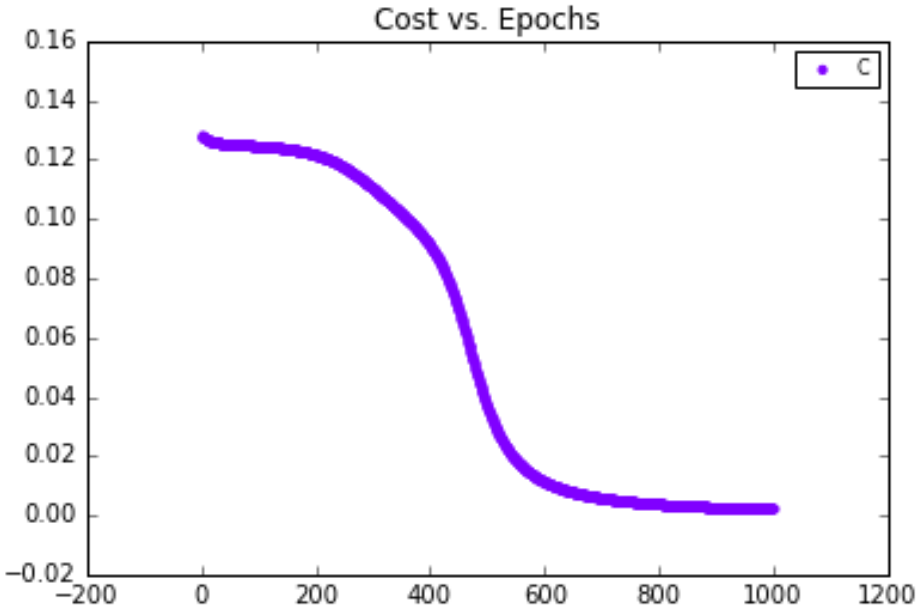
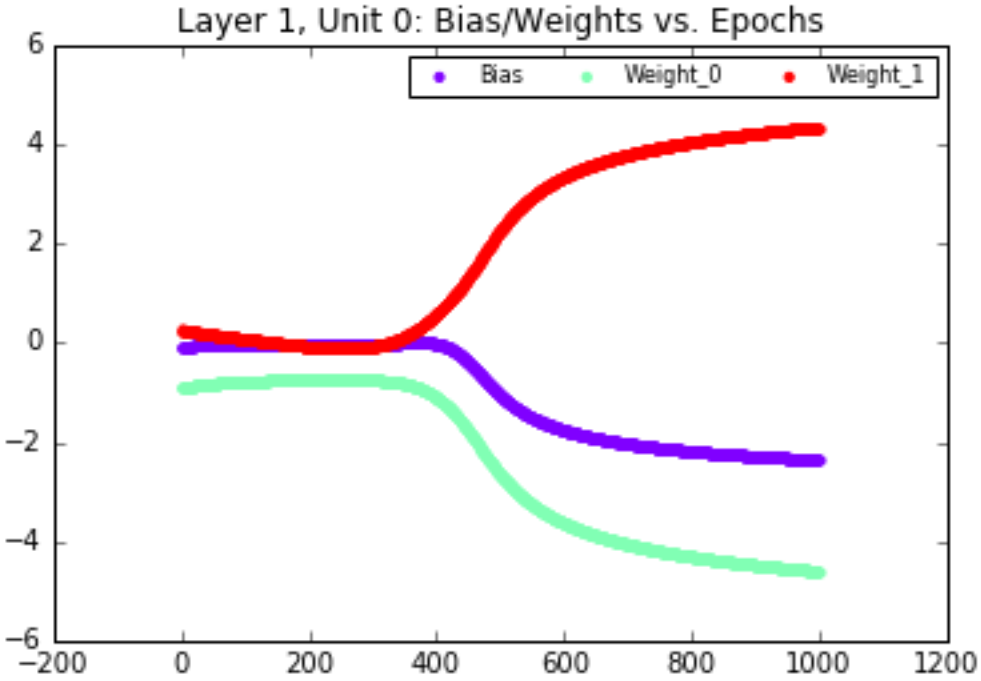
Adam Gross and Josh Palmer

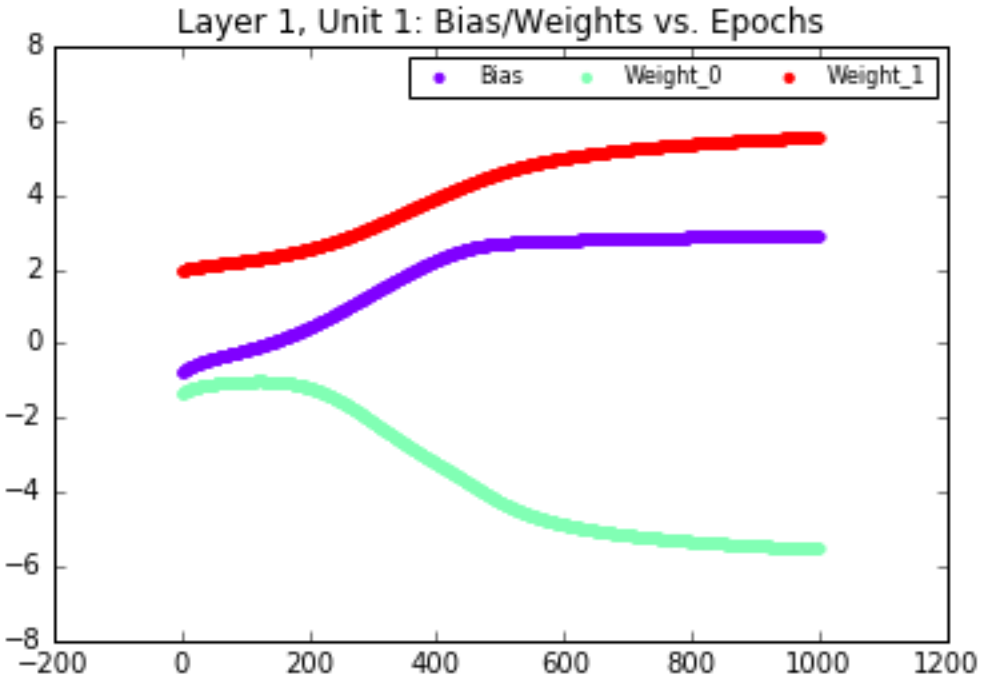
Assignment 5 Write-up

1. Tests on the XOR problem with one hidden layer with two units (topology = [2,2,1])



Hidden layer units' bias and weights changes:





The final cost was 0.00223. Thus, the network performed quite well at learning XOR, converging to this final cost by 1000 epochs. It also appears that the hidden layer weights have converged as well.

2. The table below shows the final costs for the different topologies we tested. Our minimal cost for a single hidden layer topology was with 4 hidden layer units (topology = [2,4,4]); for a double hidden layer network, the best topology was [2,2,4,4]:

Topology	**Final Cost**
[2, 1, 4]	0.12957
[2, 2, 4]	0.00322
[2, 3, 4]	0.00302
[2, 4, 4]	0.00265 *** Best for single hidden layer
[2, 1, 1, 4]	0.13444
[2, 1, 2, 4]	0.11083
[2, 1, 3, 4]	0.00694
[2, 1, 4, 4]	0.00387
[2, 2, 1, 4]	0.13343
[2, 2, 2, 4]	0.00326
[2, 2, 3, 4]	0.00373
[2, 2, 4, 4]	0.00199 *** Best for double hidden layer
[2, 3, 1, 4]	0.12998
[2, 3, 2, 4]	0.00367
[2, 3, 3, 4]	0.00255
[2, 3, 4, 4]	0.00355
[2, 4, 1, 4]	0.12590

[2, 4, 2, 4]	0.00436
[2, 4, 3, 4]	0.00242
[2, 4, 4, 4]	0.00225

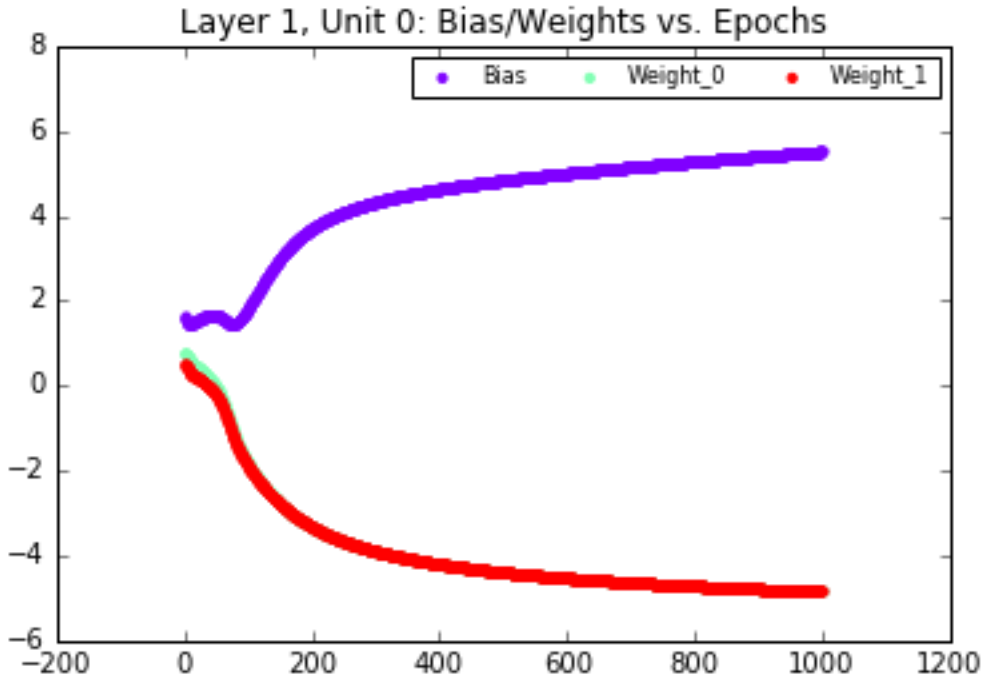
As expected, for single hidden layer networks, the topology [2, 1, 4] performed the worst, by a significant amount. Only having 1 hidden unit leads to a "loss of information" because two inputs are "squashed down" into one unit. The rest of the topologies [2, 2, 4], [2, 3, 4], [2, 4, 4] all had similar performances that were much better, with the latter having a slight performance edge (however such slight differences could be attributed to randomization in weight/bias initialization).

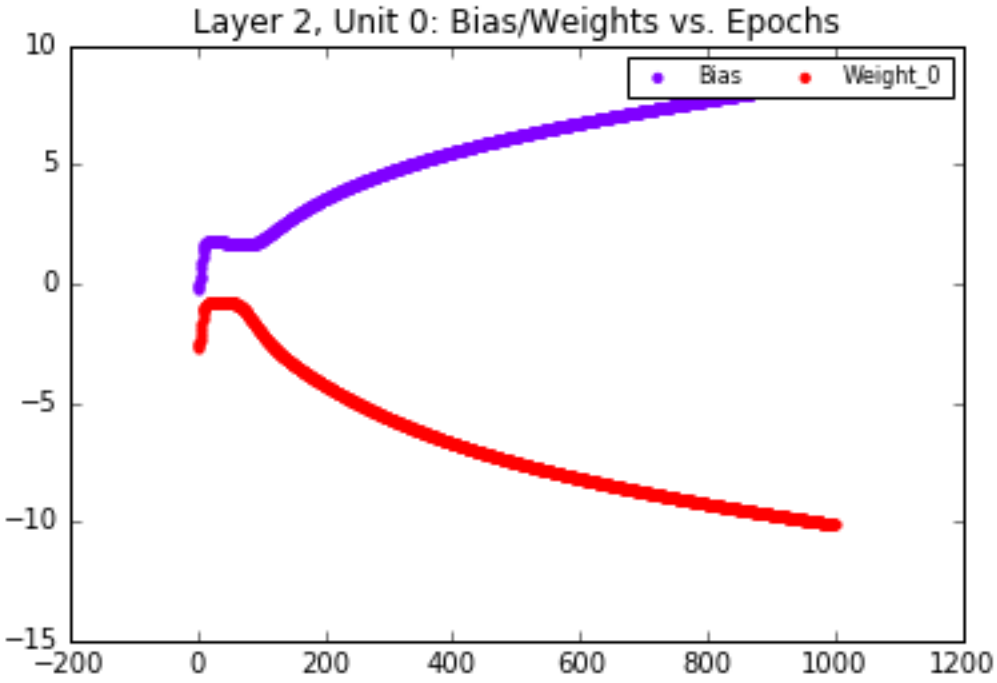
As for double hidden layer networks, the worst performance was seen when the second hidden layer (i.e. the one immediately before the output layer) only had one unit. Irrespective of the number of units in the first hidden layer, subsequently adding units to the second hidden layer improved performance dramatically. There are slight differences that can be chalked up to random weight/bias initialization; also there seems to bias towards consecutive layers having equivalent (or close to equivalent) numbers of units. For example, [2, 3, 3, 4] performed slightly better than [2, 3, 2, 4]. This makes intuitive sense because layer transitions that preserve sizes are more likely to preserve information. There is a satisfying symmetry to the winning topology: [2, 2, 4, 4]. The hidden layers match in unit size to the input and output layers respectively.

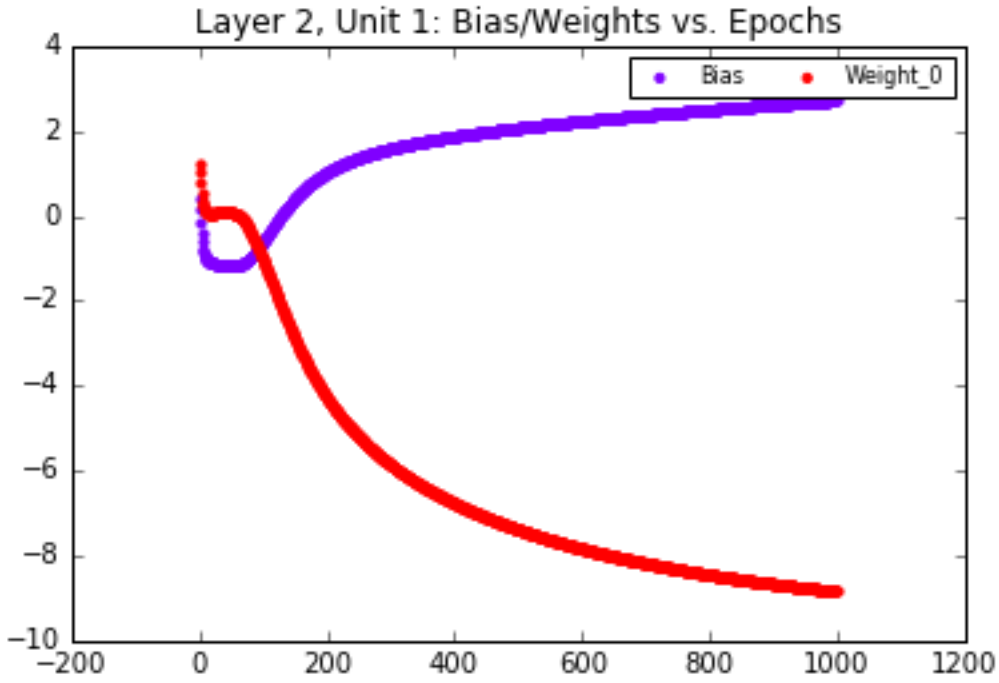
We inspected the weights of the winners ([2,4,4] on page 15 and [2,2,4,4] on page 53) to see if we could ascertain what was being learned in each layer. With [2,4,4], the units in the hidden layer alternate between sending both of its weights to positive and negative regimes (i.e. unit0 and unit2 weights converge to positive values whereas unit1 and unit3 weights converge to negative values). Similarly, with [2,2,4,4] the weights in the first hidden layer converge positively for unit0 and negatively for unit1. Then with the second hidden layer, the weights diverge, one going positive, the other negative. This is effectively a logical circuit propagating 0s and 1s to each appropriate logical gate (AND, OR, NAND, XOR). In the final layer, the weights are flipped for the AND and NAND gates as expected.

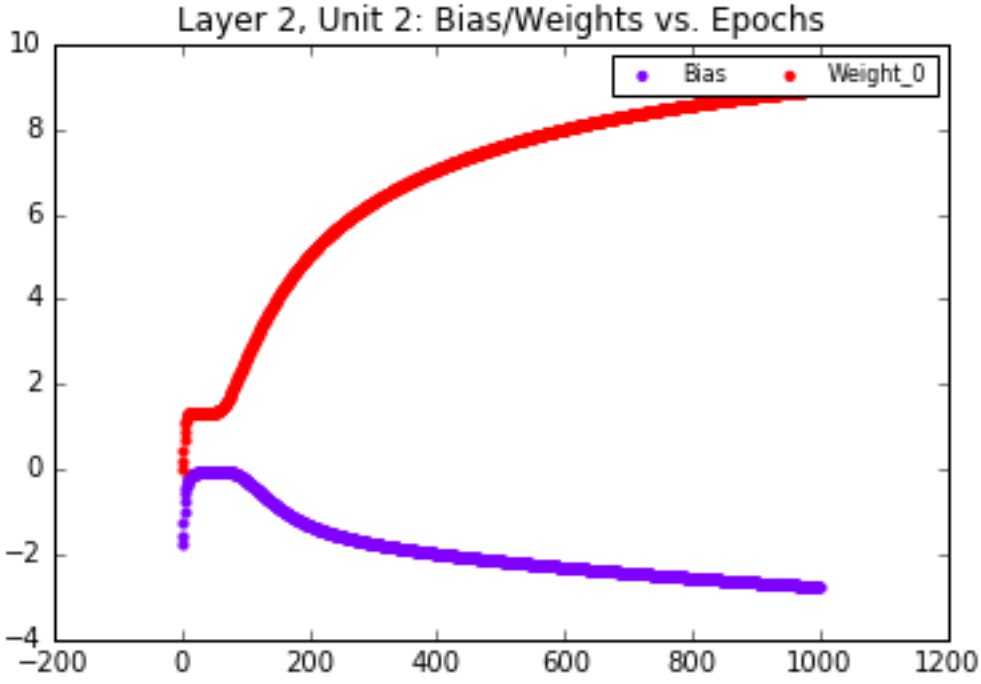
Plots of our different parameters tested are provided on the remaining pages. We had two input values and four outputs representing OR, AND, NAND, and XOR. The variables were the number of hidden layers and the number of units in those hidden layers. We performed tests with all combinations of 1, or 2 hidden layers and 1, 2, 3, or 4 hidden layer units. The double hidden layer winner.

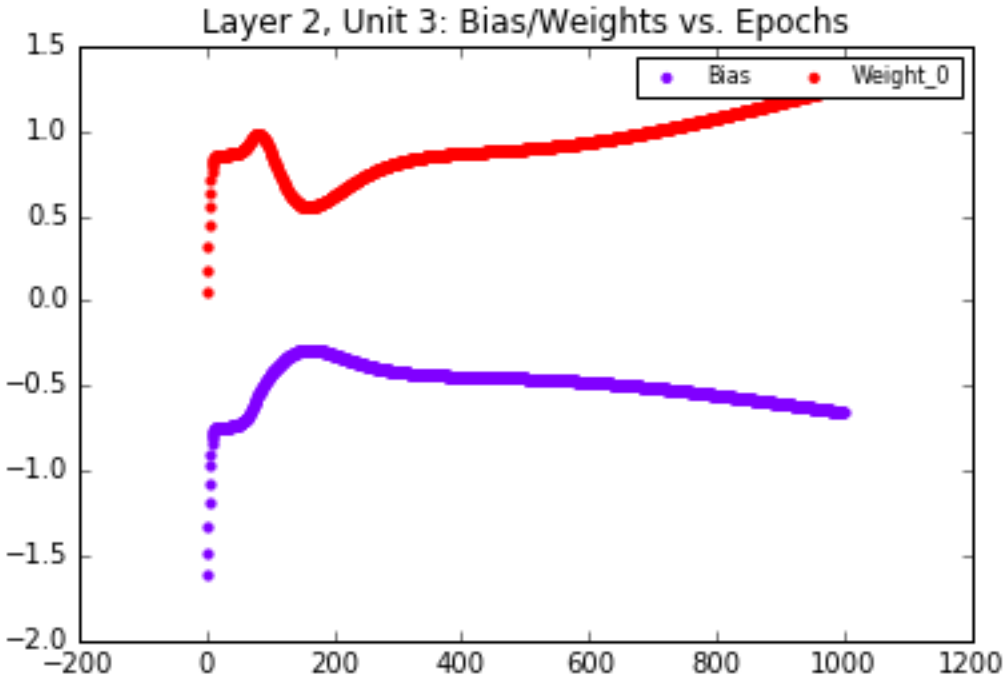
2 inputs, 1 hidden layer unit, 4 outputs

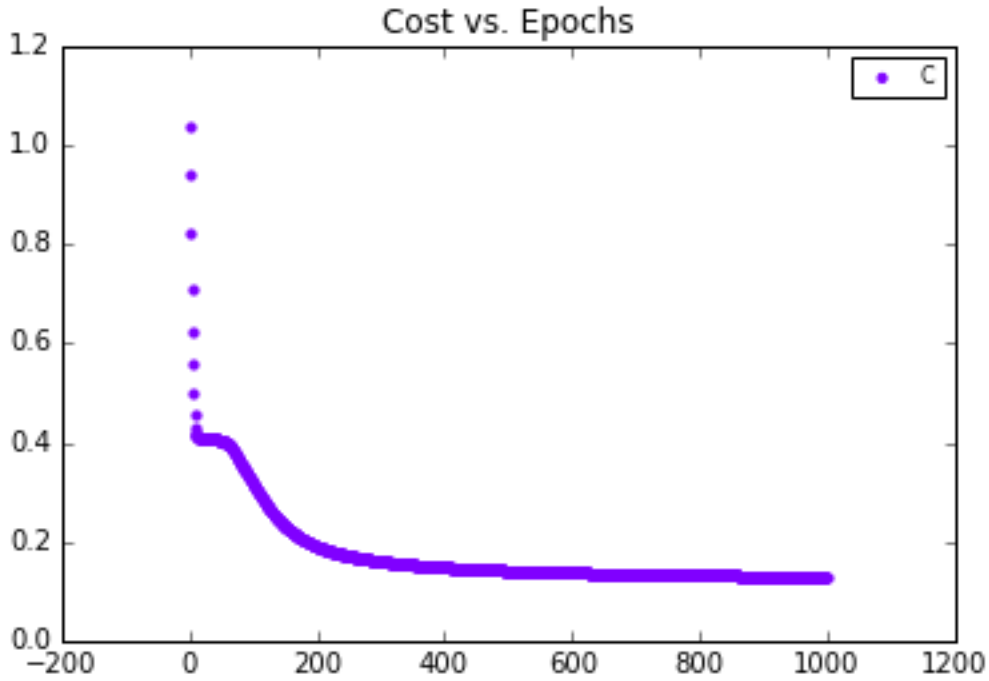

Layer 1, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200


Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
10
5
0
−5
−10
−15
−200
0
200
400
600
800
1000
1200

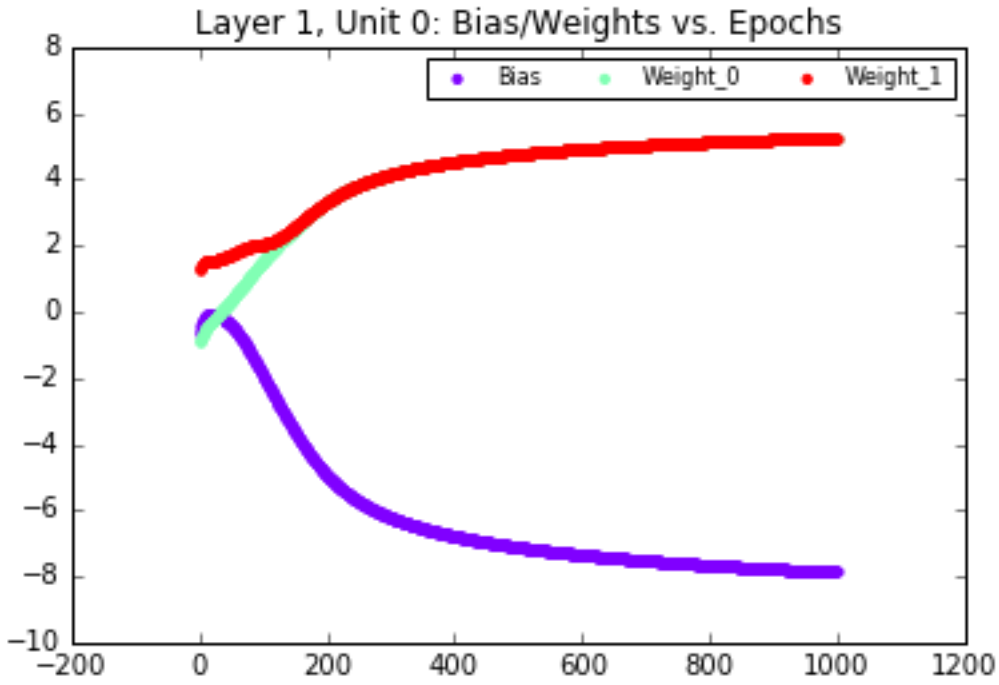

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

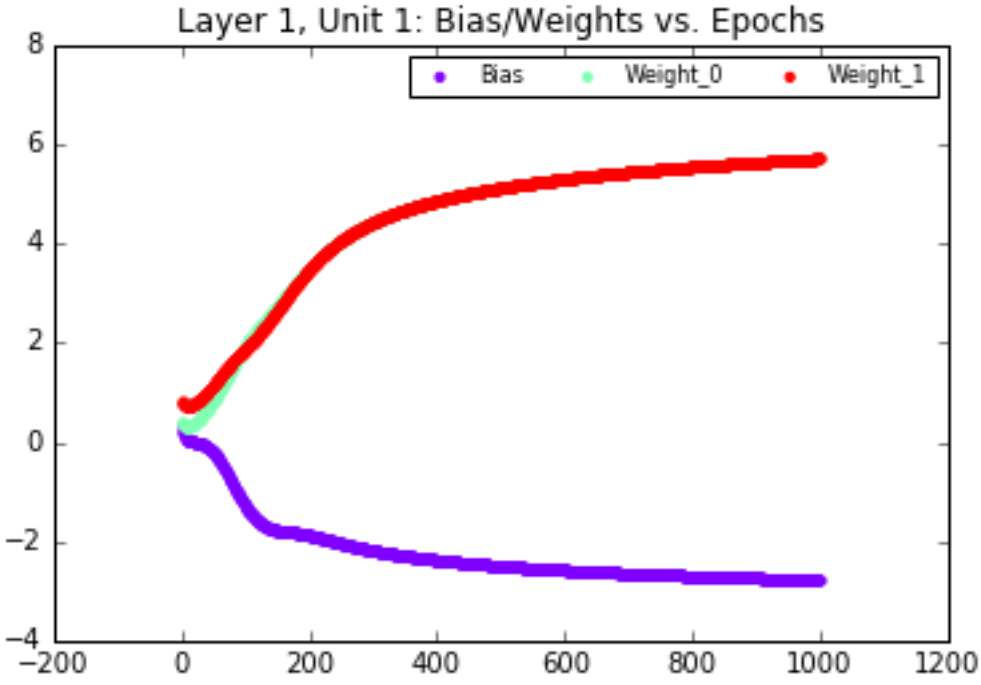

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
10
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

