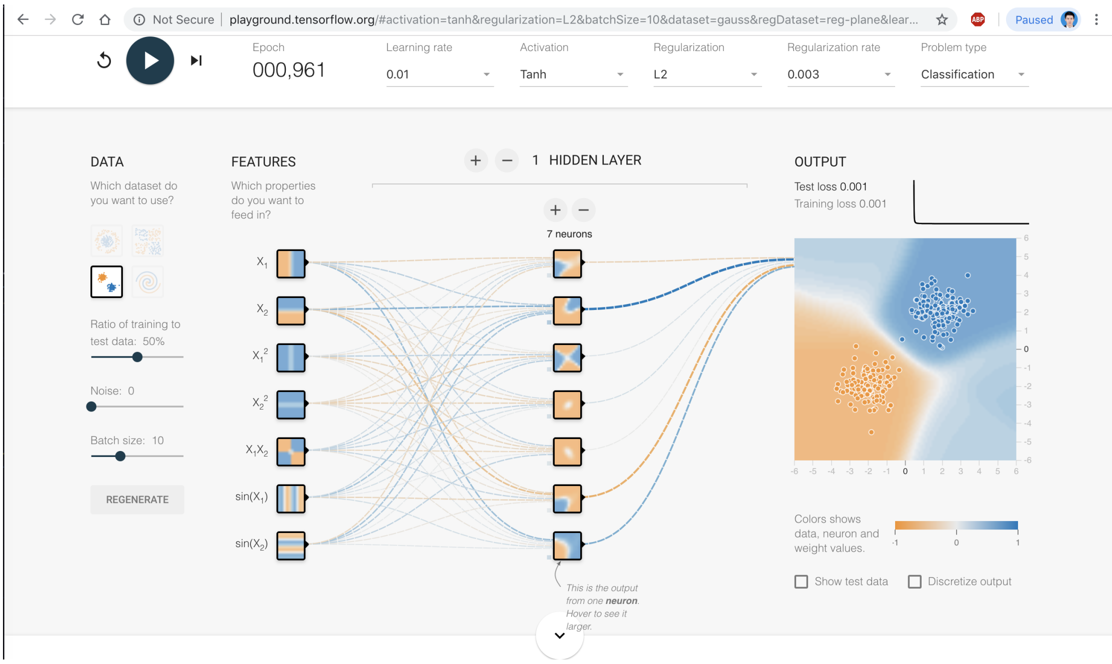


Figure 3.4: dataset 3

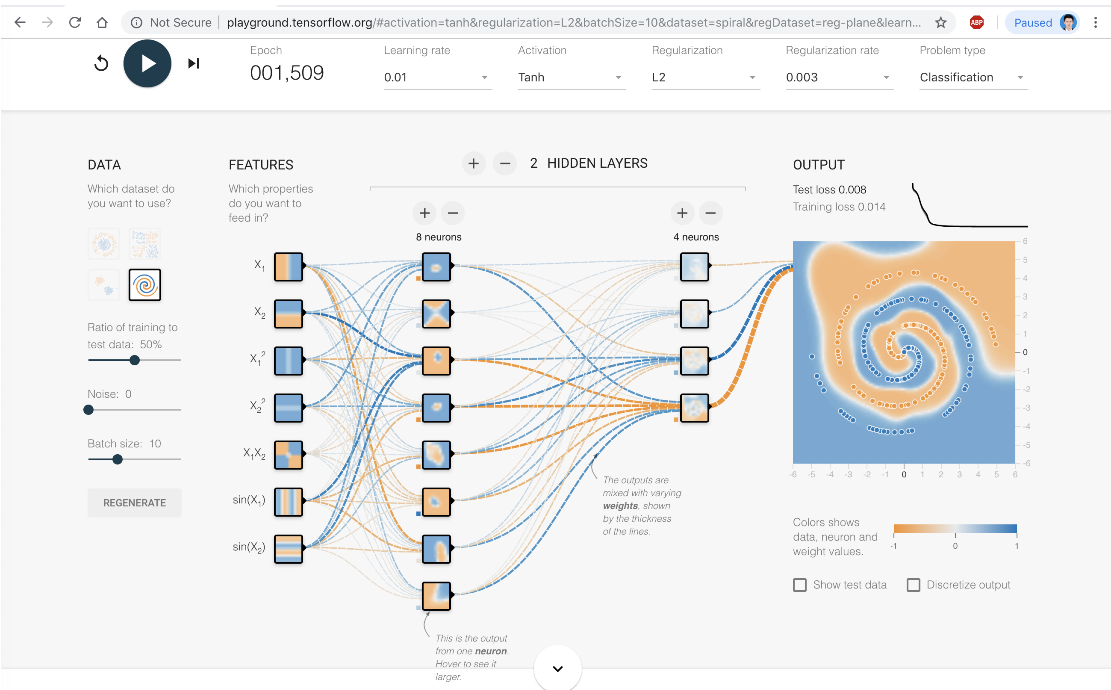


Figure 3.5: dataset 4