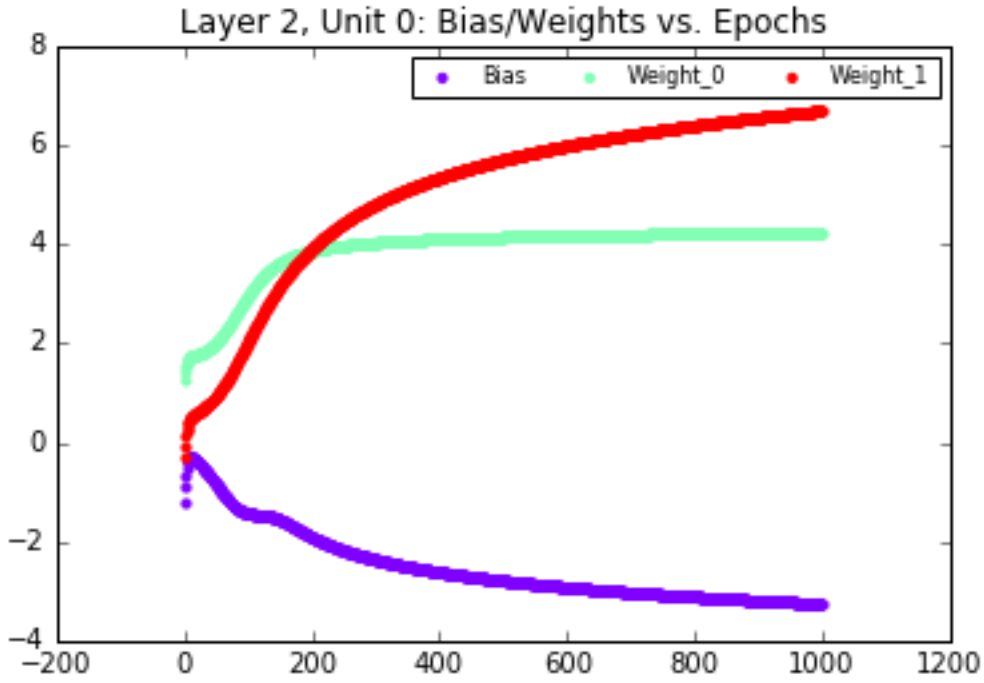

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−200
0
200
400
600
800
1000
1200

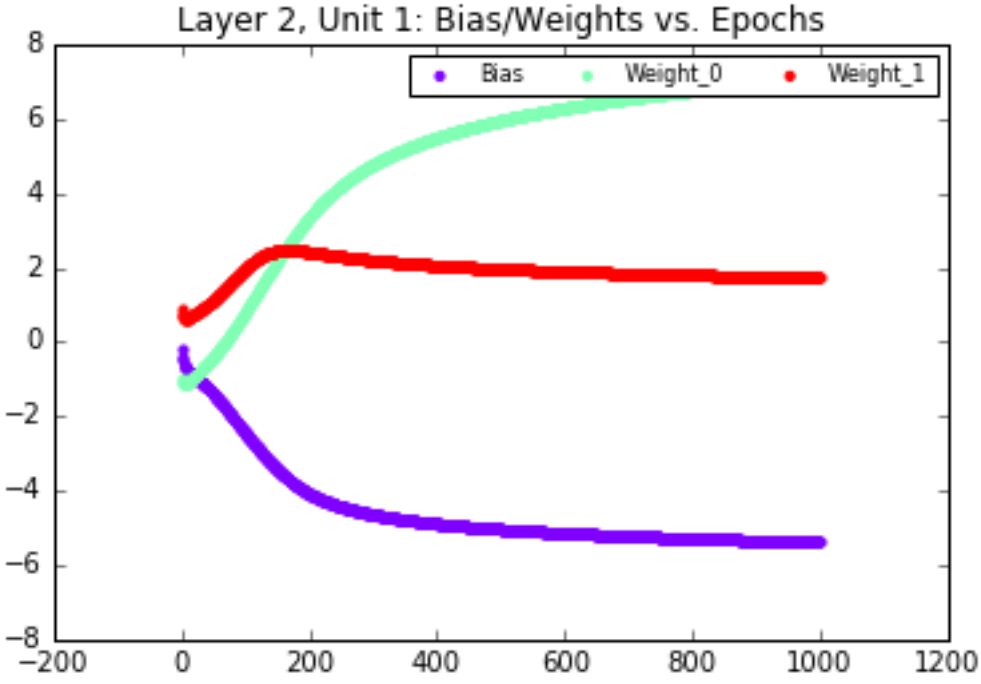

Cost vs. Epochs
C
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−200
0
200
400
600
800
1000
1200

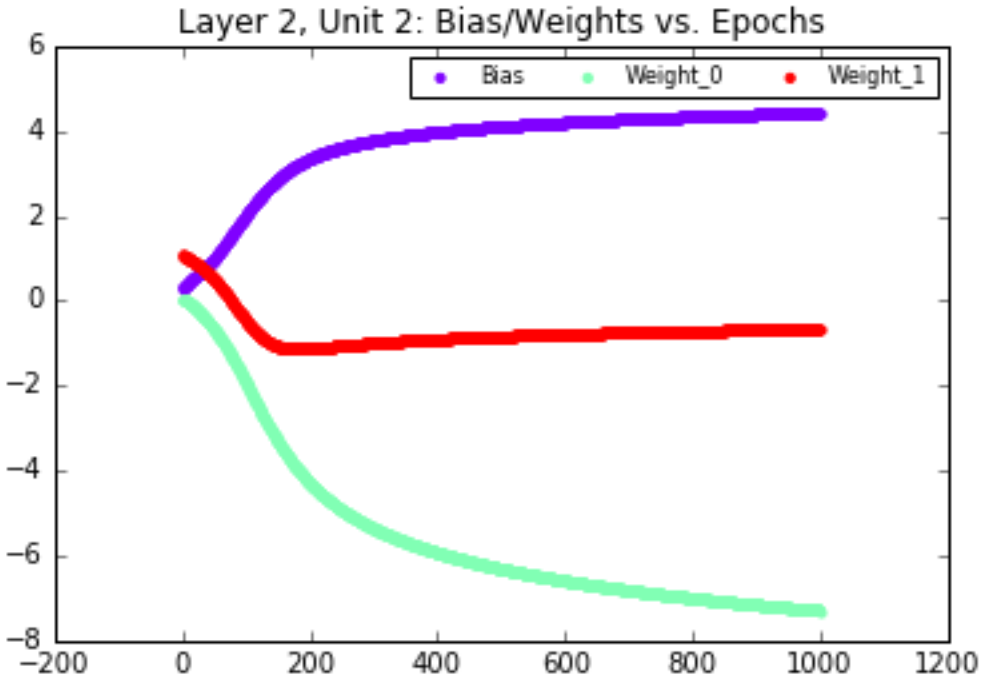
2 inputs, 2 hidden layer units, 4 outputs

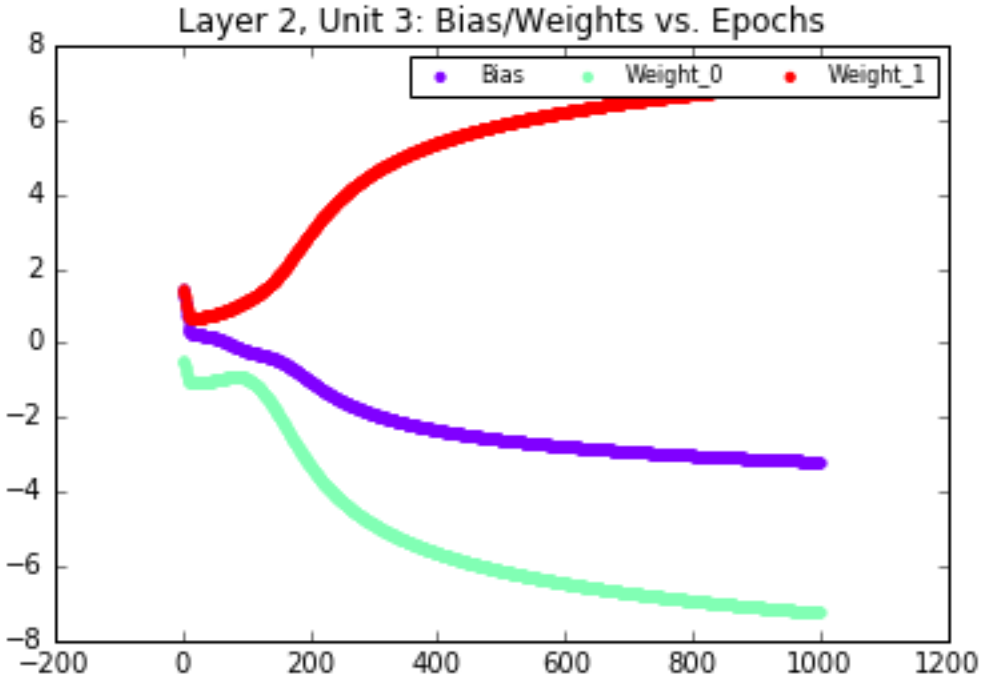


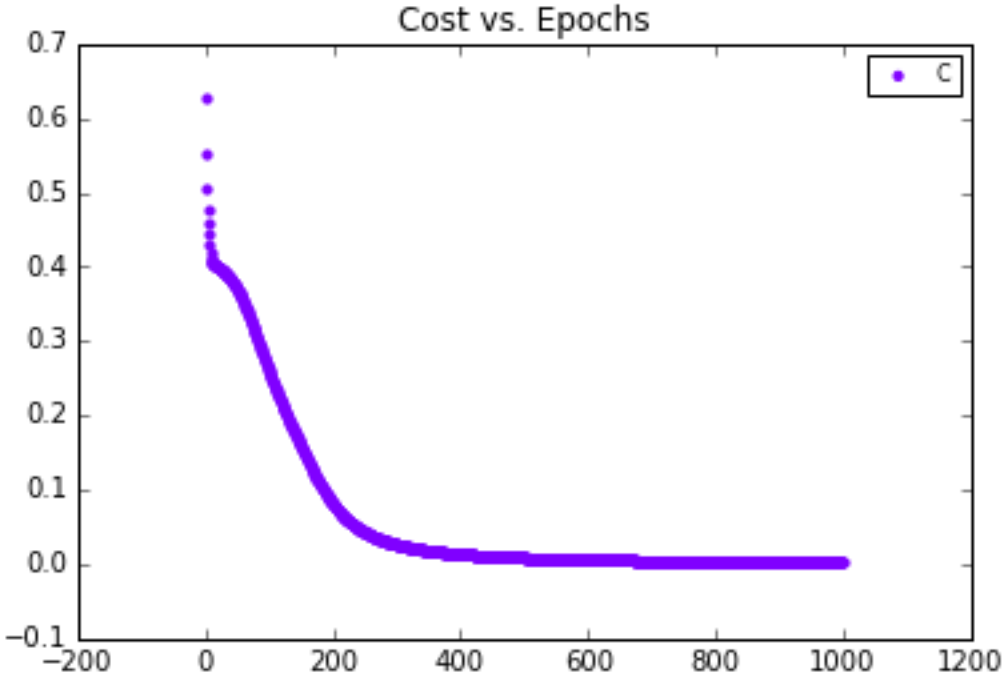



Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

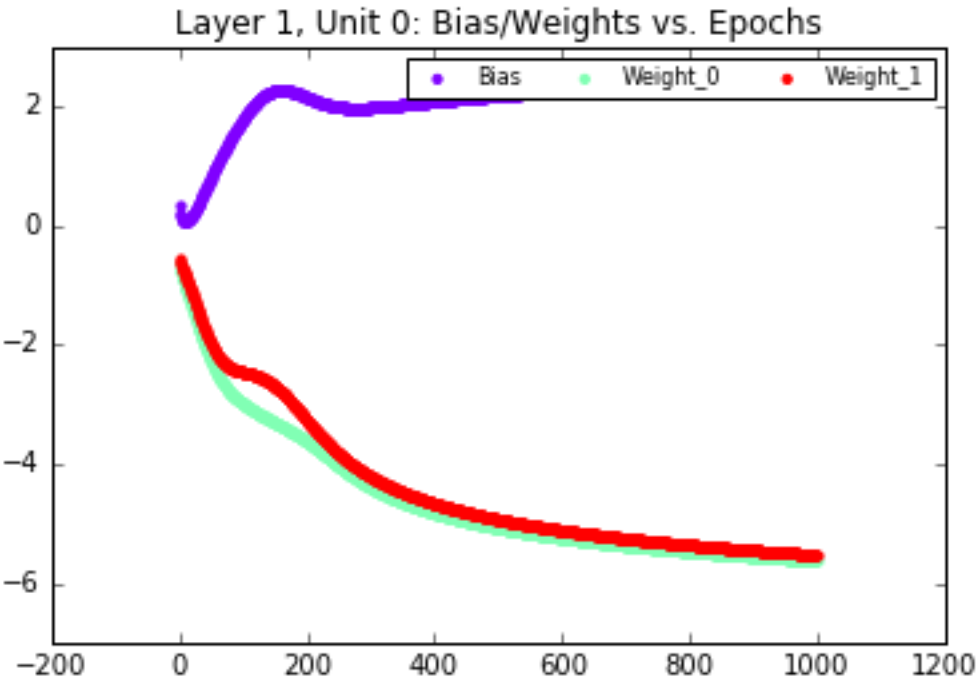

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

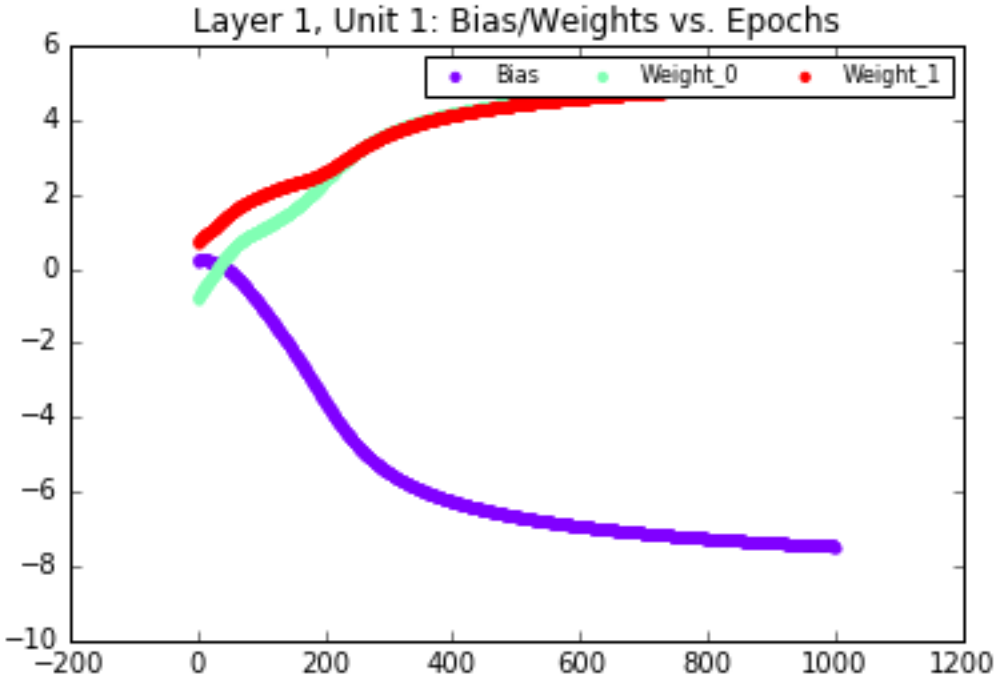

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

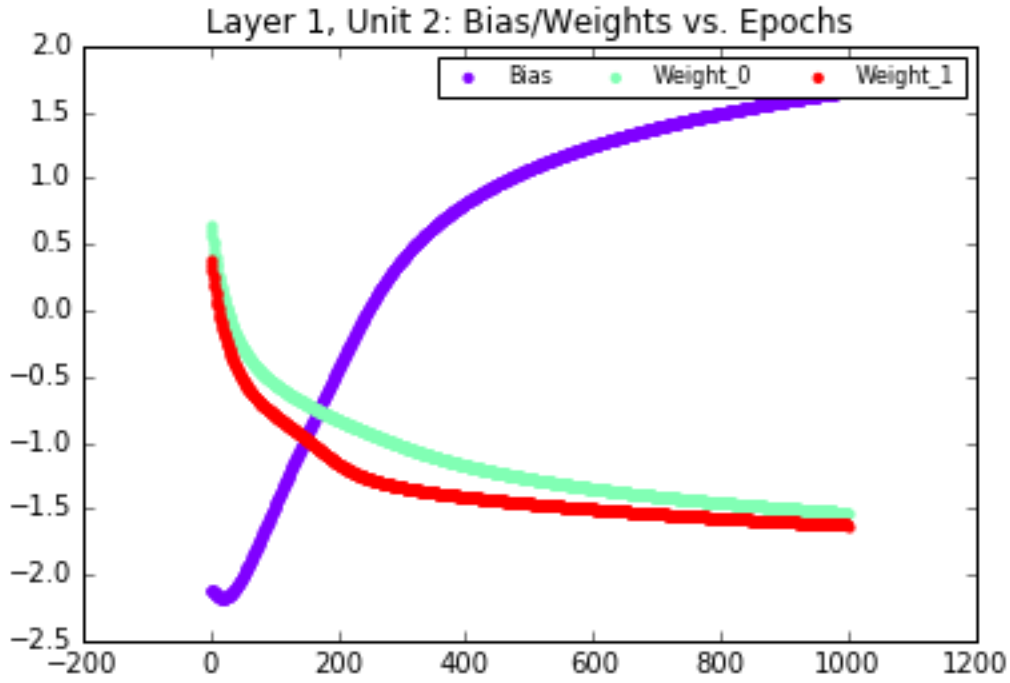

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

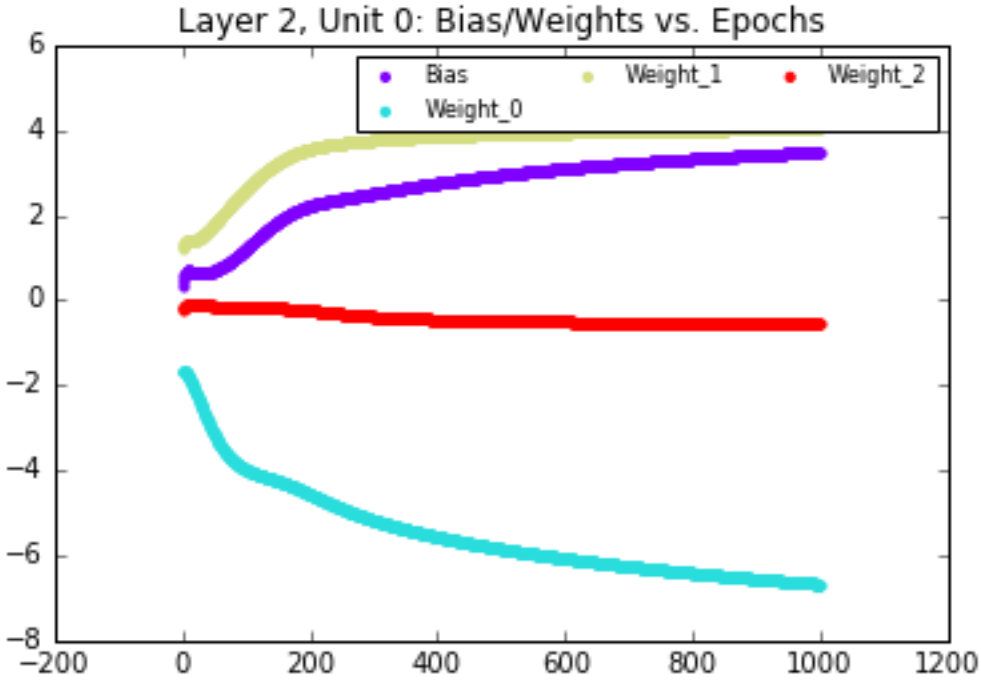

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

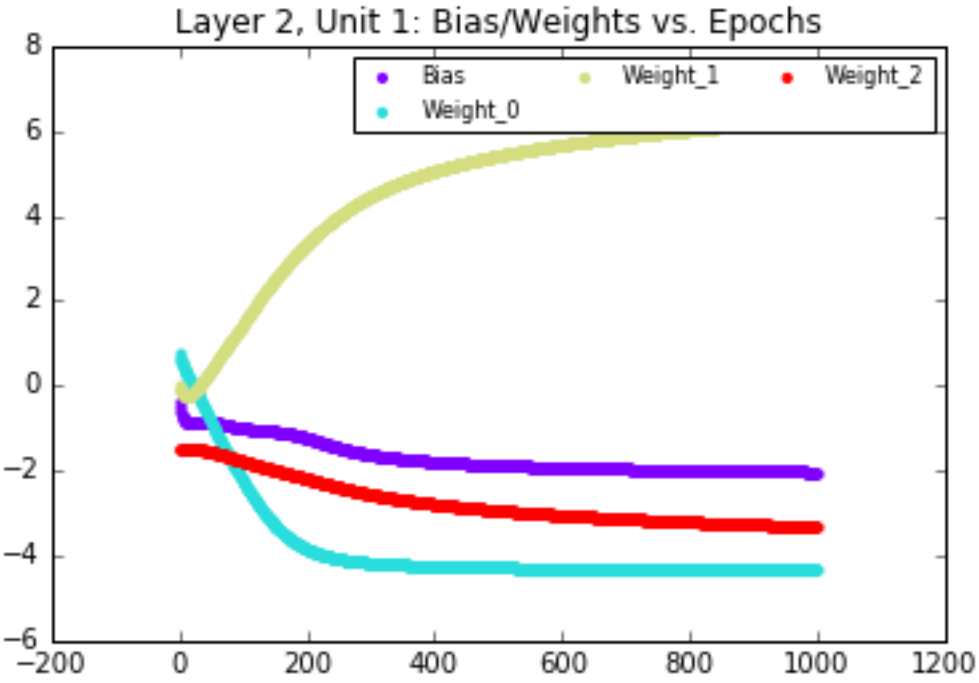
2 inputs, 3 hidden layer units, 4 outputs

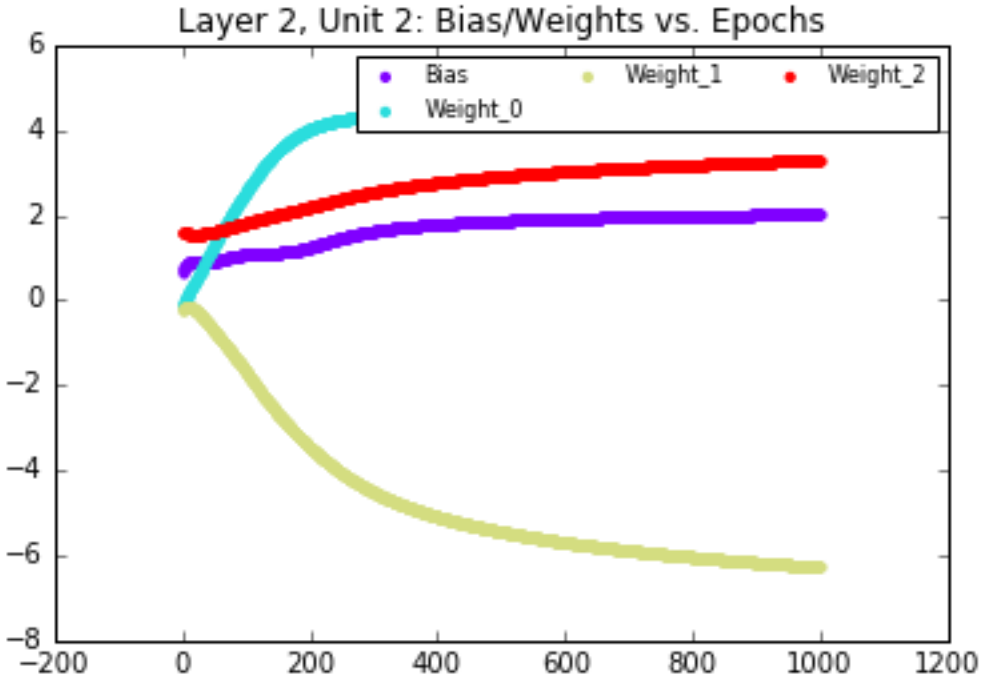


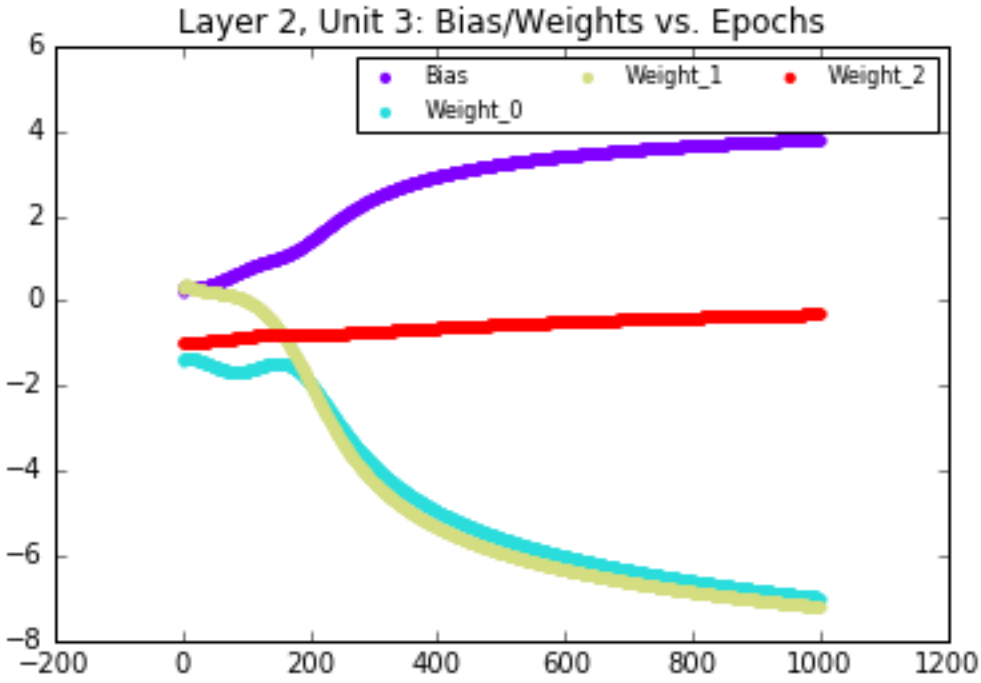


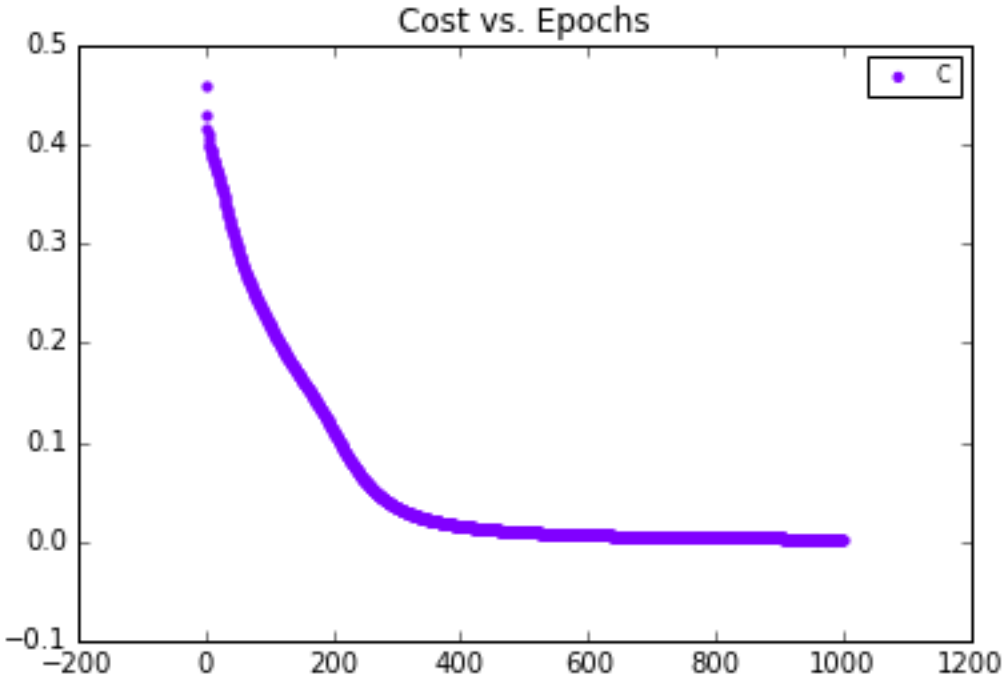

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−200
0
200
400
600
800
1000
1200


Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

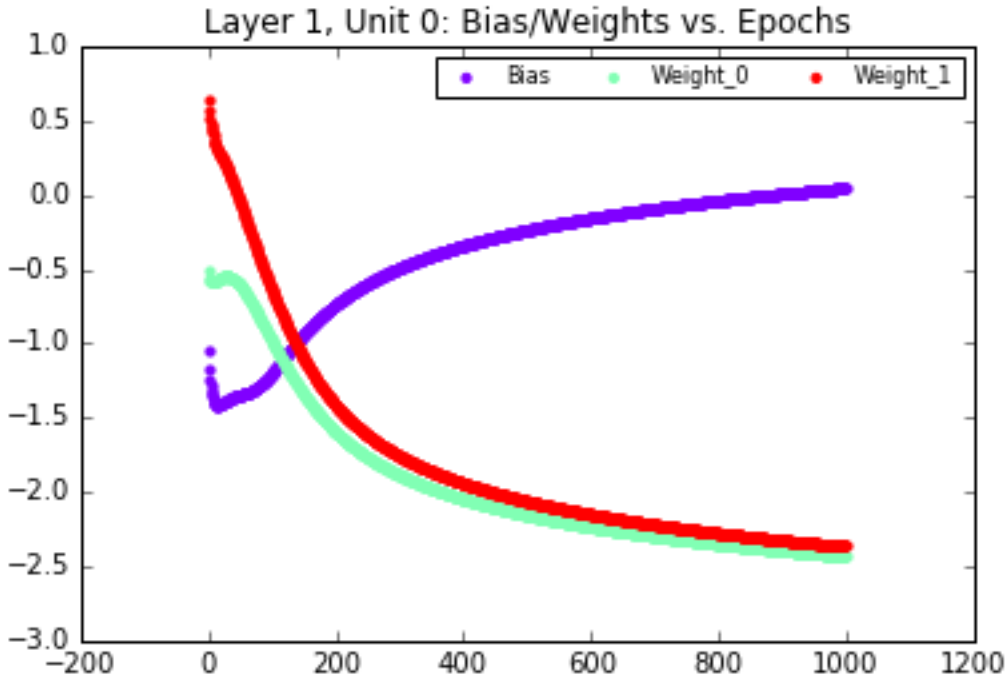

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

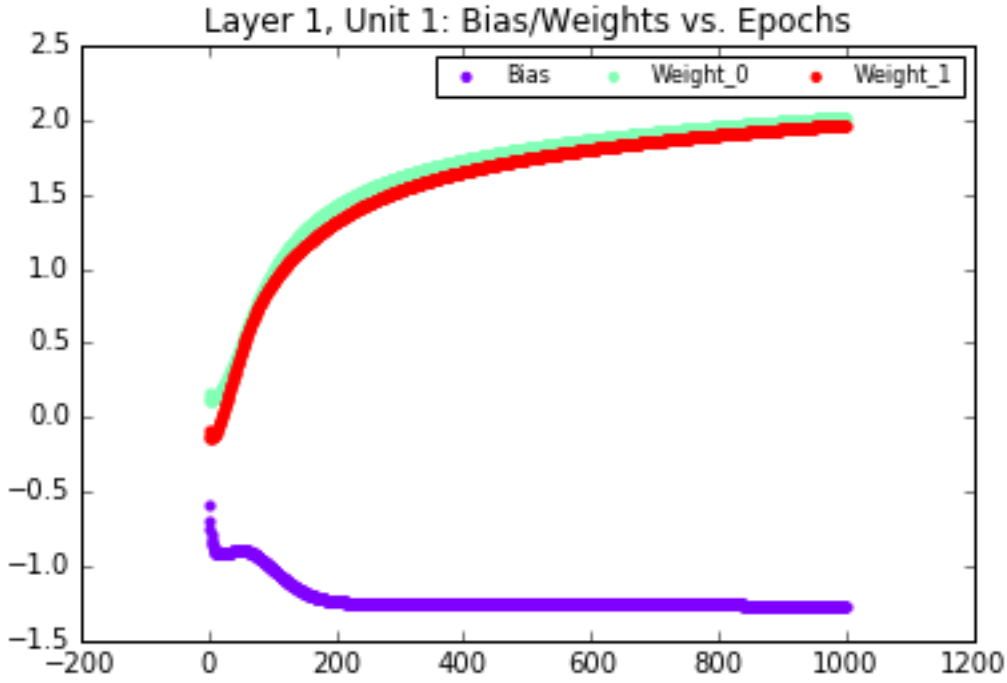

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

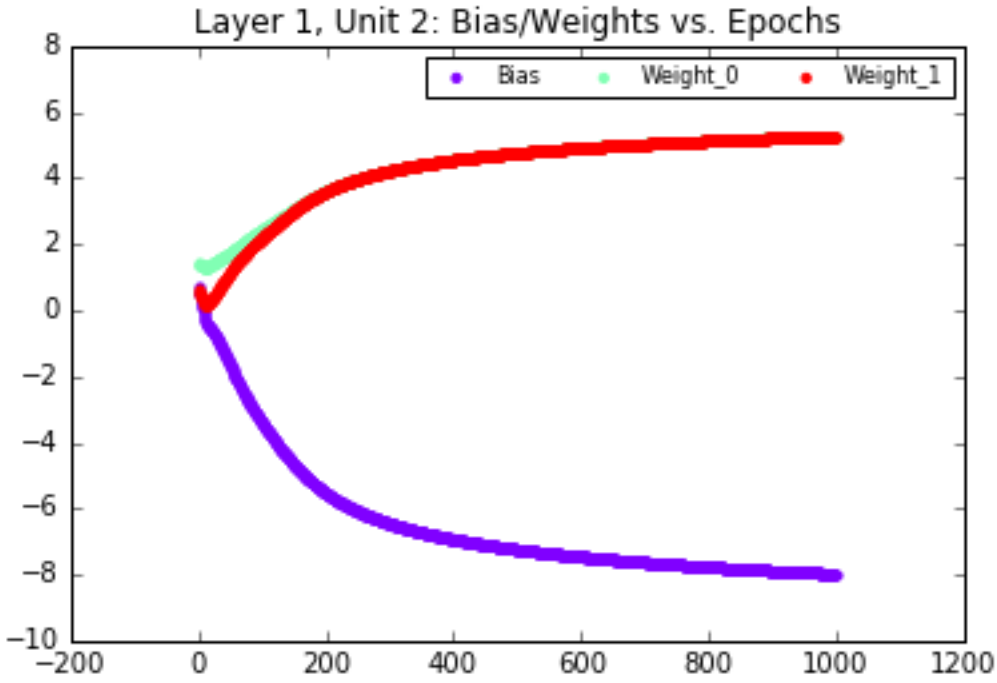


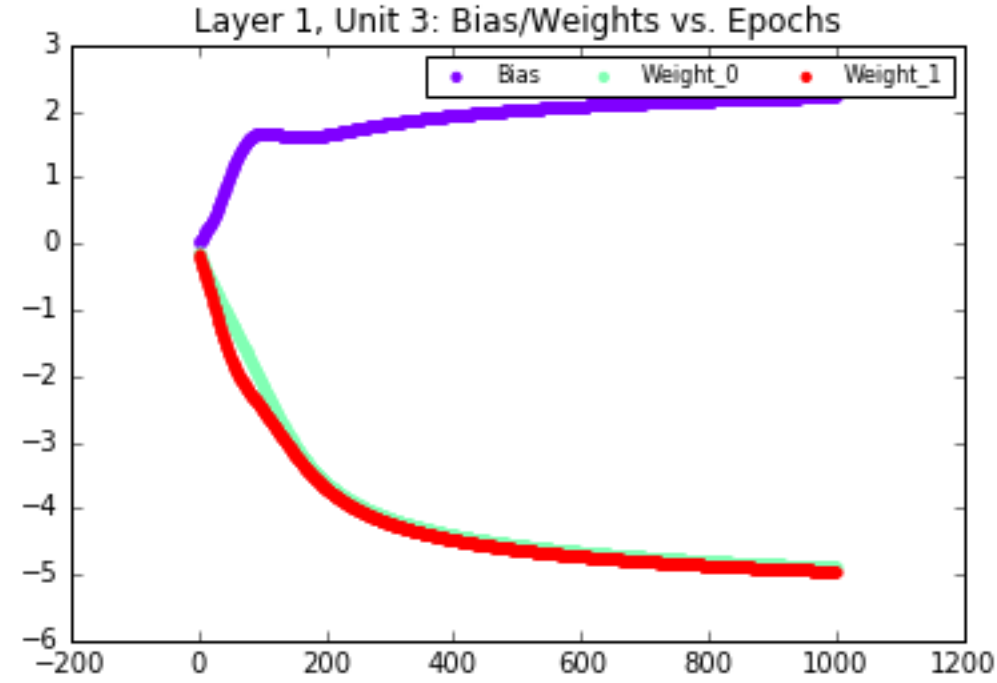


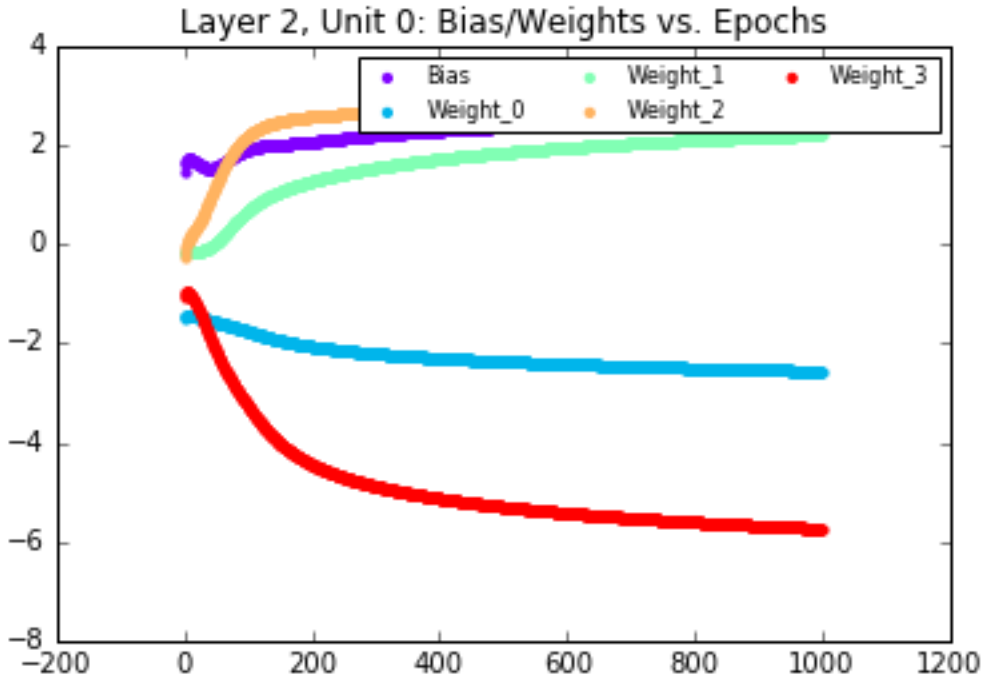
2 inputs, 4 hidden layer units, 4 outputs

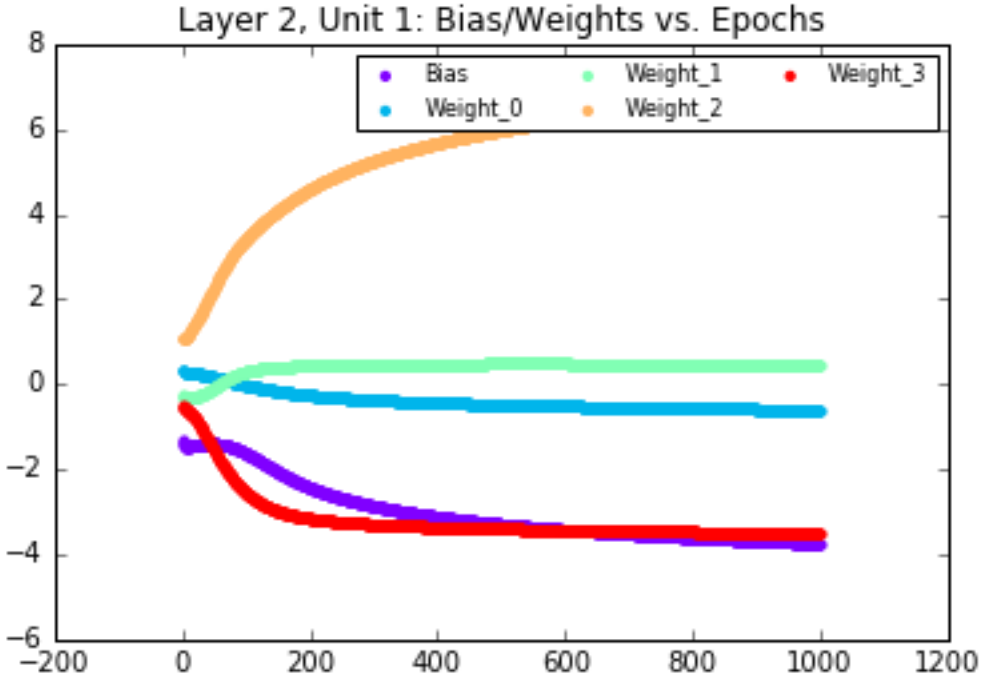

Layer 1, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−200
0
200
400
600
800
1000
1200

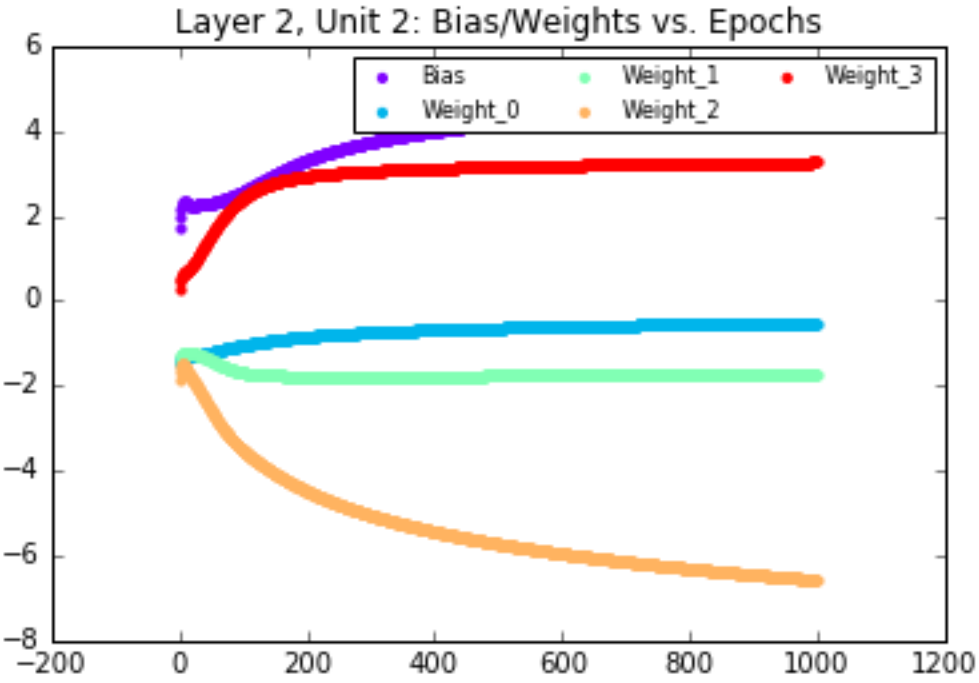

Layer 1, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−200
0
200
400
600
800
1000
1200

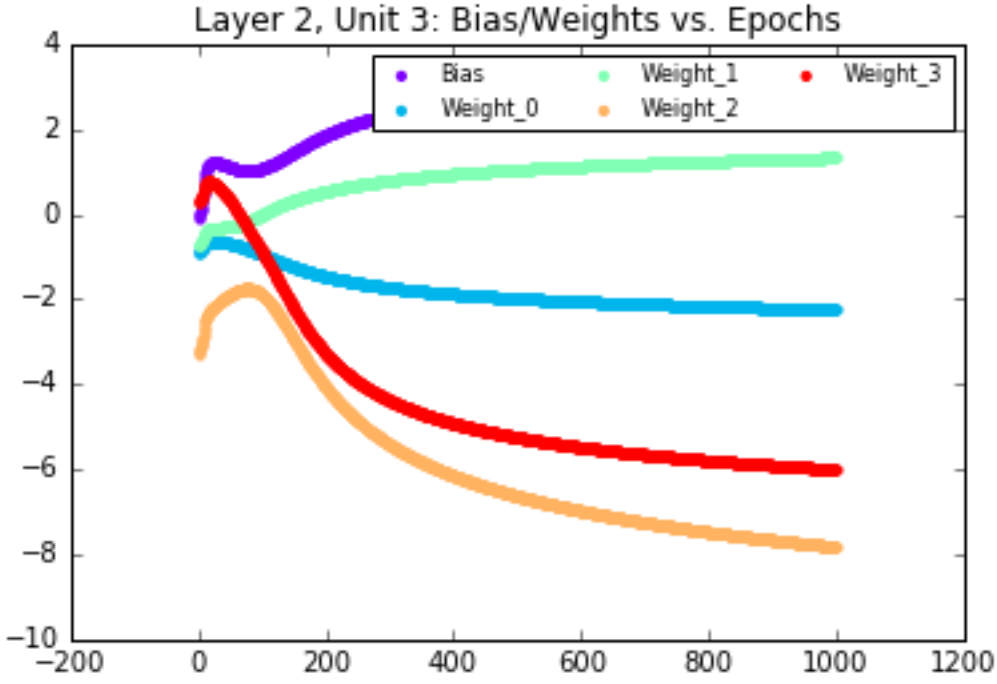

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

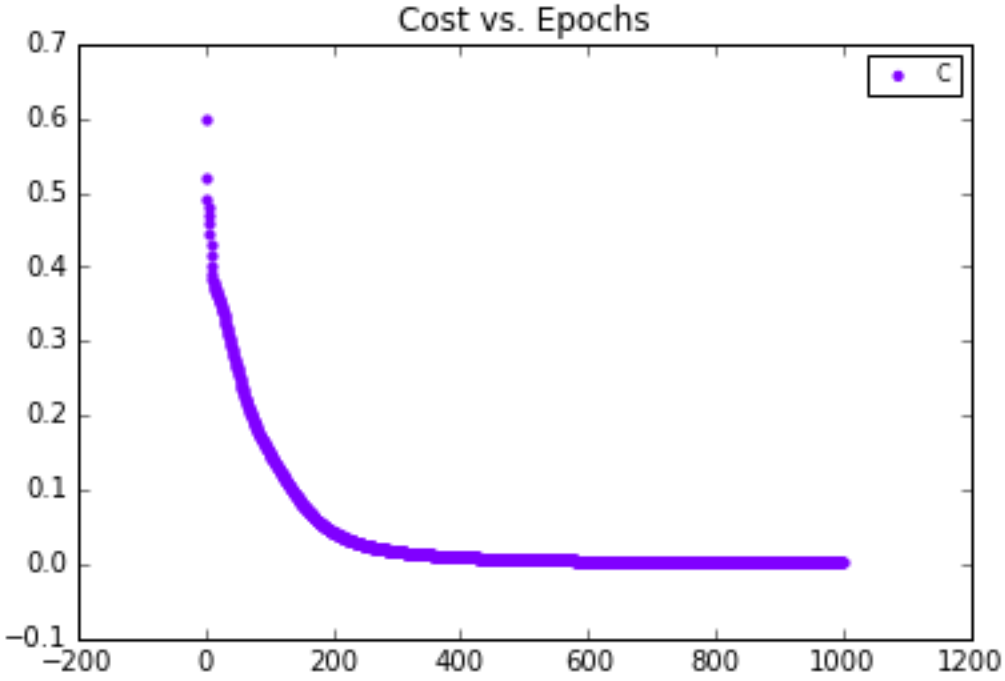

Layer 1, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
3
2
1
0
−1
−2
−3
−4
−5
−6
−200
0
200
400
600
800
1000
1200


Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

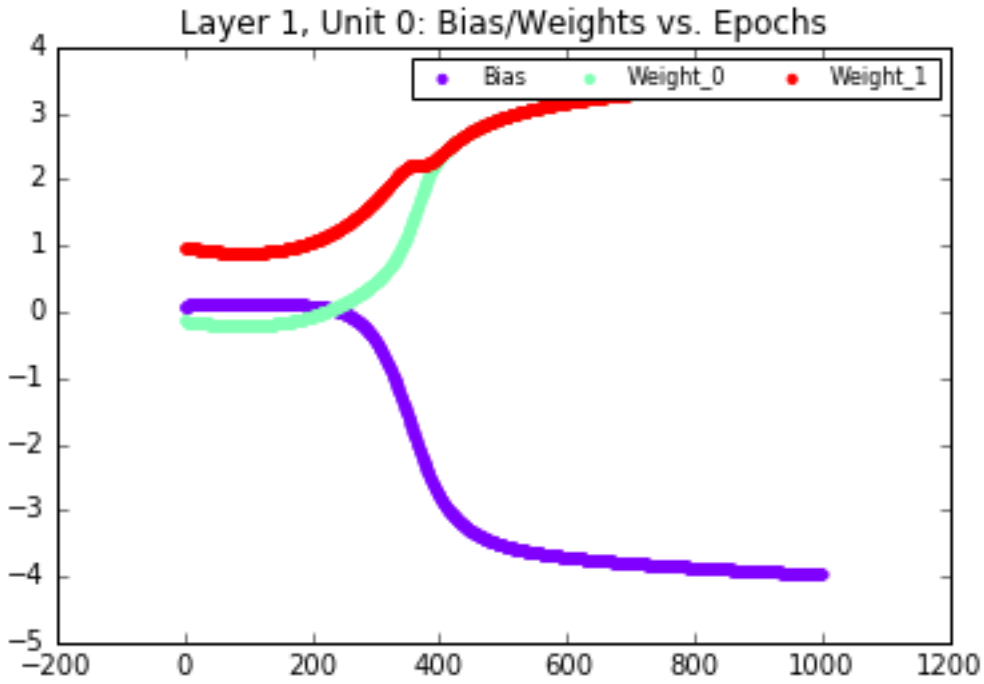

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

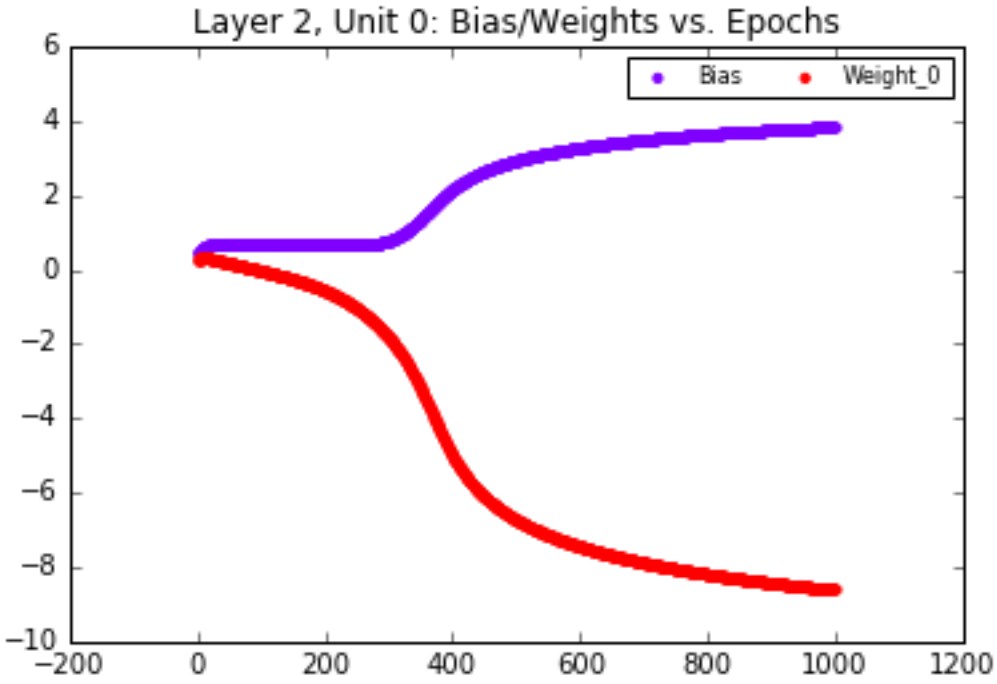

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

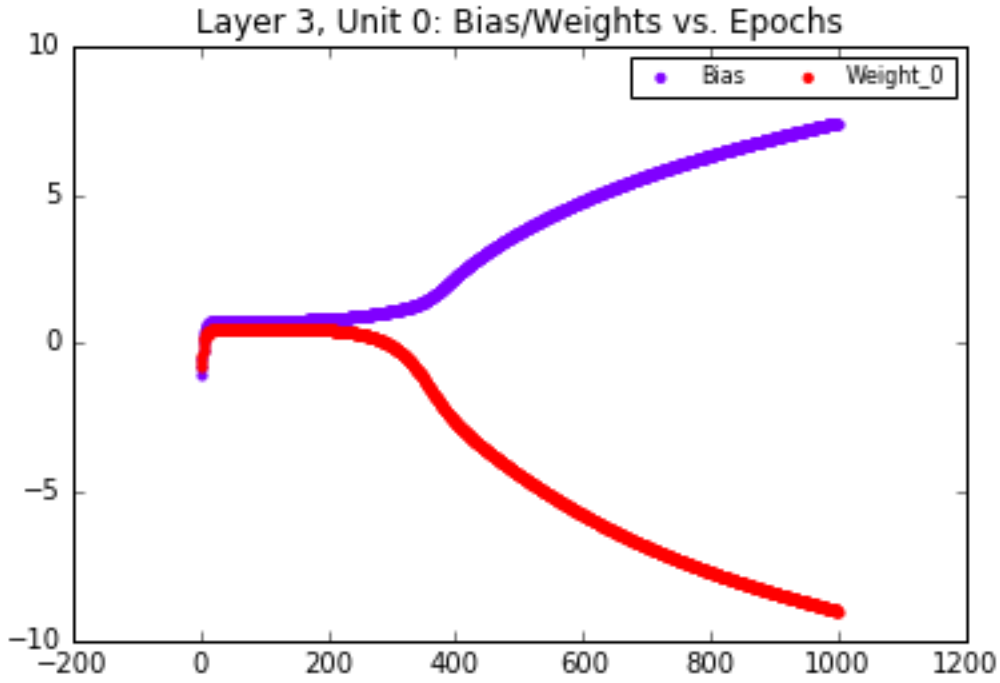

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

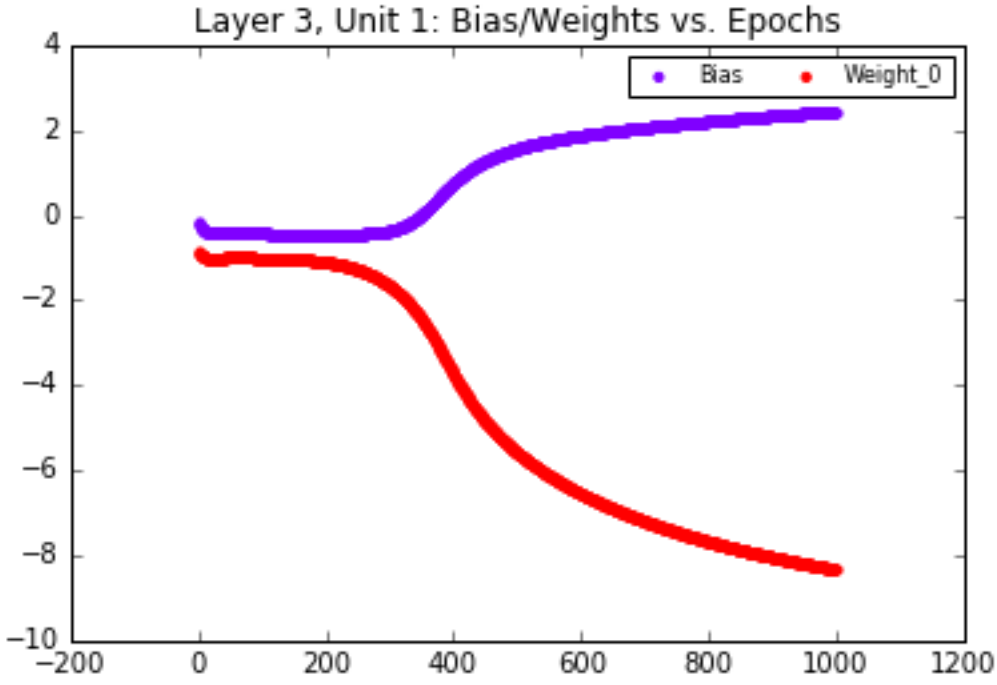

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

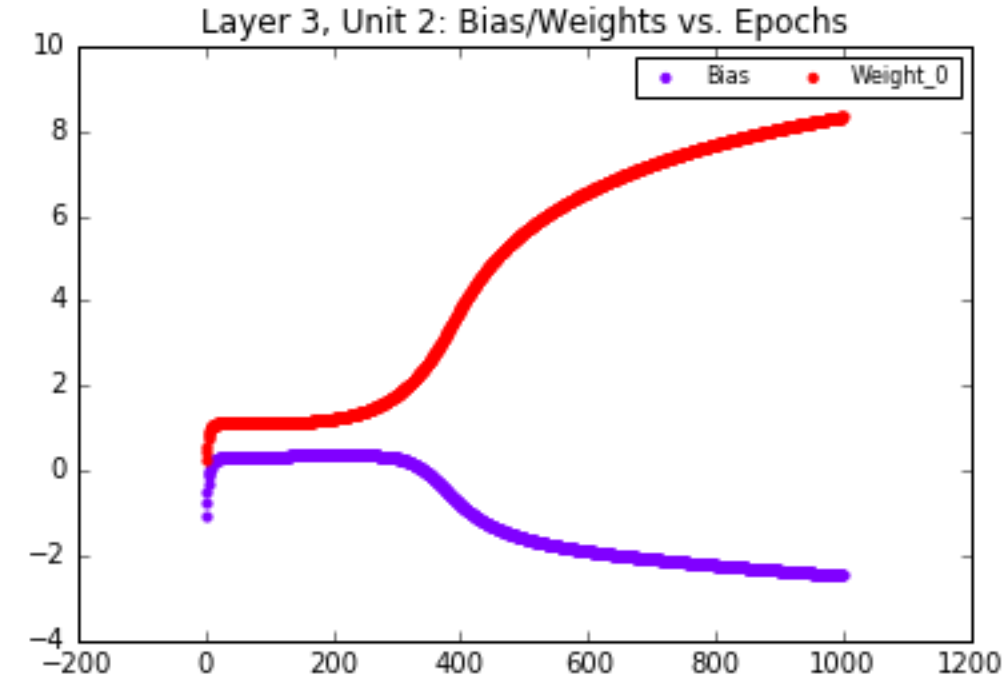
2 inputs, 2 hidden layers each of size 1, four outputs

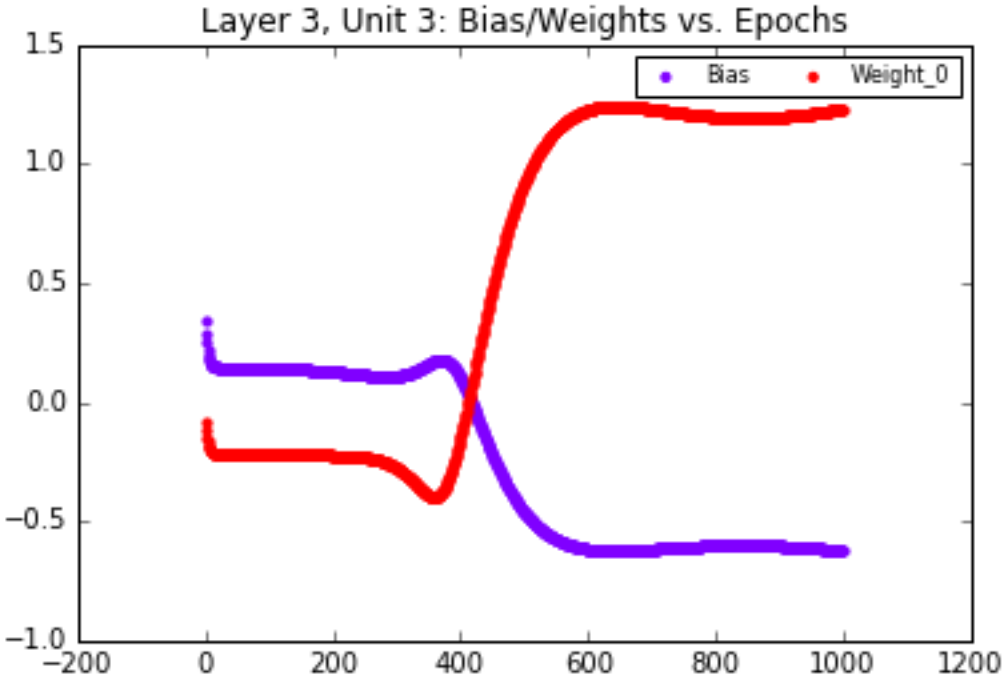


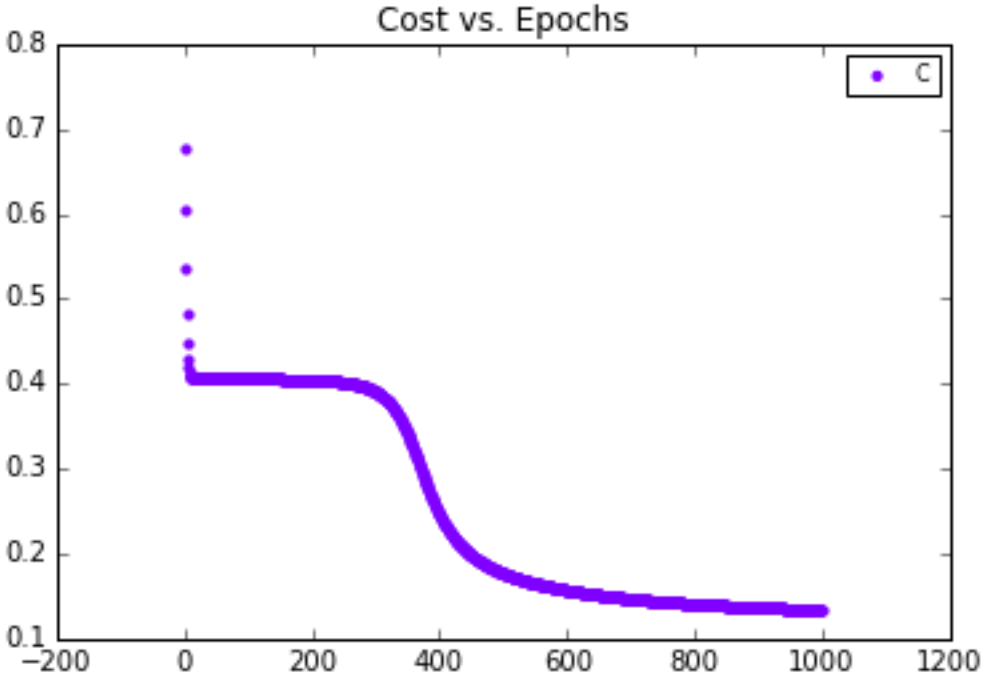



Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
10
5
0
−5
−10
−200
0
200
400
600
800
1000
1200

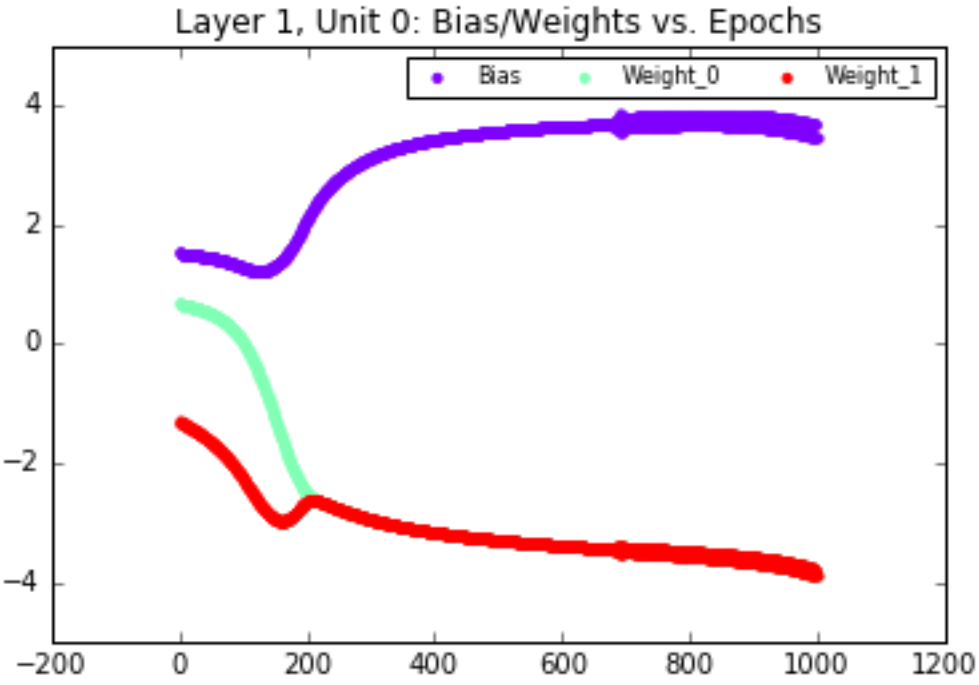

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

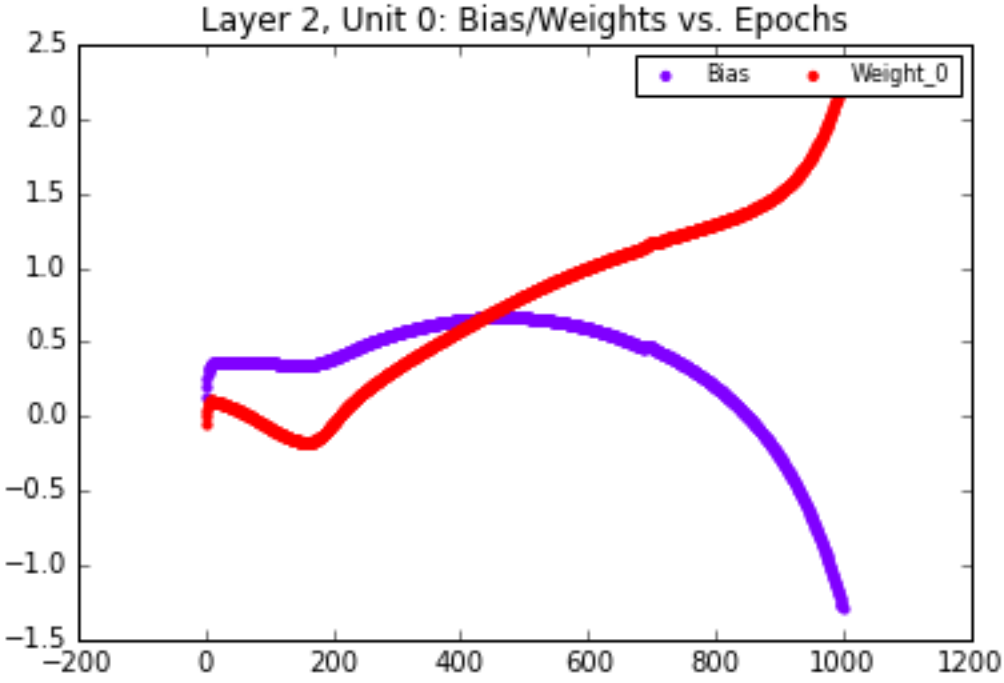

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
10
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

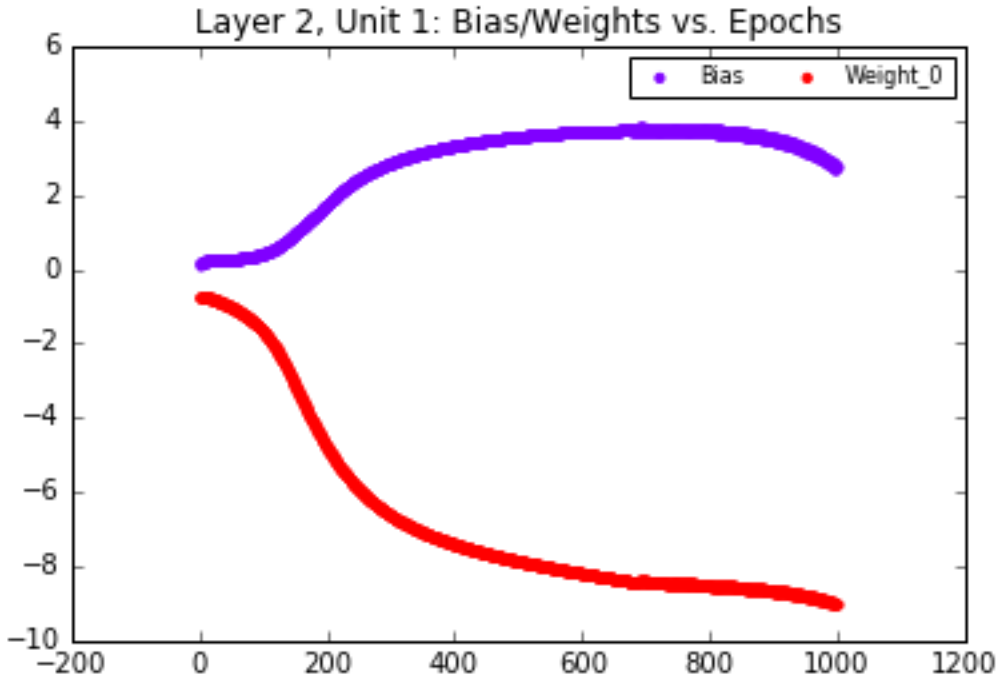

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
1.5
1.0
0.5
0.0
−0.5
−1.0
−200
0
200
400
600
800
1000
1200

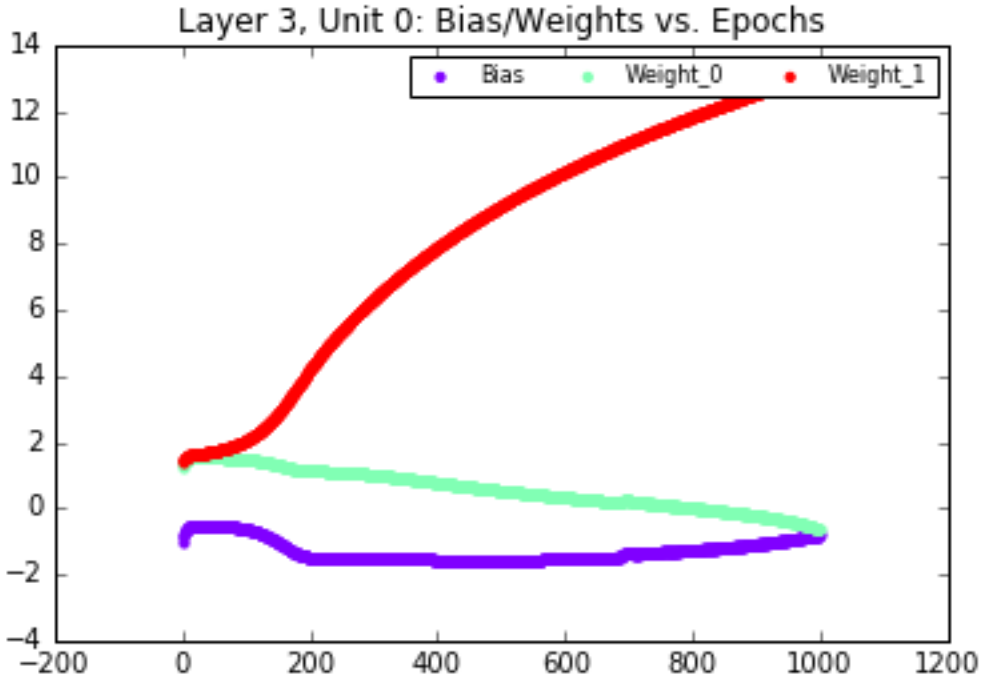

Cost vs. Epochs
C
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
−200
0
200
400
600
800
1000
1200

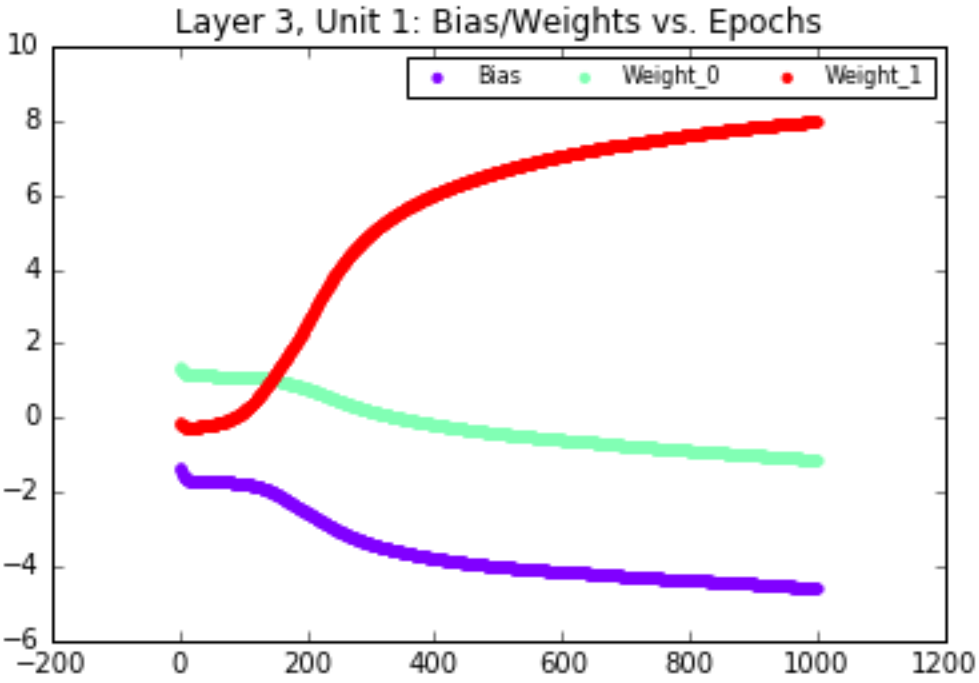
2 inputs, 2 hidden layers, first of size one, second of size 2, four outputs

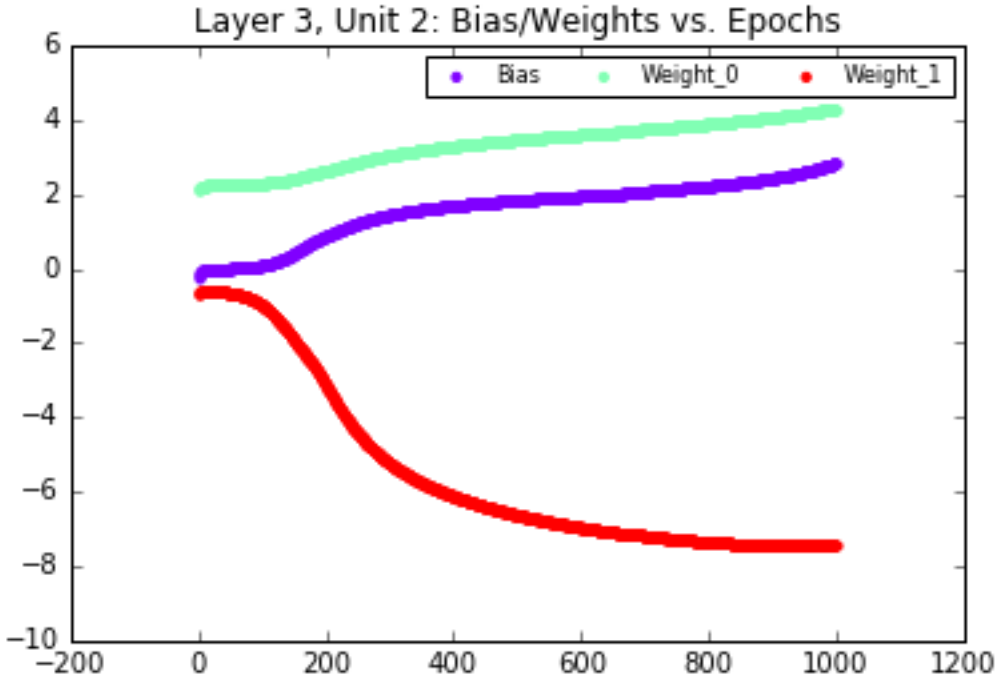


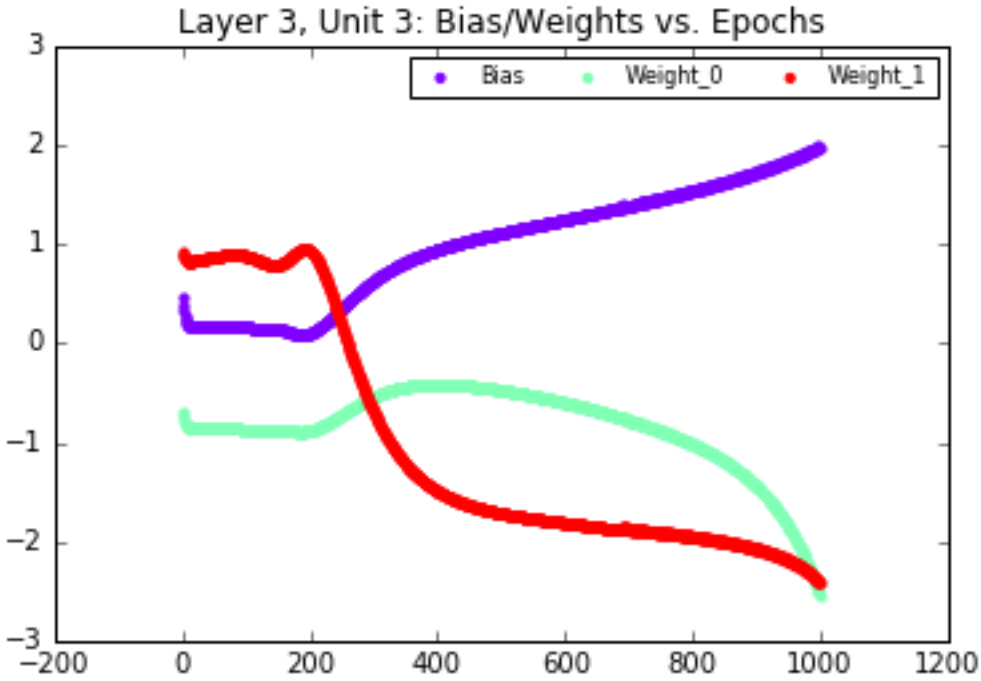


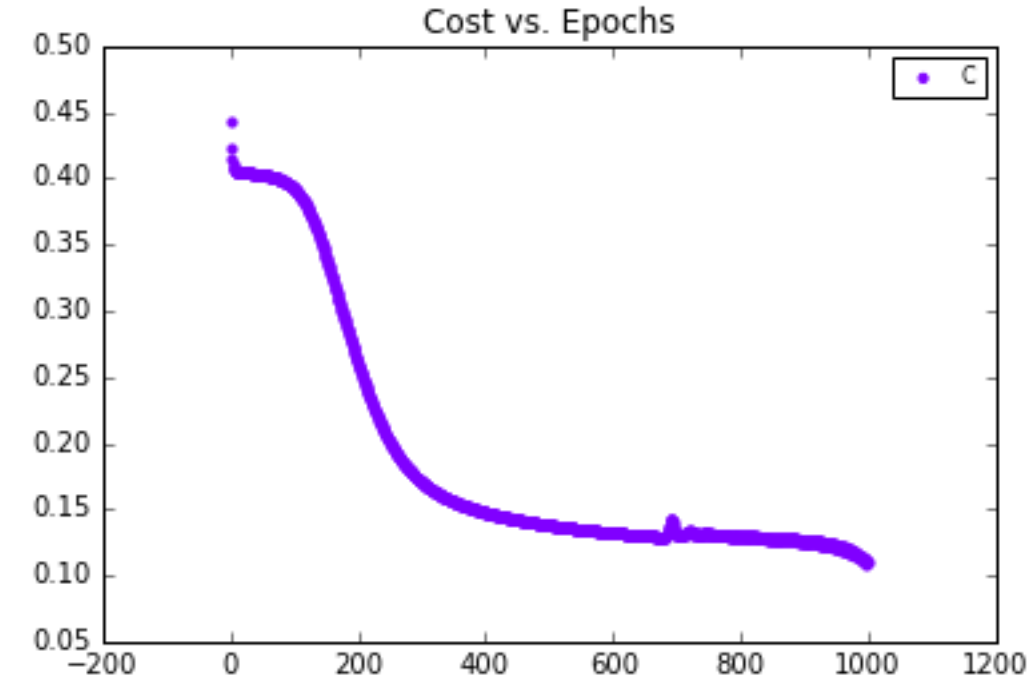

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
14
12
10
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

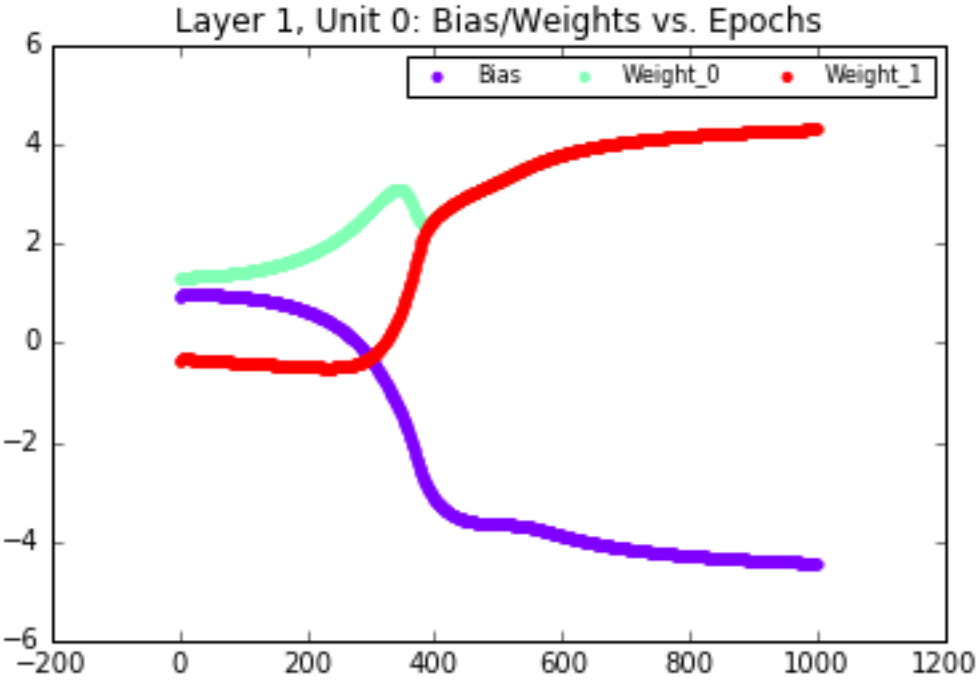

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
10
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

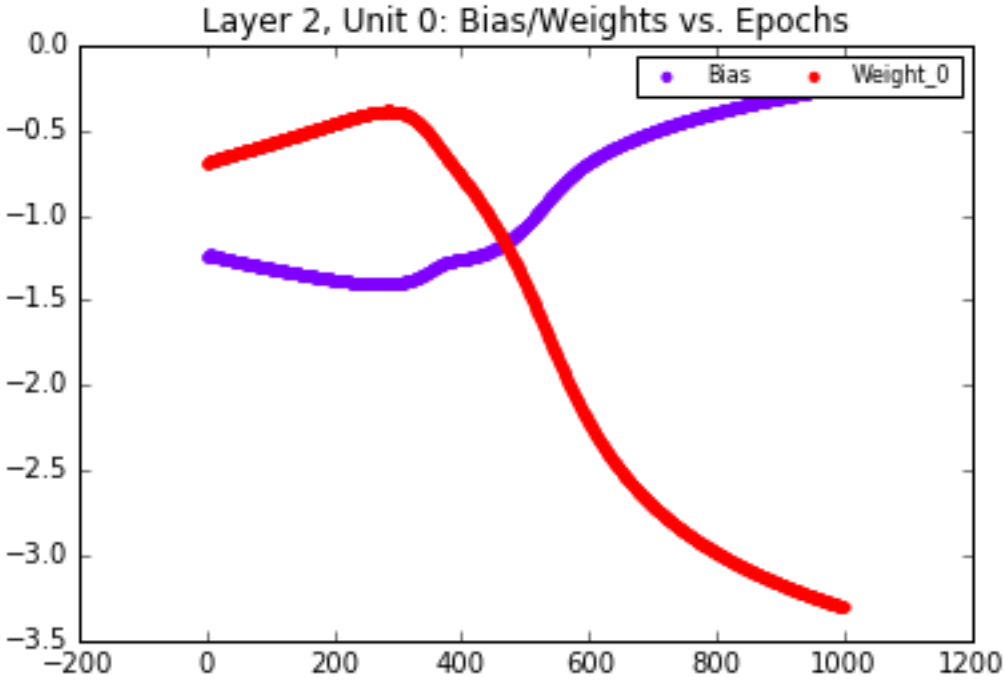

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

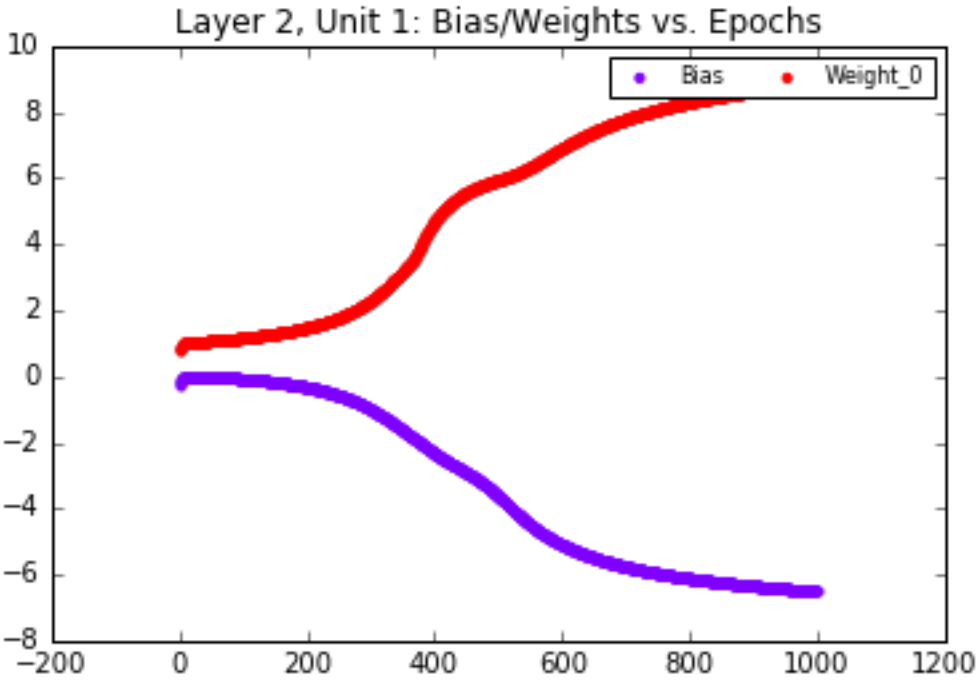

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
3
2
1
0
−1
−2
−3
−200
0
200
400
600
800
1000
1200

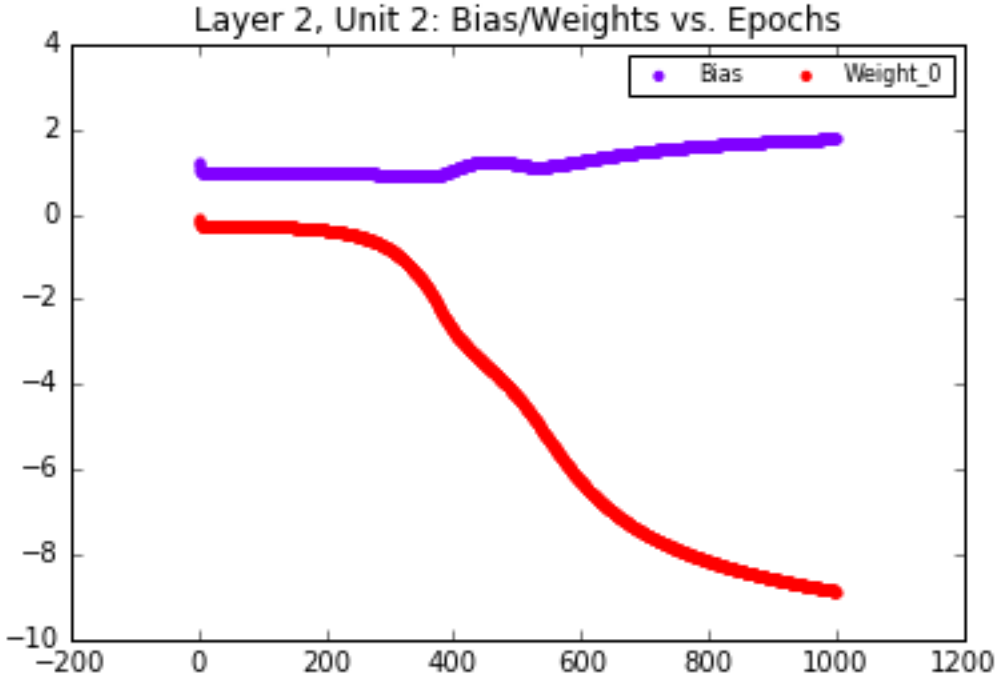

Cost vs. Epochs
C
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
−200
0
200
400
600
800
1000
1200

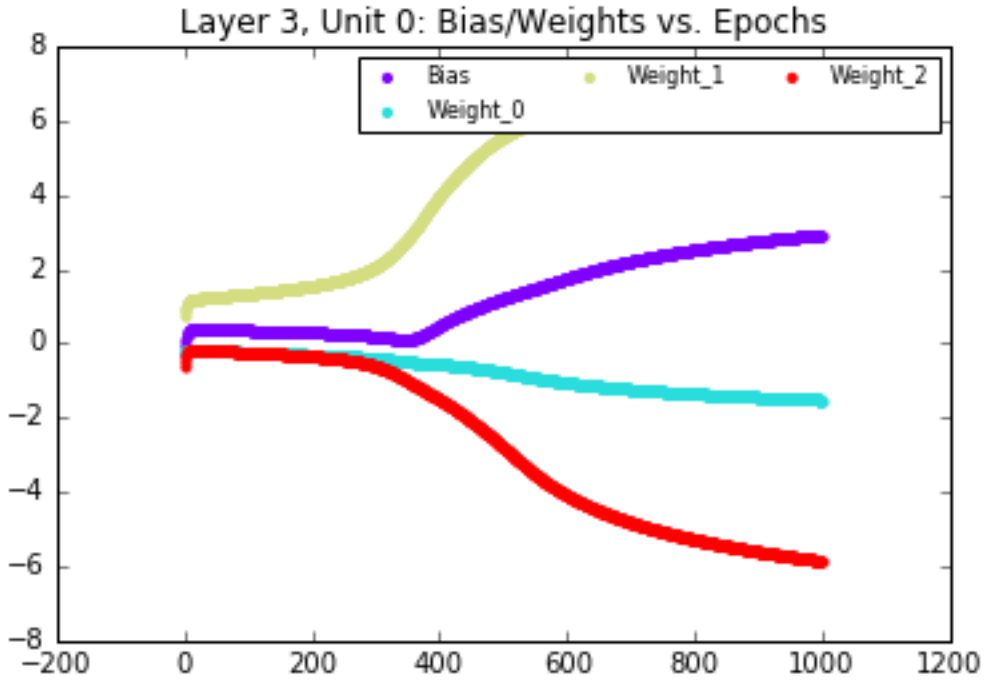
2 inputs, 2 hidden layers, first of size one, second of size 3, four outputs

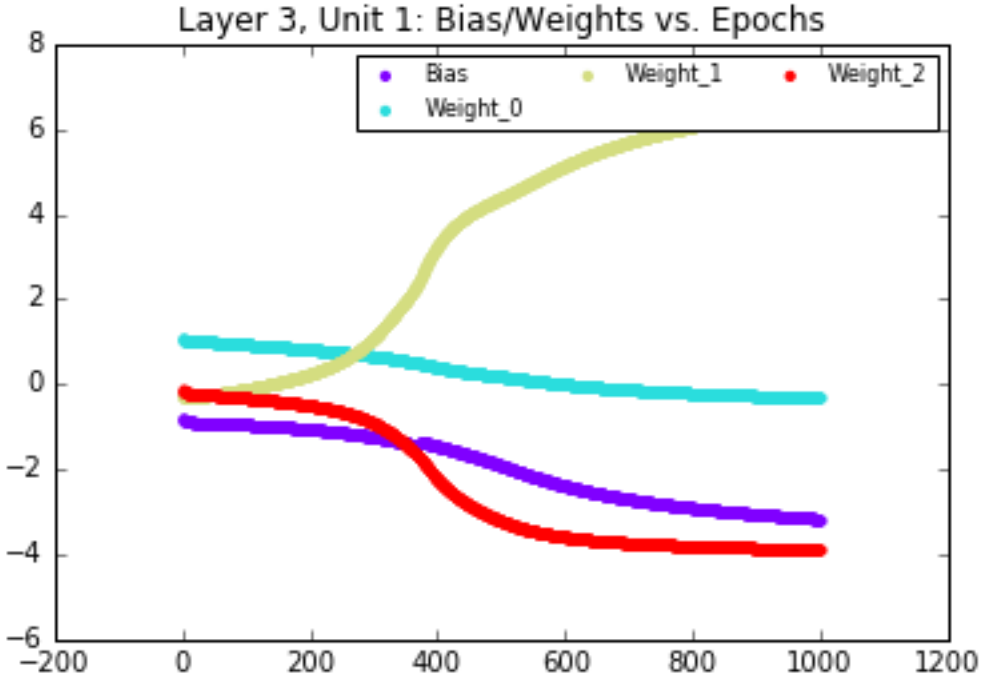


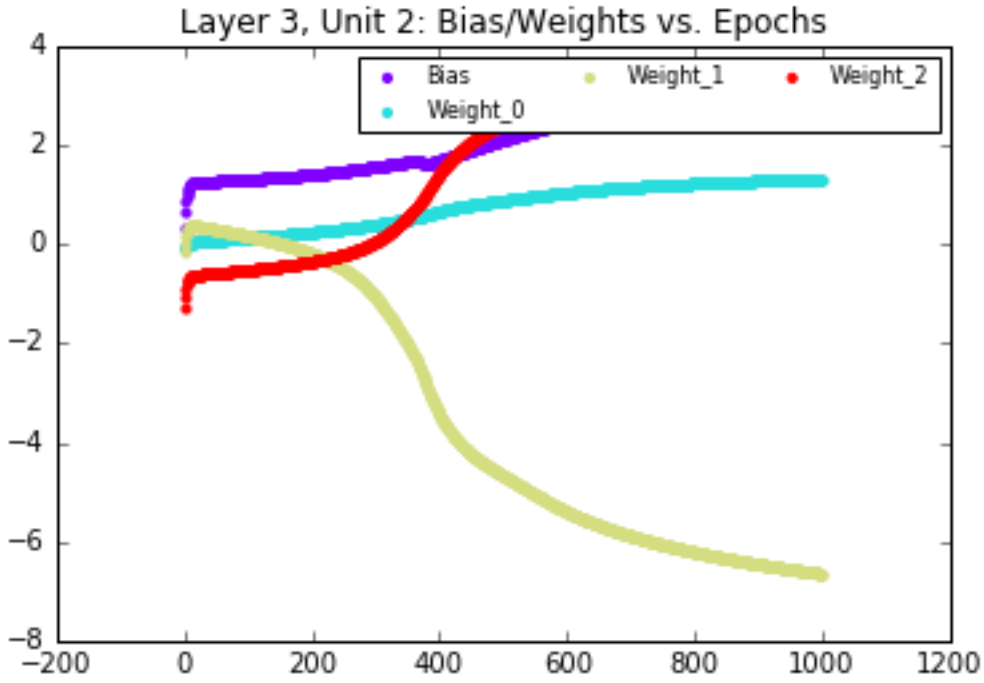


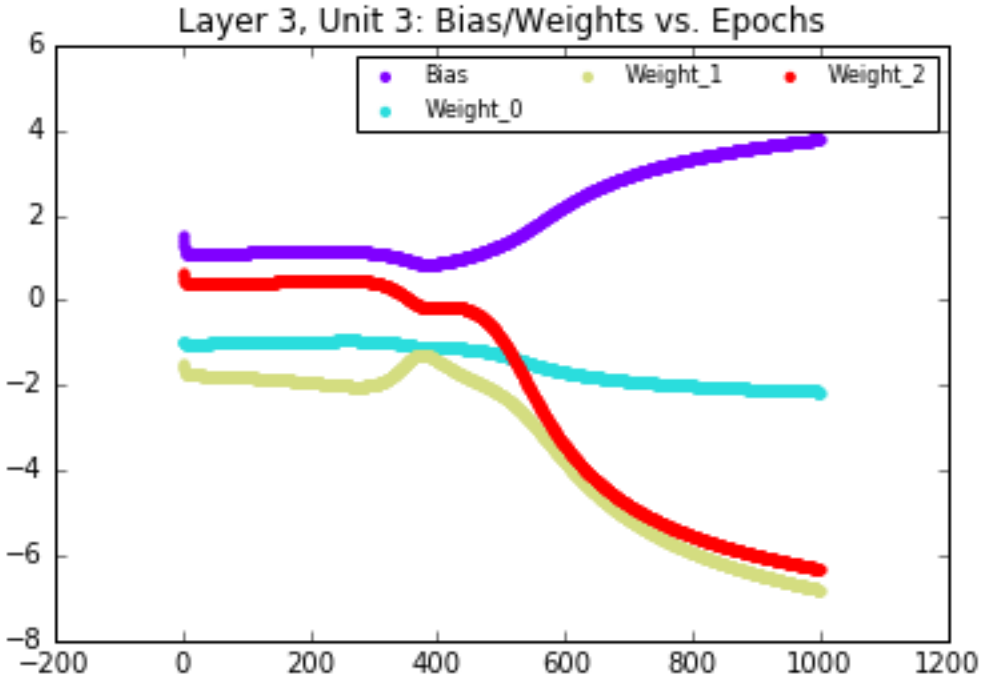

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
10
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

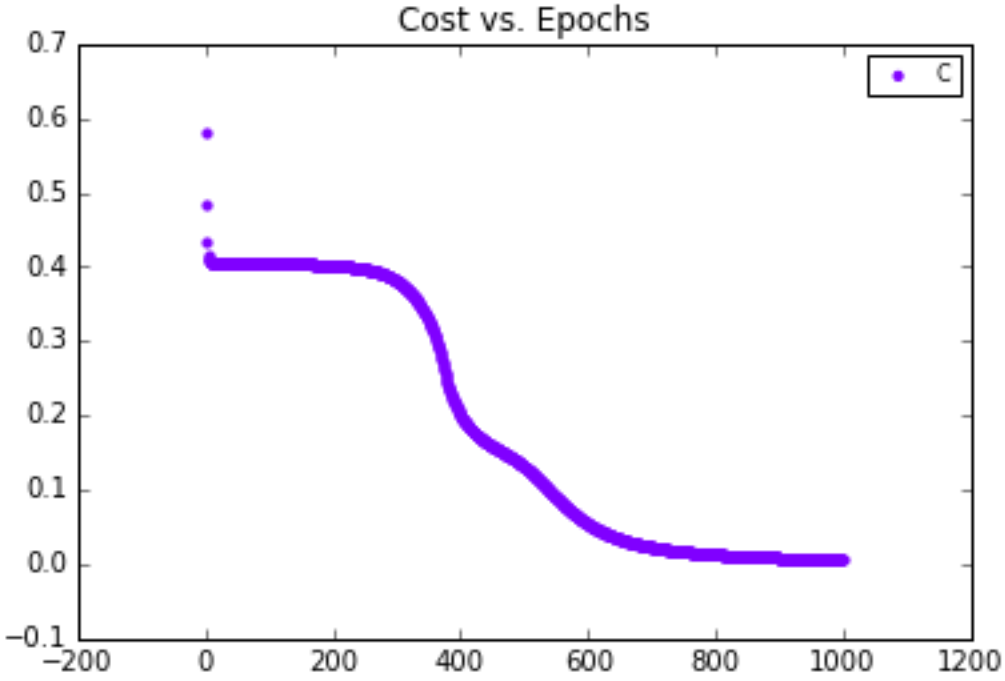

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

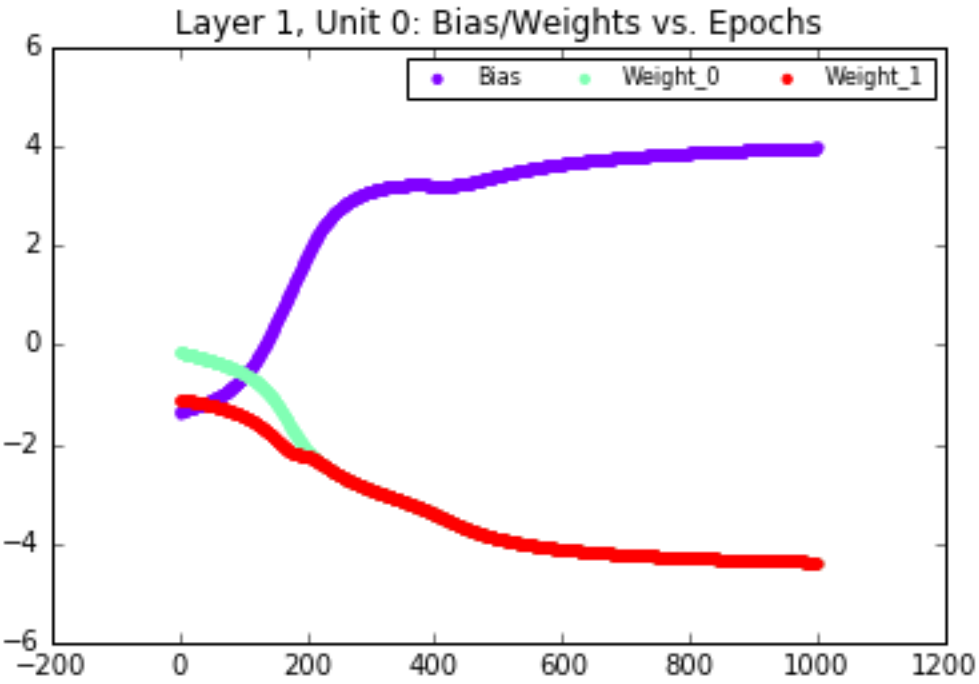

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

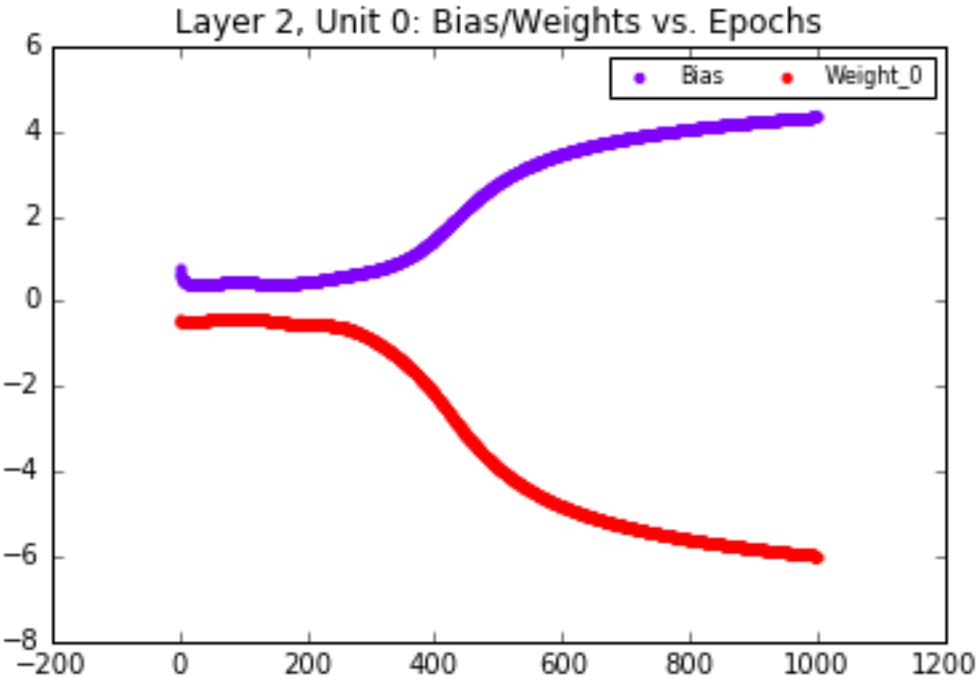

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

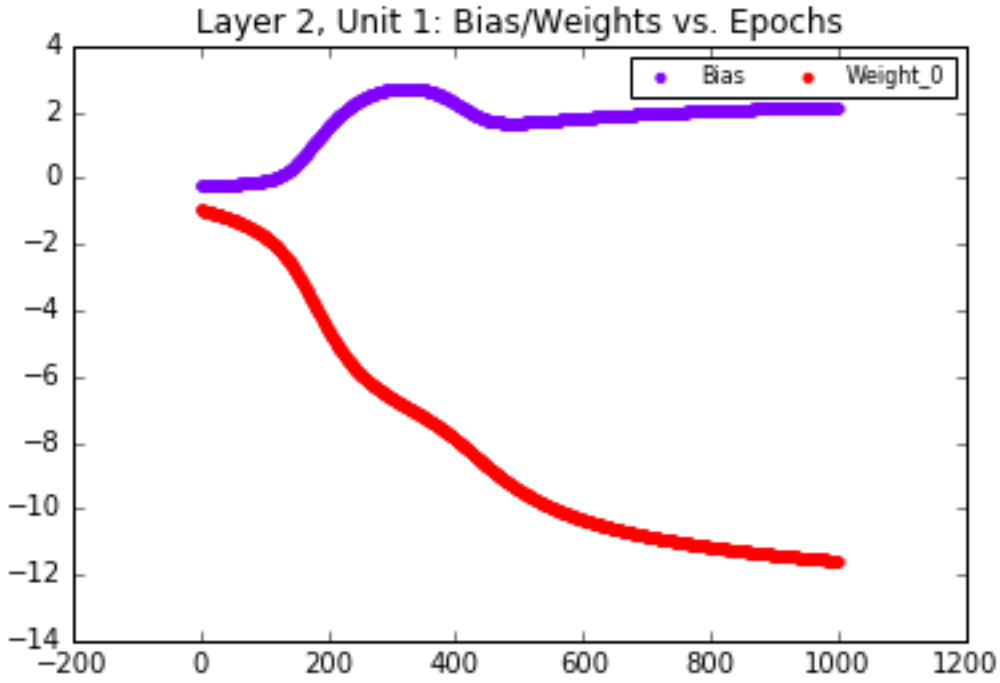

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

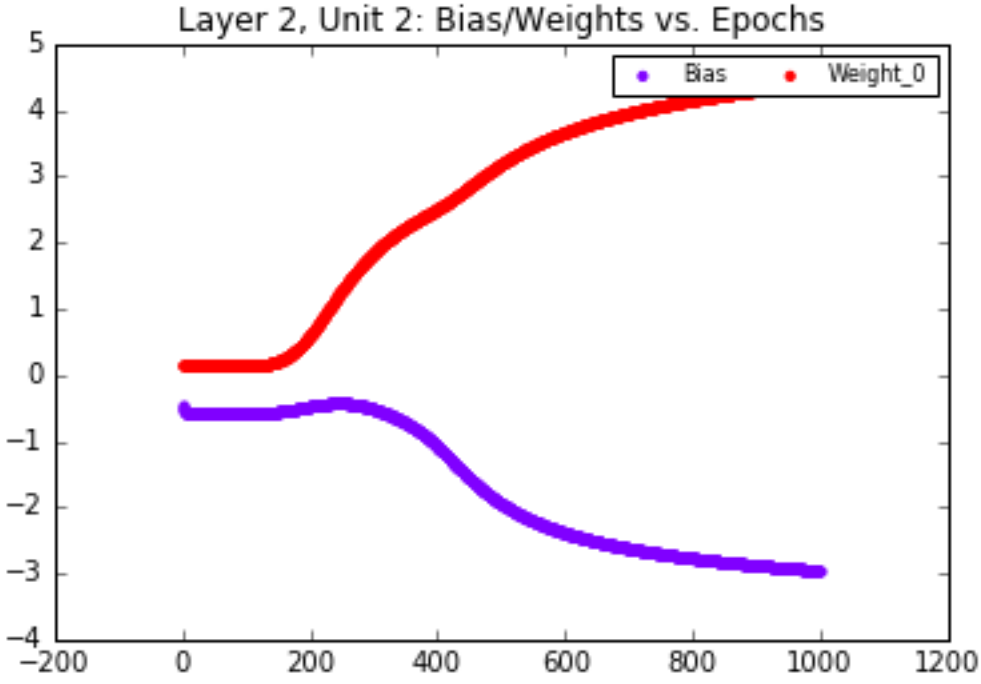

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

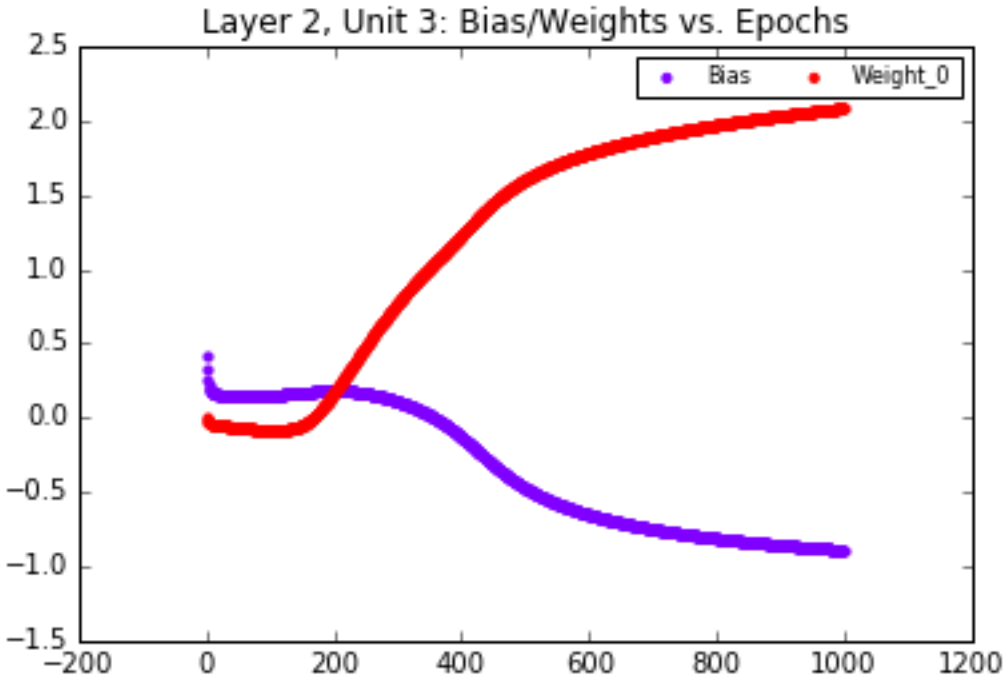
2 inputs, 2 hidden layers, first of size one, second of size 4, four outputs

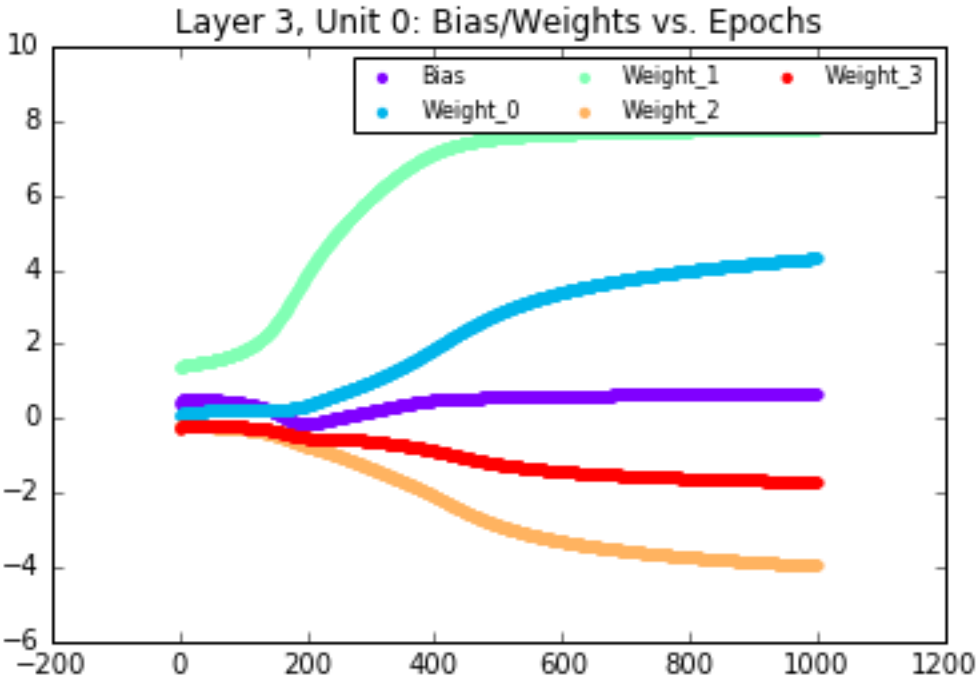


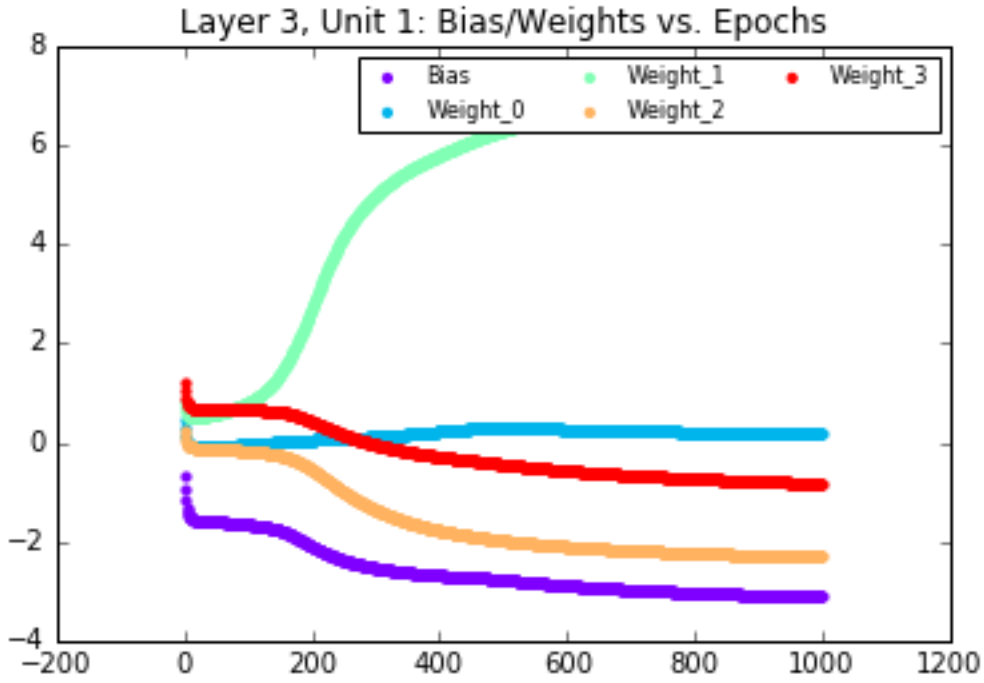


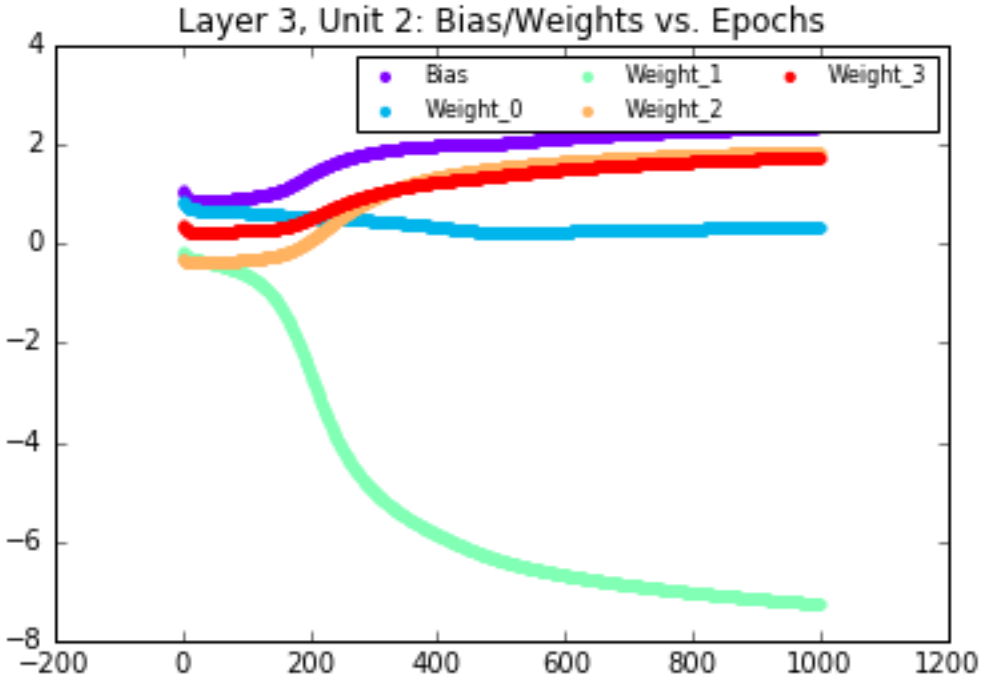

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
4
2
0
−2
−4
−6
−8
−10
−12
−14
−200
0
200
400
600
800
1000
1200

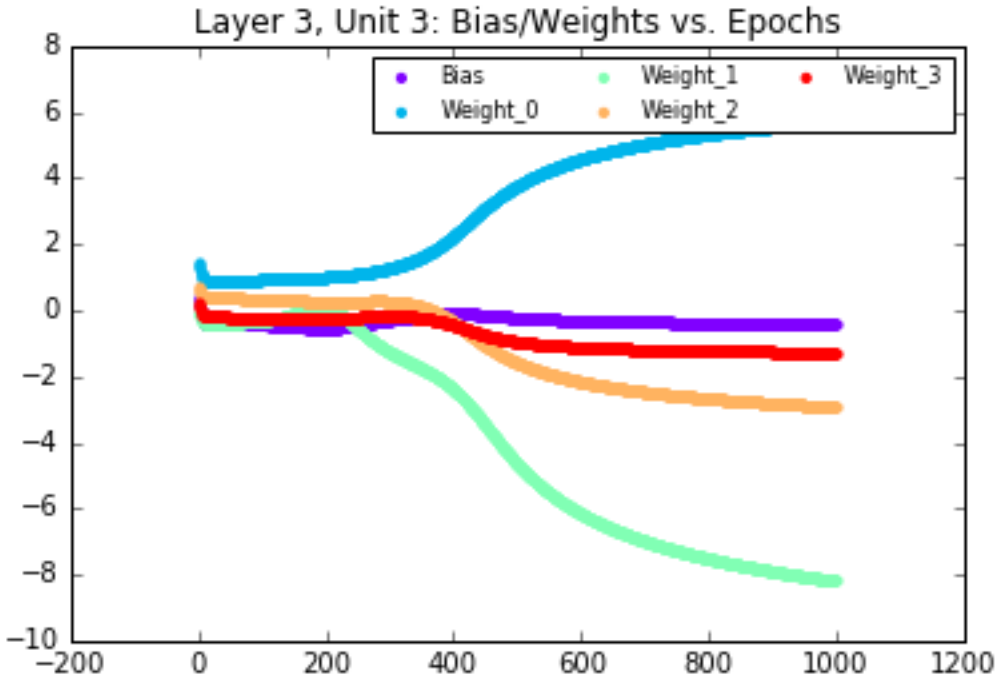

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
5
4
3
2
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

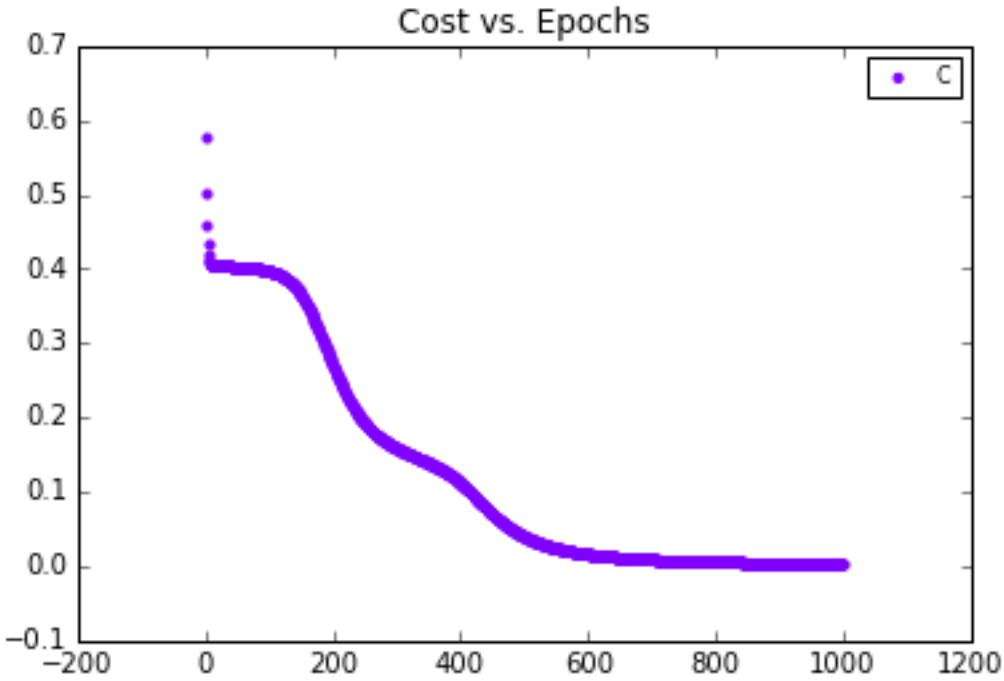

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
10
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

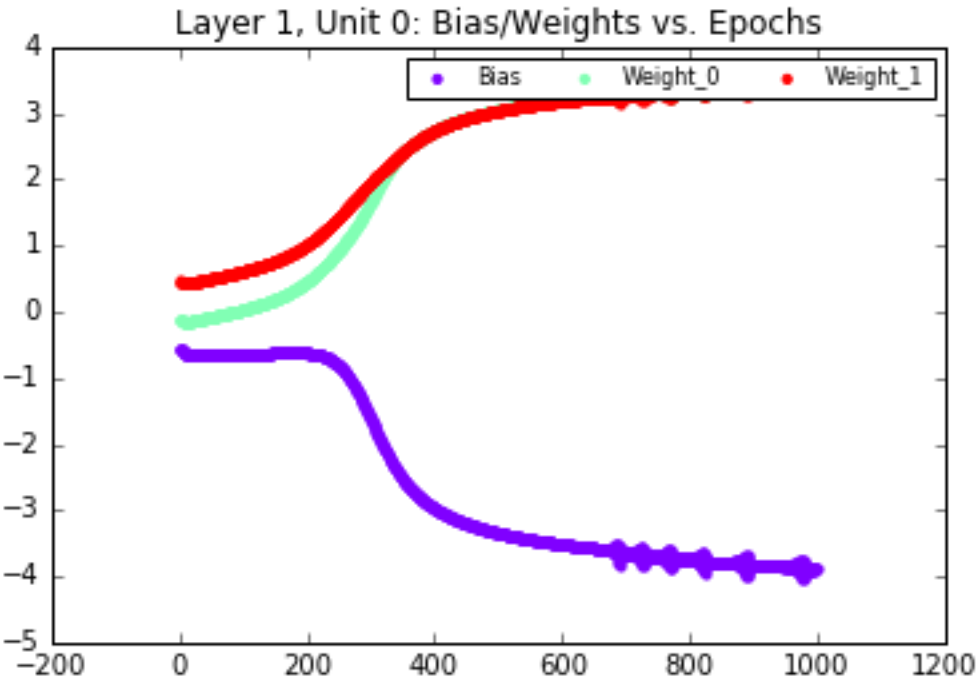

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

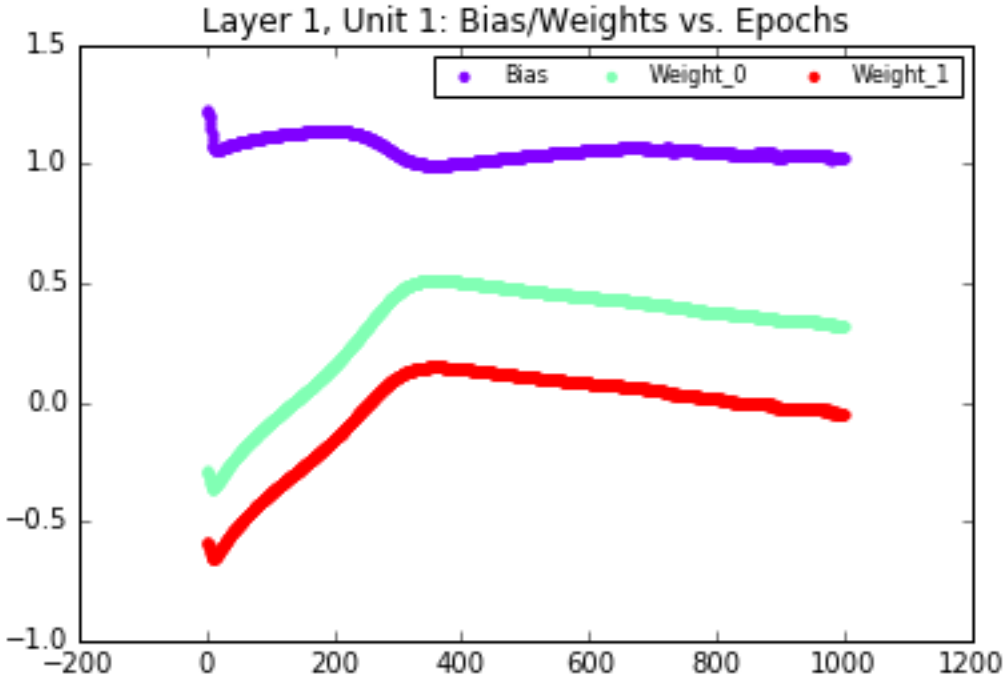

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

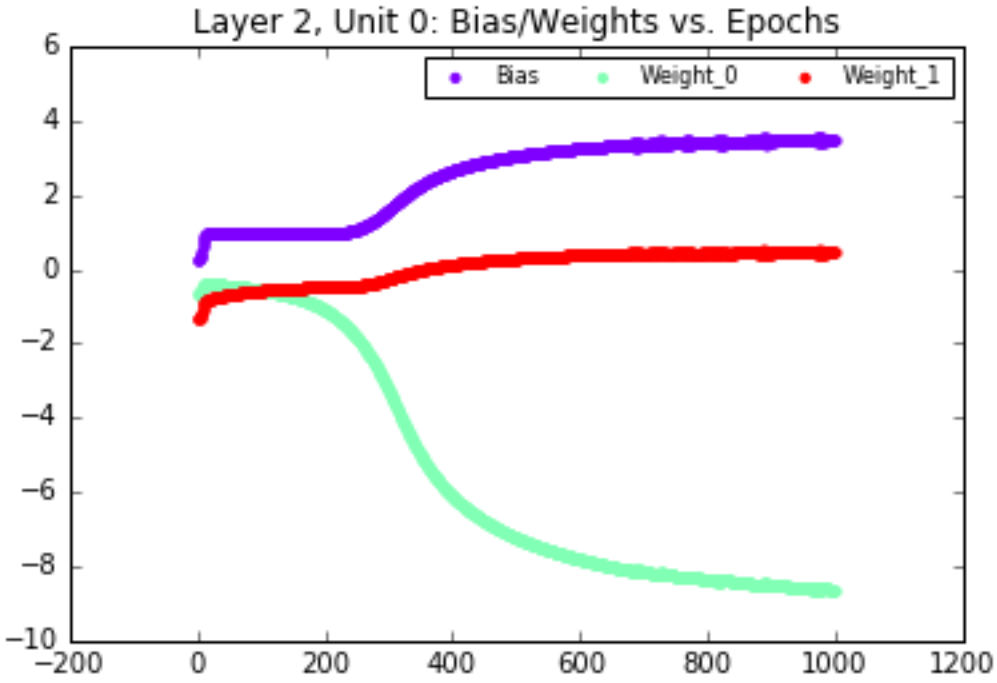

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

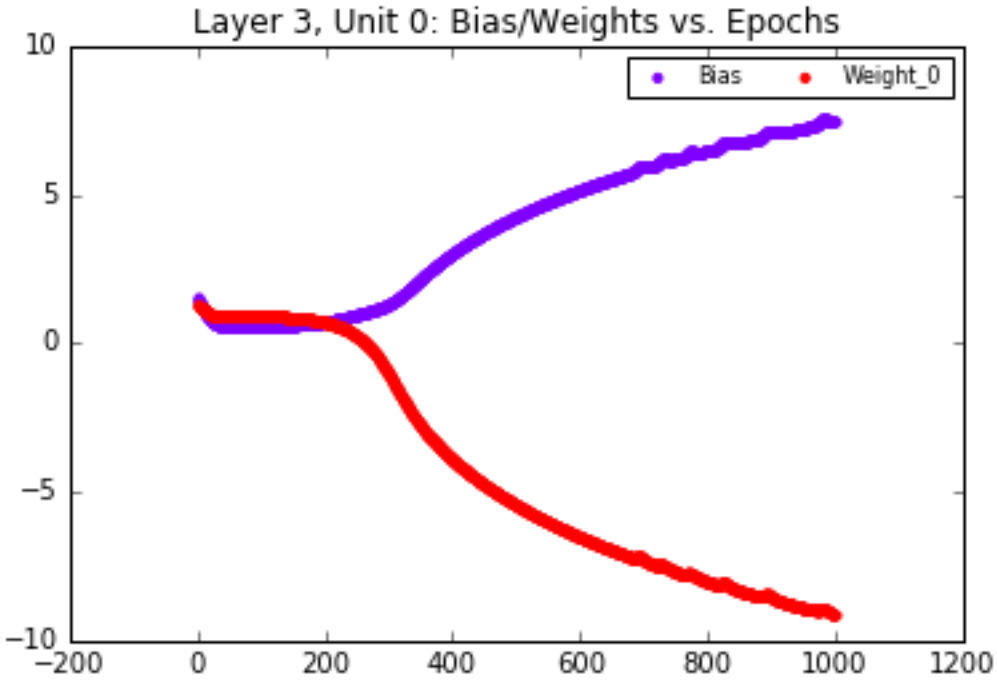

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

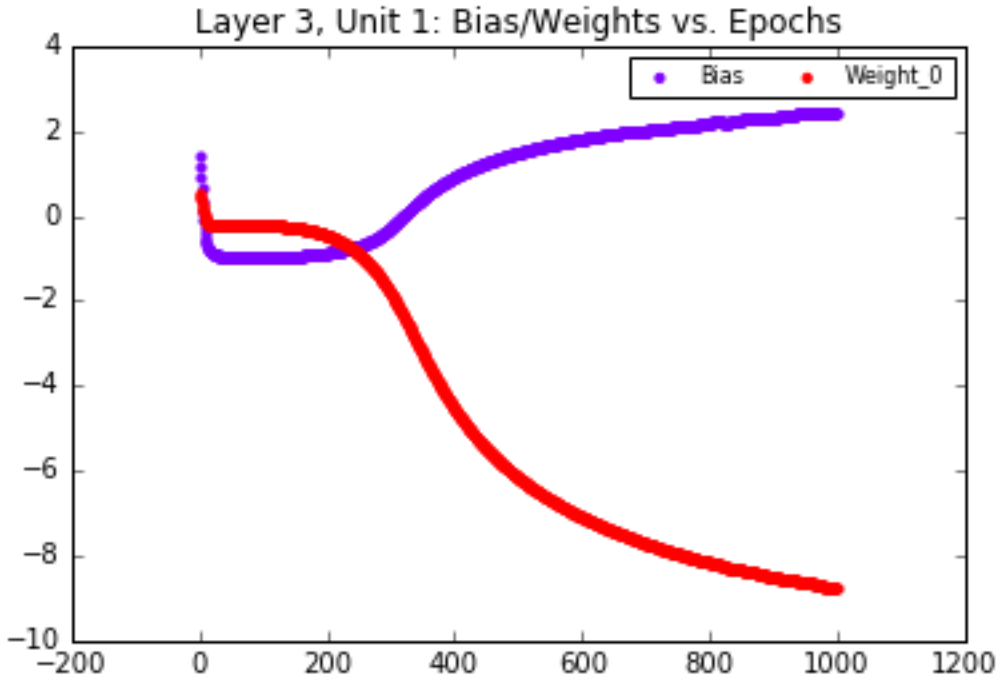
2 input units, 2 hidden layers, first of size 2 second of size 1, 4 output units

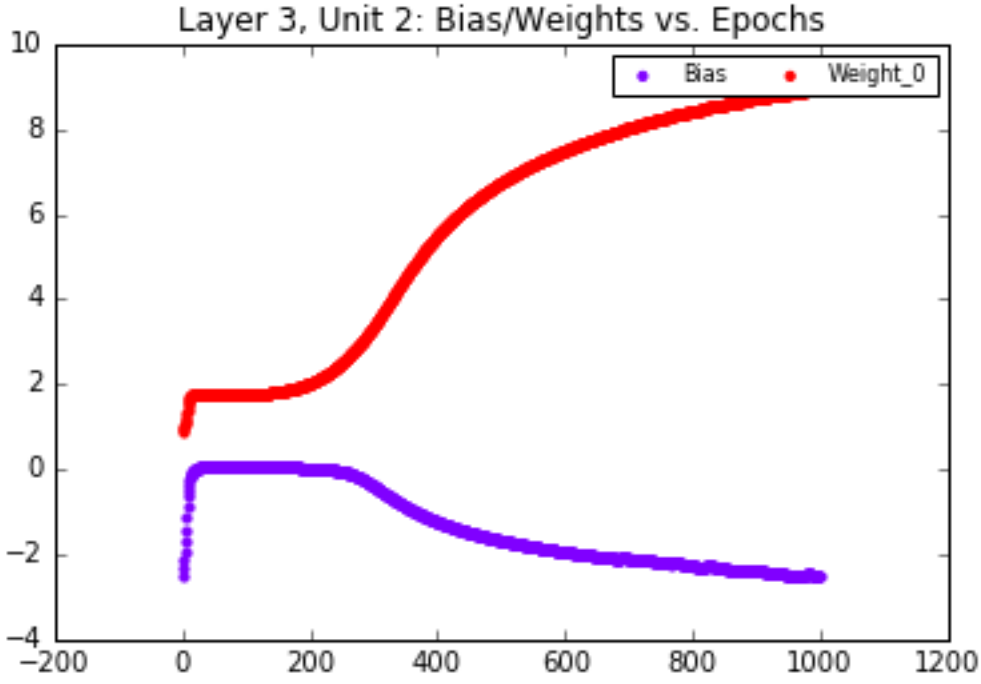


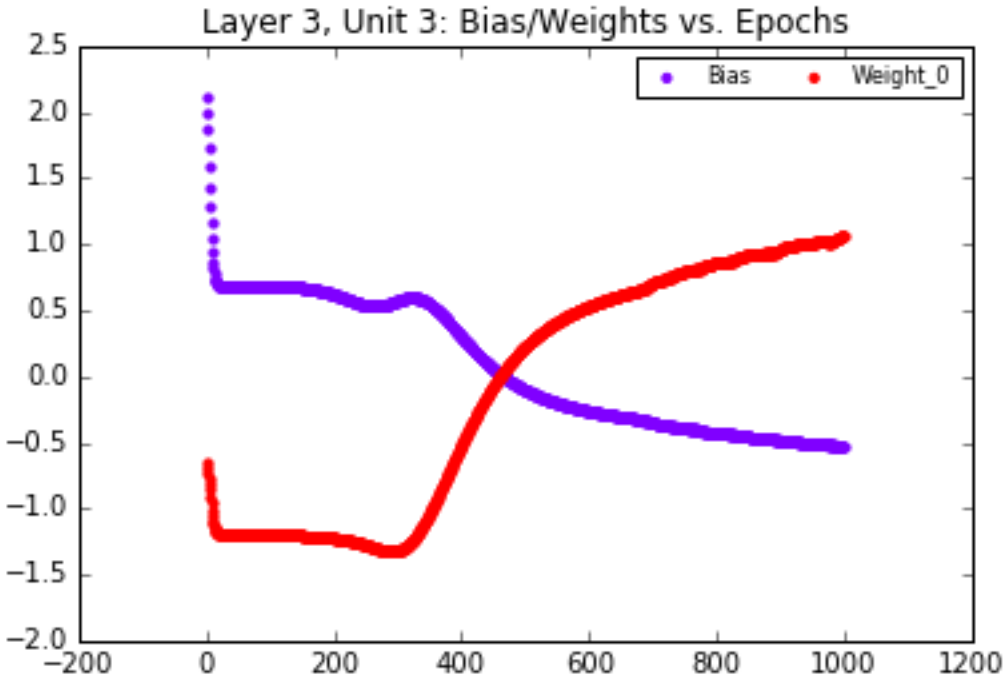


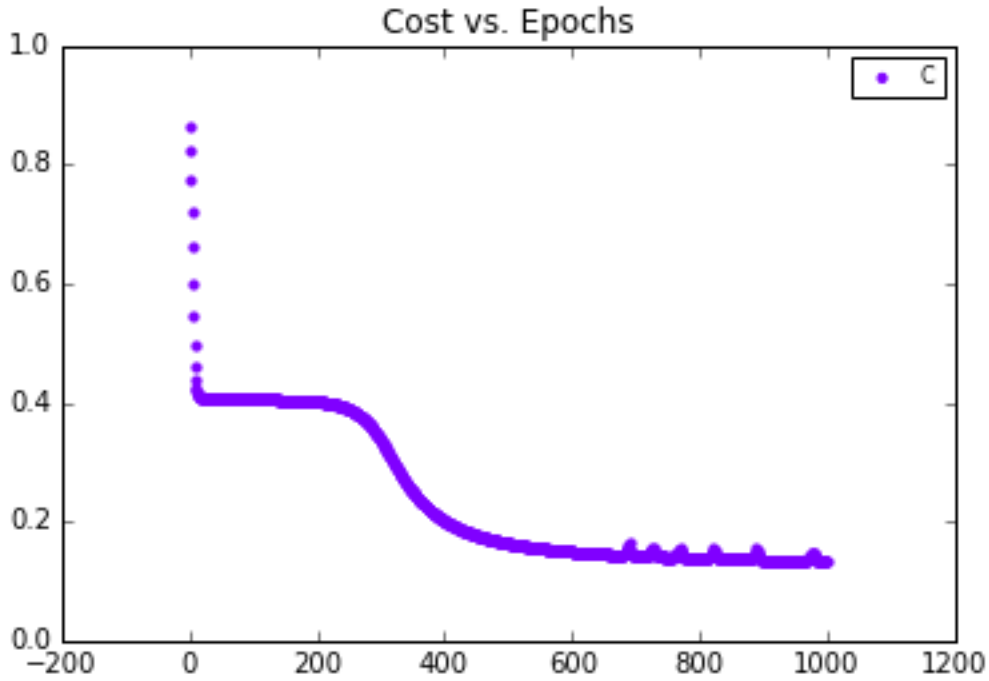

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
10
5
0
−5
−10
−200
0
200
400
600
800
1000
1200

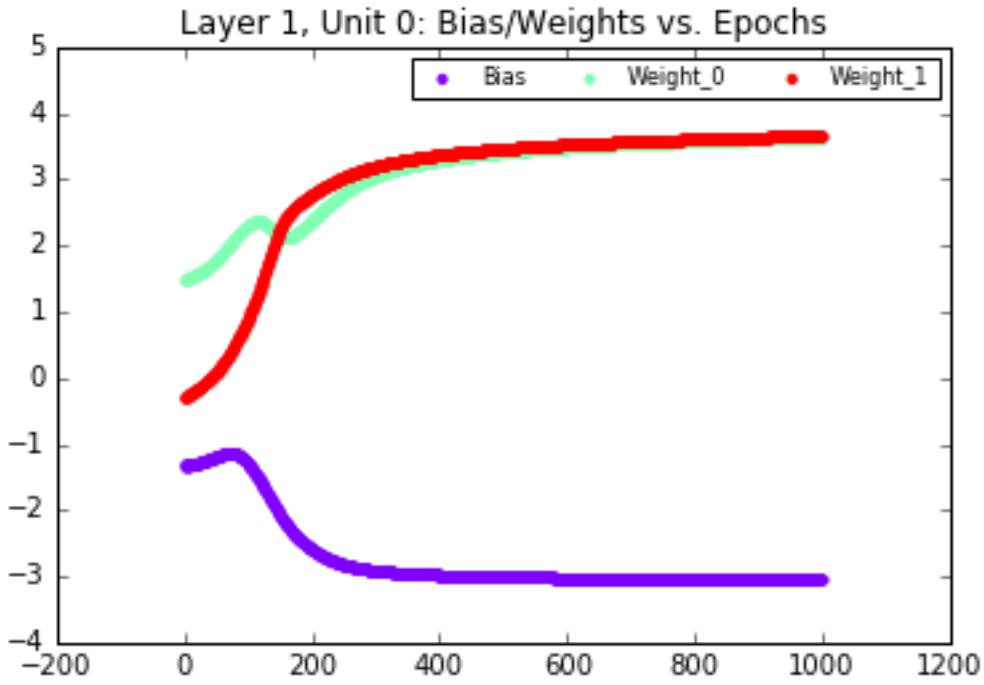

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

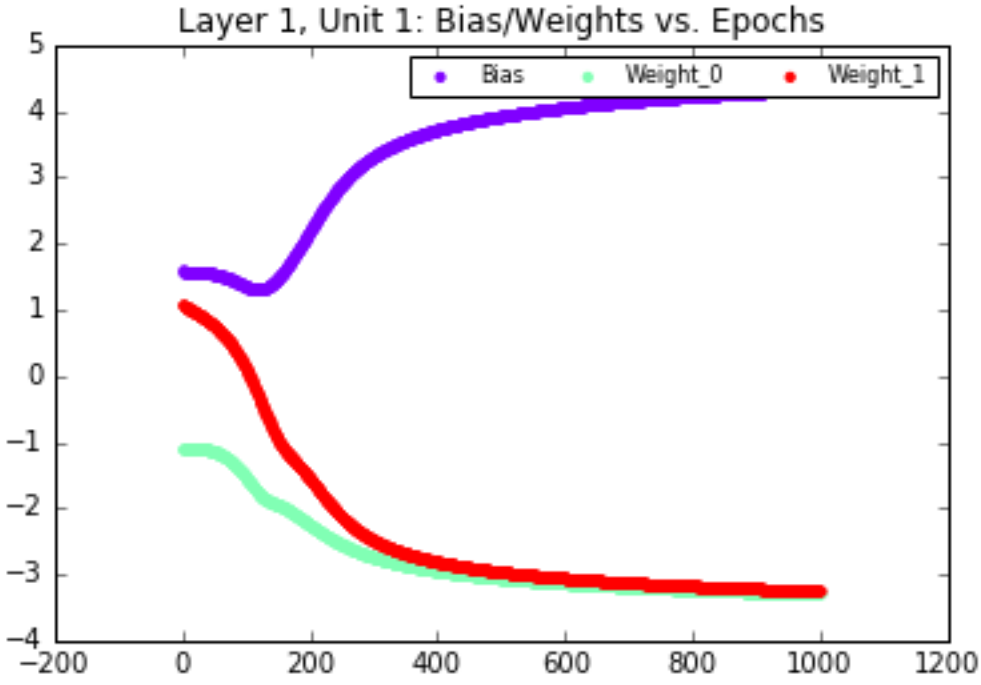

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
10
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

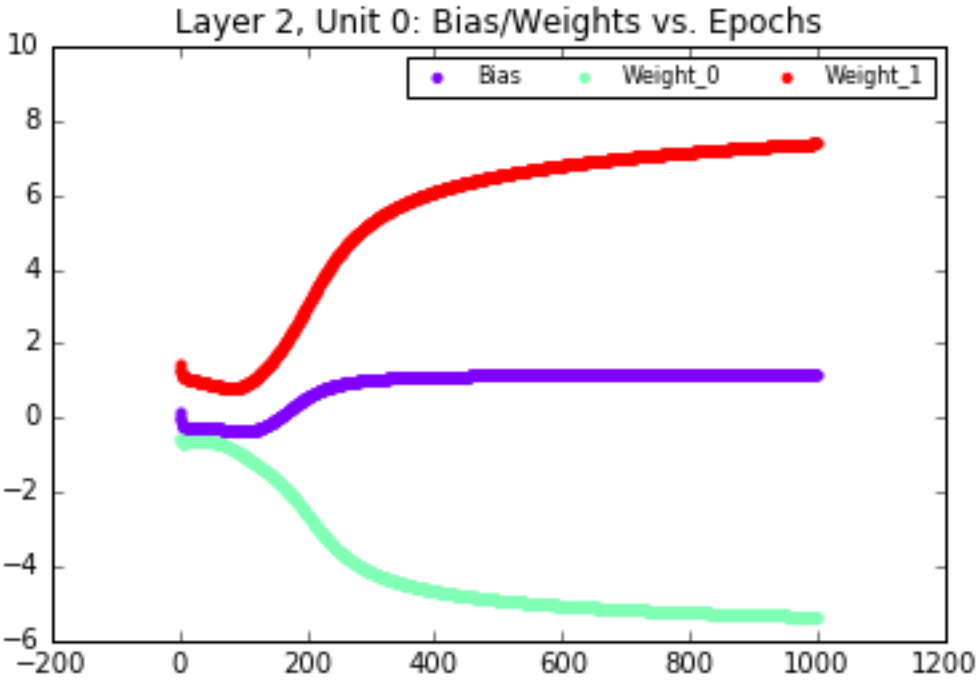

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−200
0
200
400
600
800
1000
1200

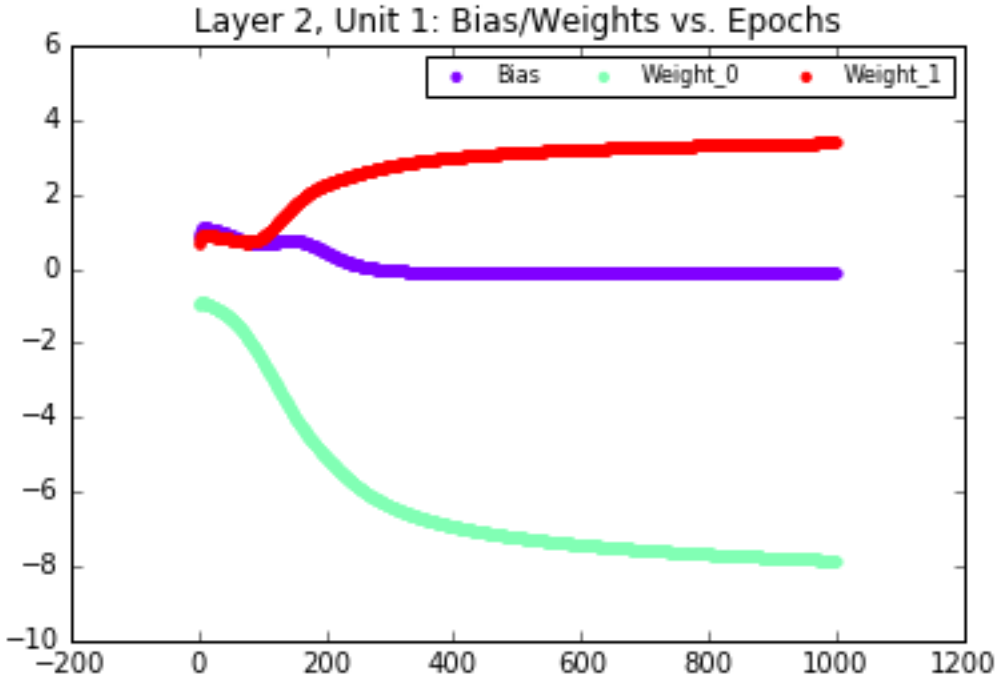

Cost vs. Epochs
C
1.0
0.8
0.6
0.4
0.2
0.0
−200
0
200
400
600
800
1000
1200

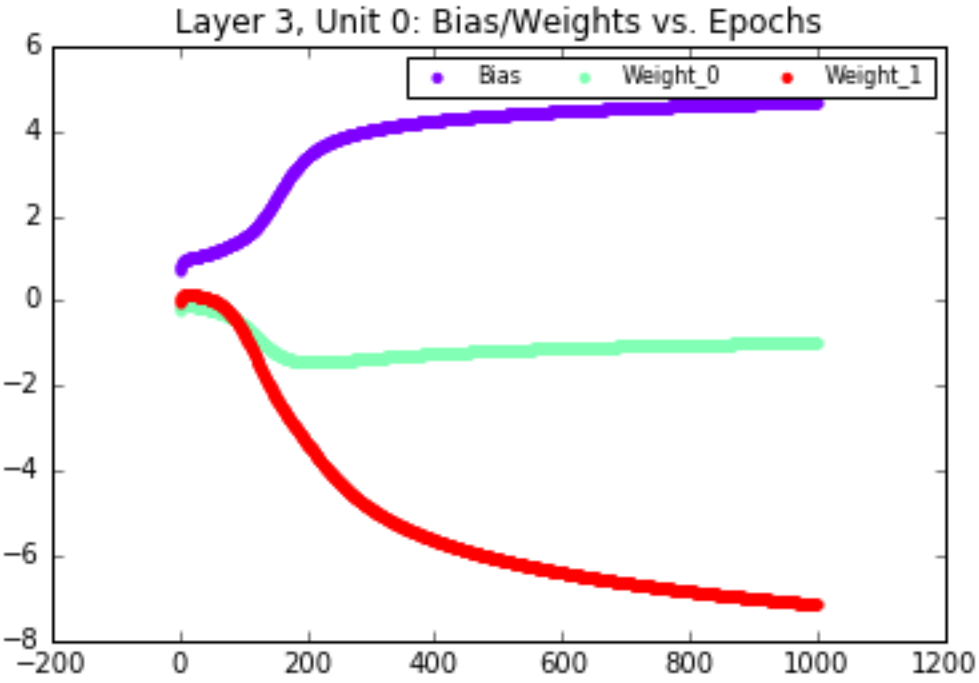
2 input units, 2 hidden layers, first of size 2 second of size 2, 4 output units

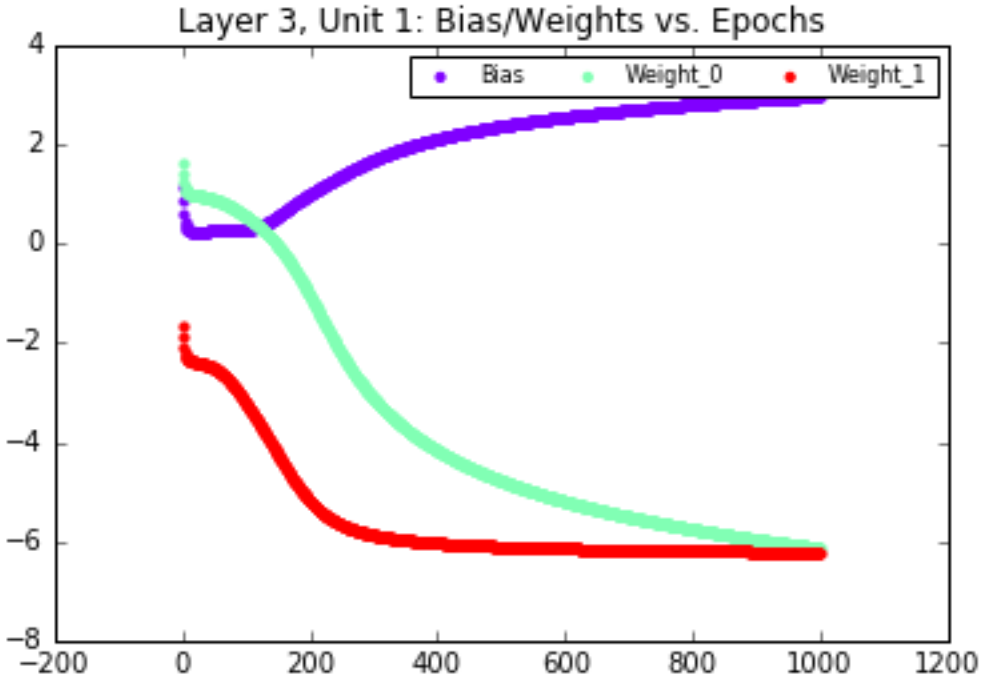


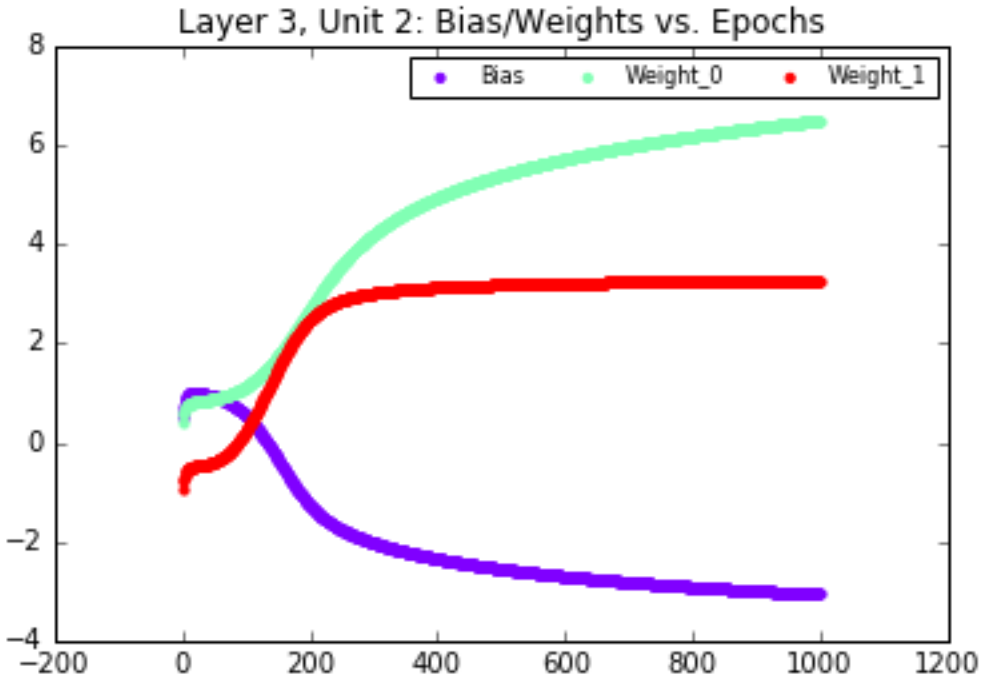


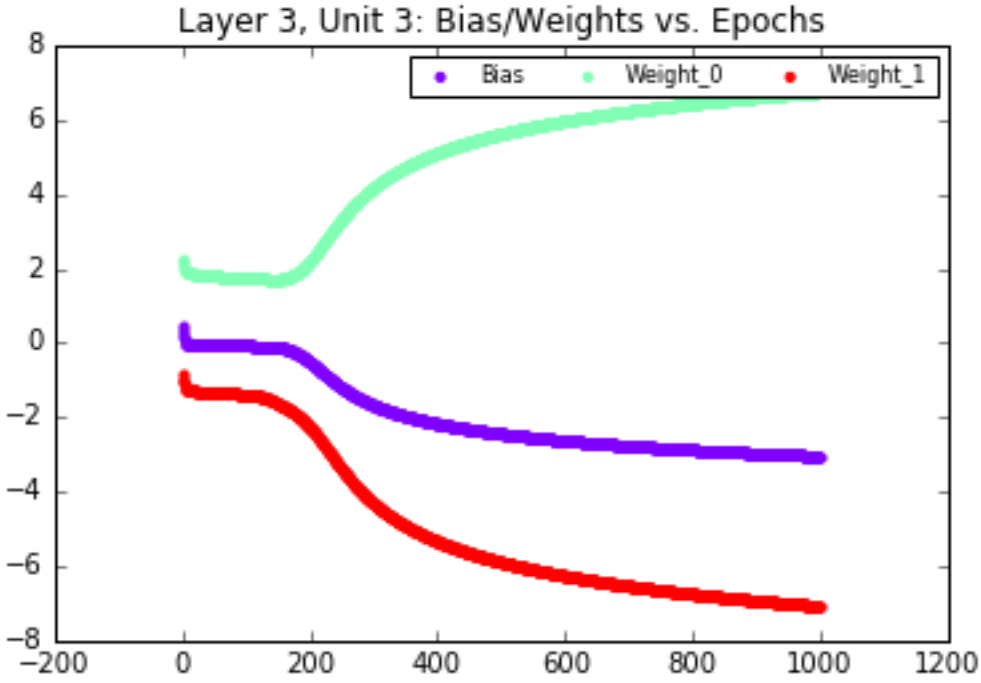

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
10
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

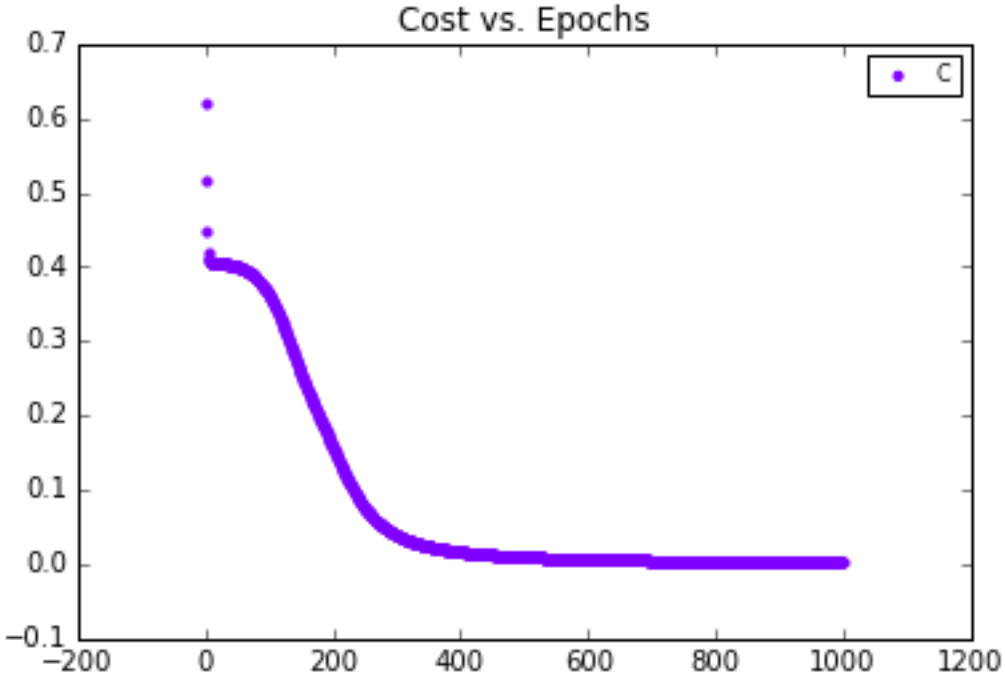

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

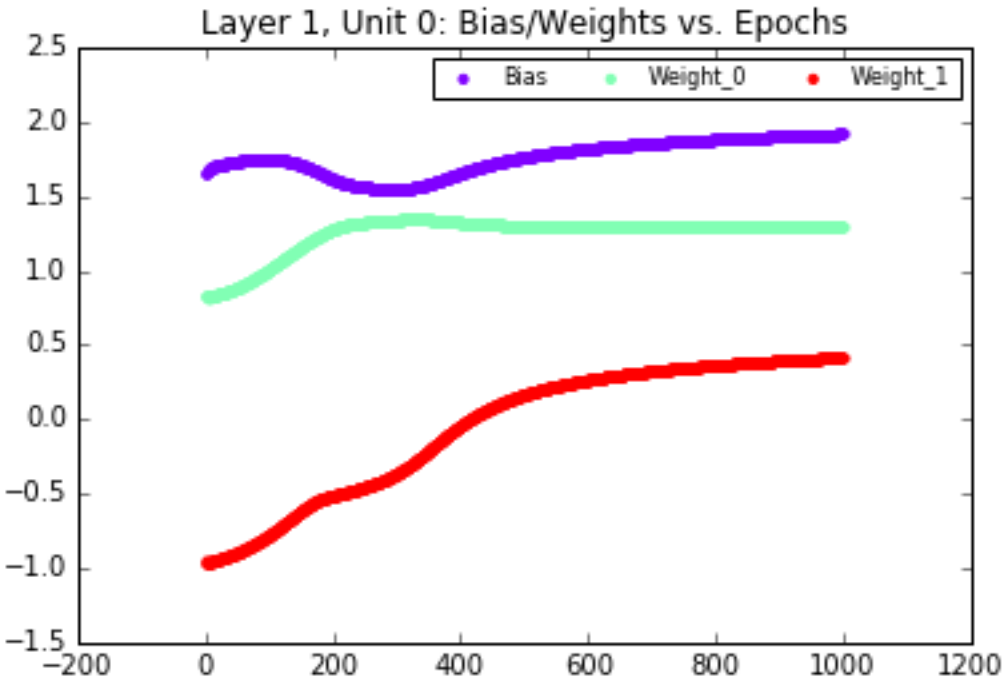

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

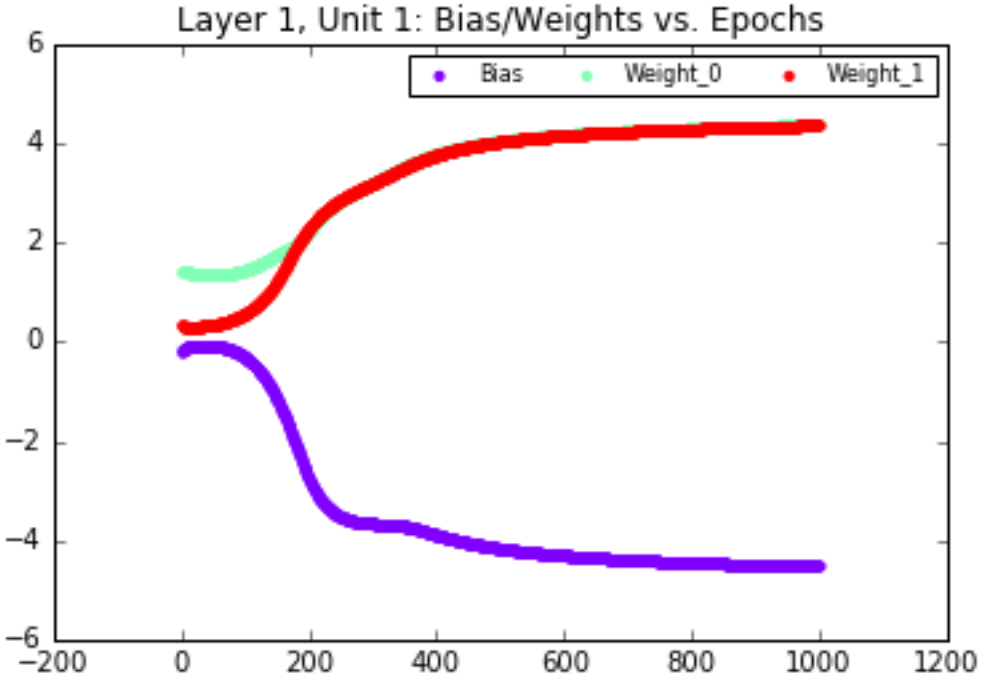

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

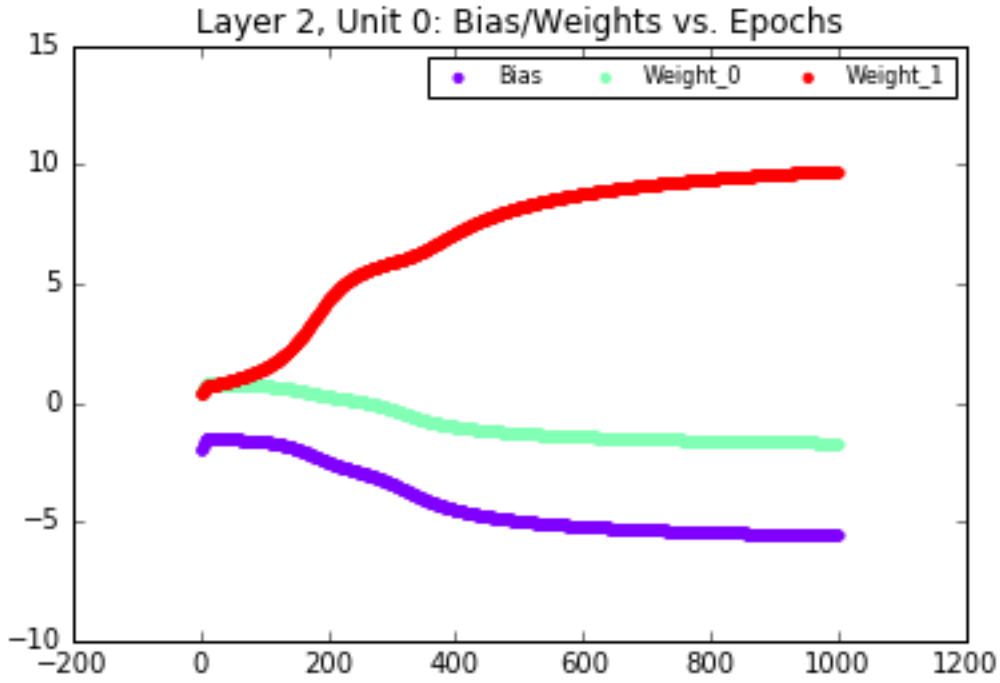

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

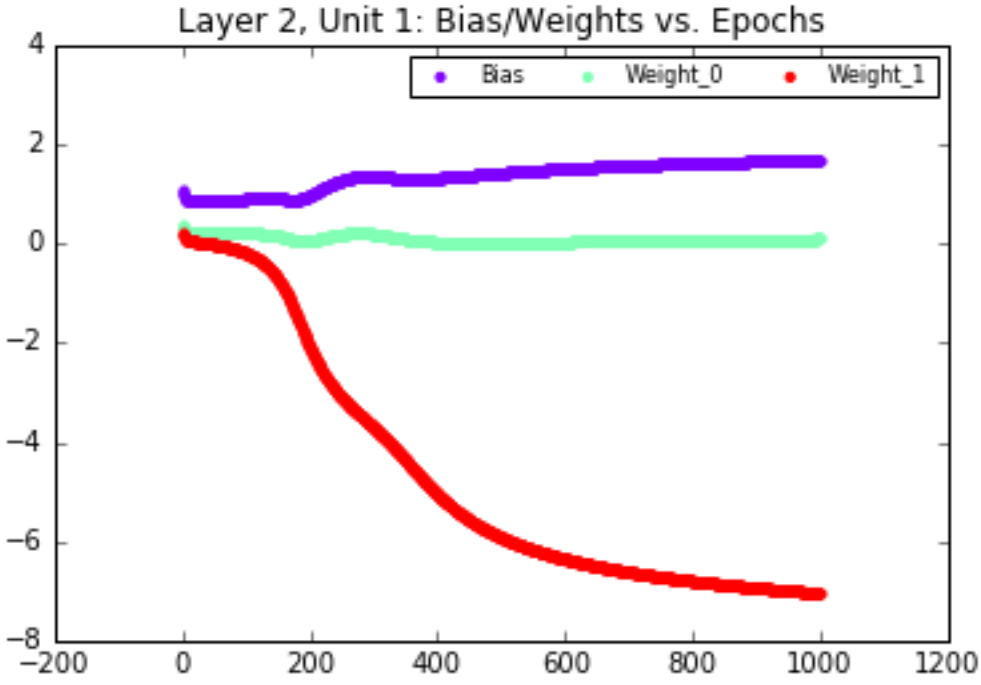

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

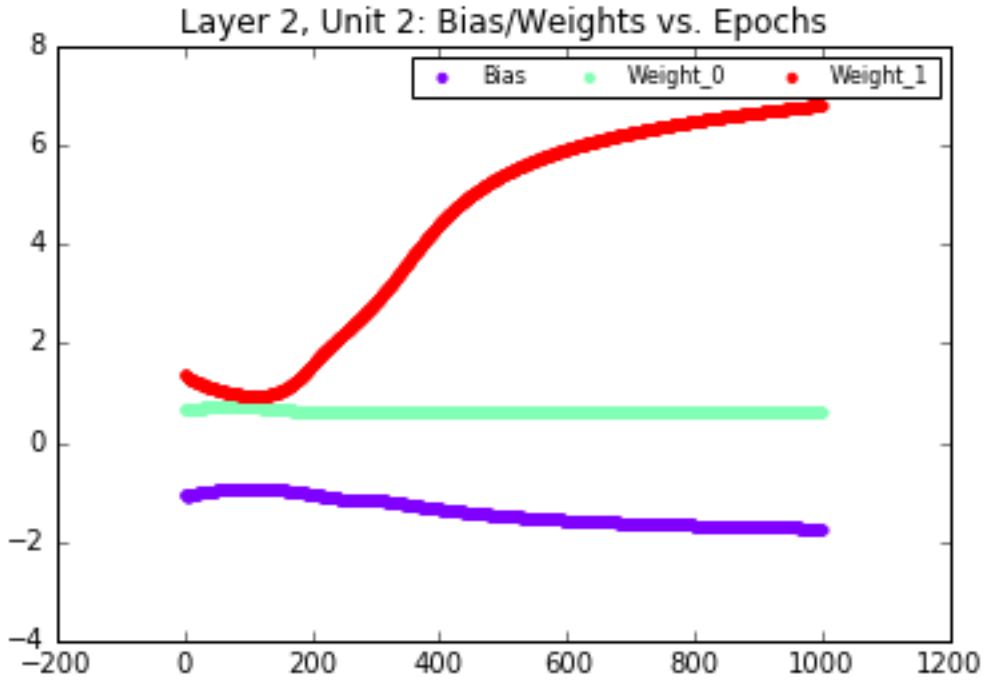
2 input units, 2 hidden layers, first of size 2 second of size 3, 4 output units

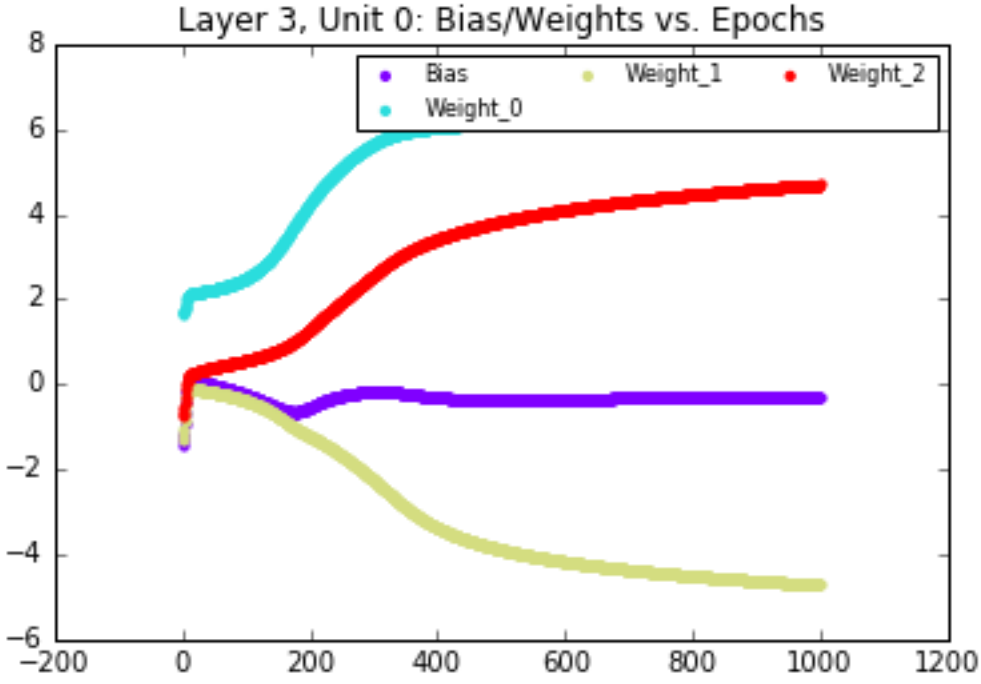


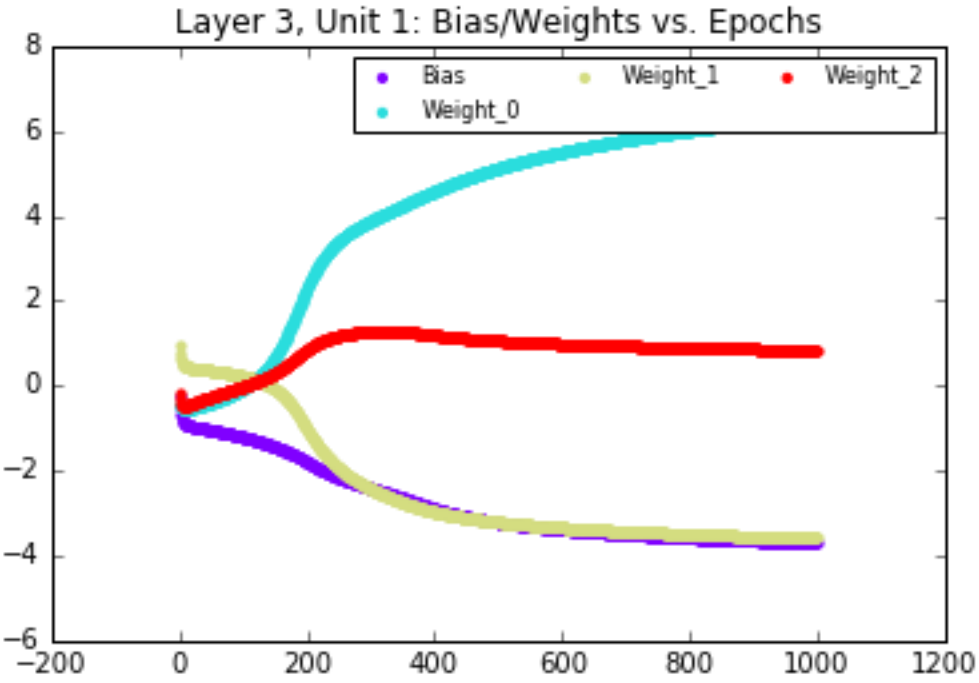


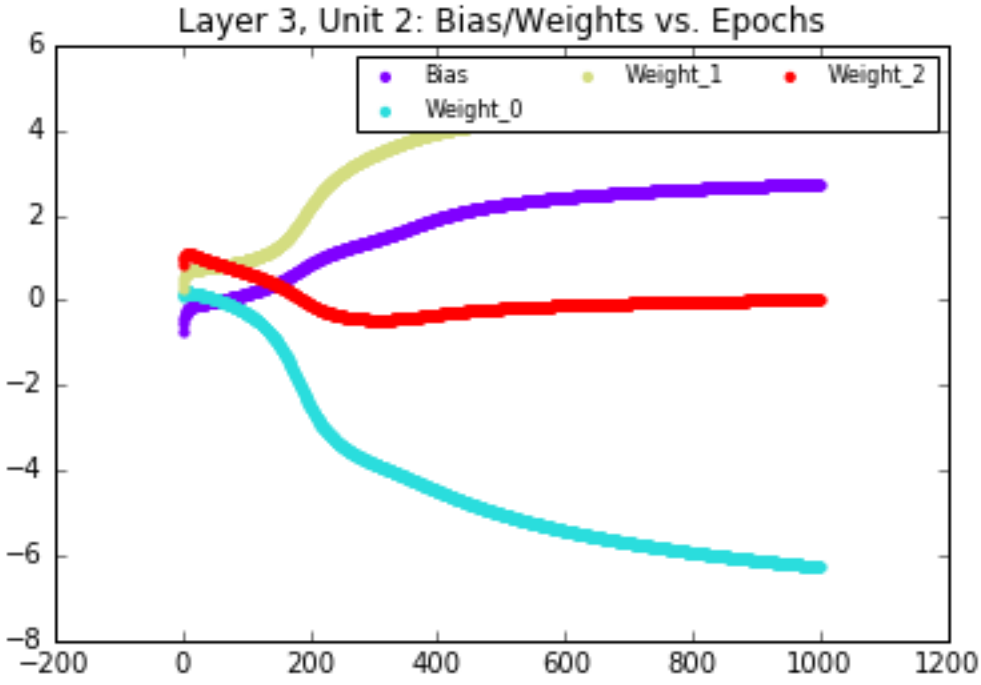

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
15
10
5
0
−5
−10
−200
0
200
400
600
800
1000
1200

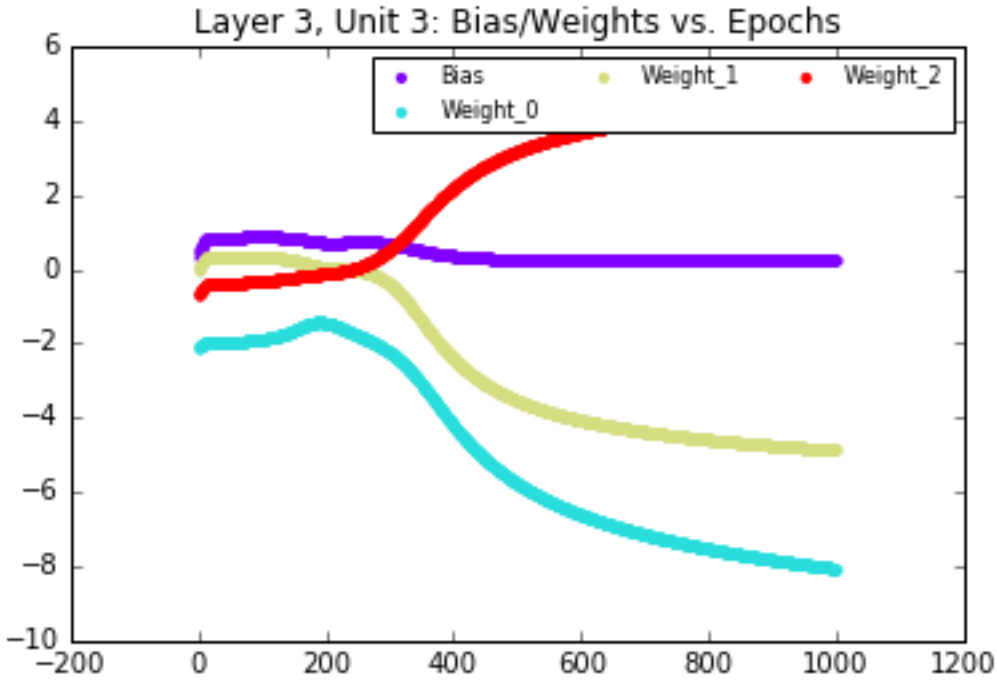

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

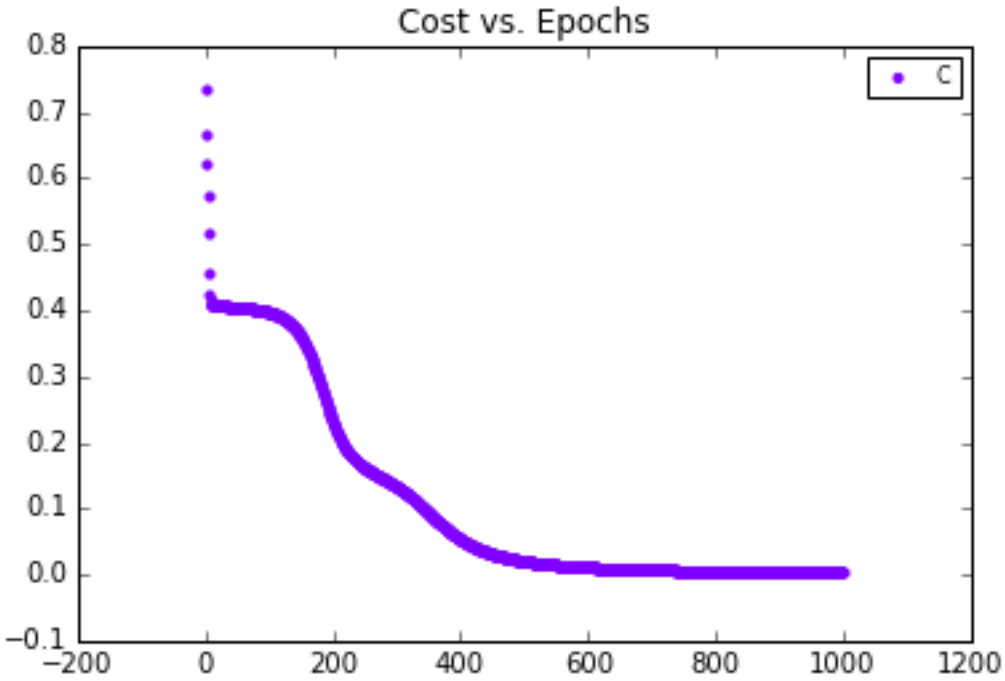

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

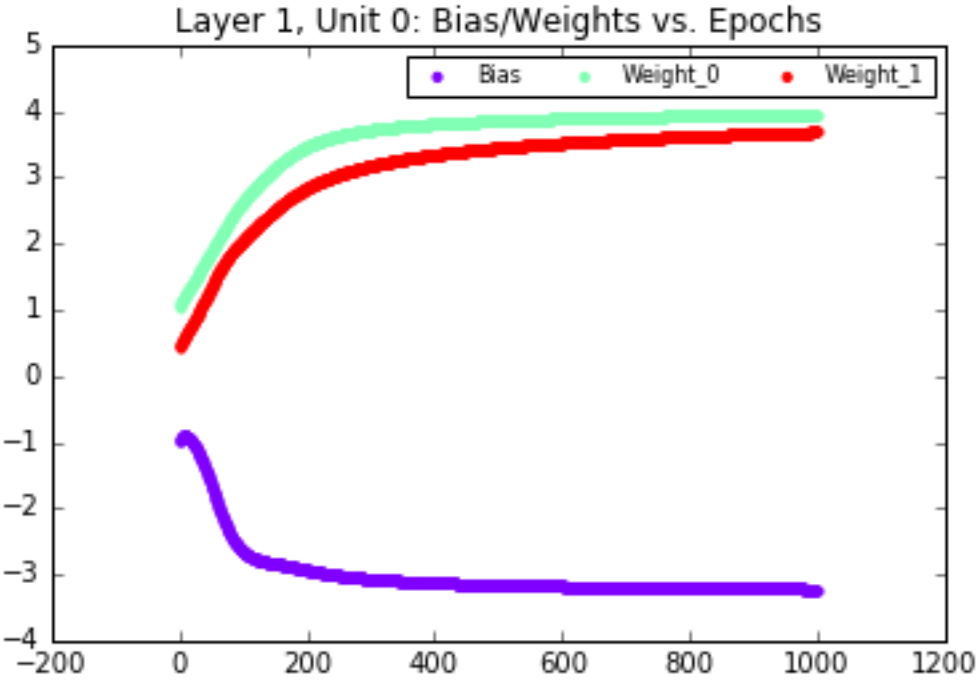

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

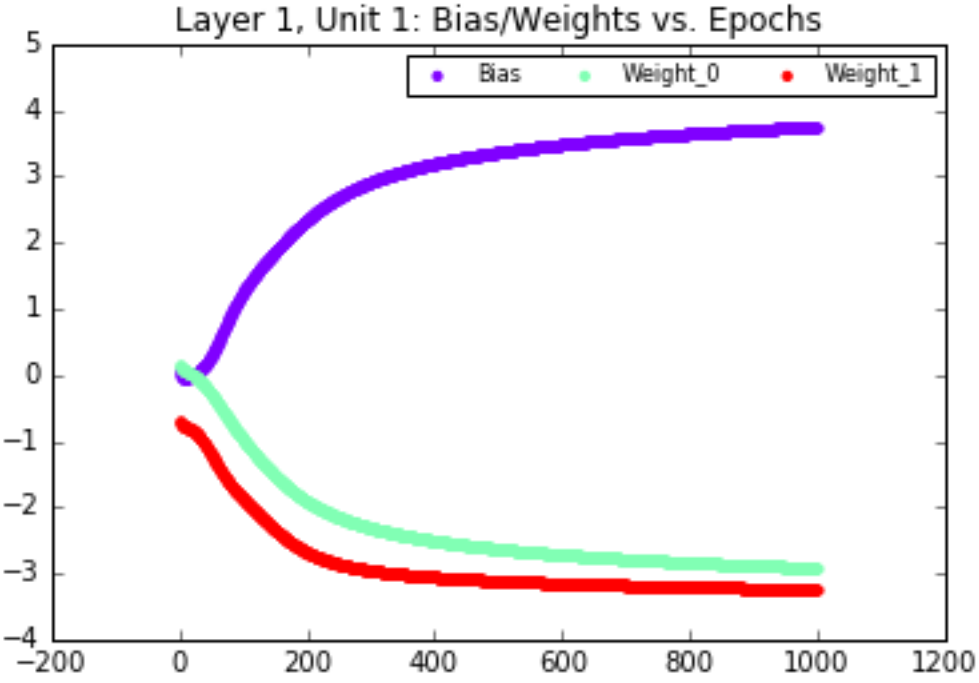

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

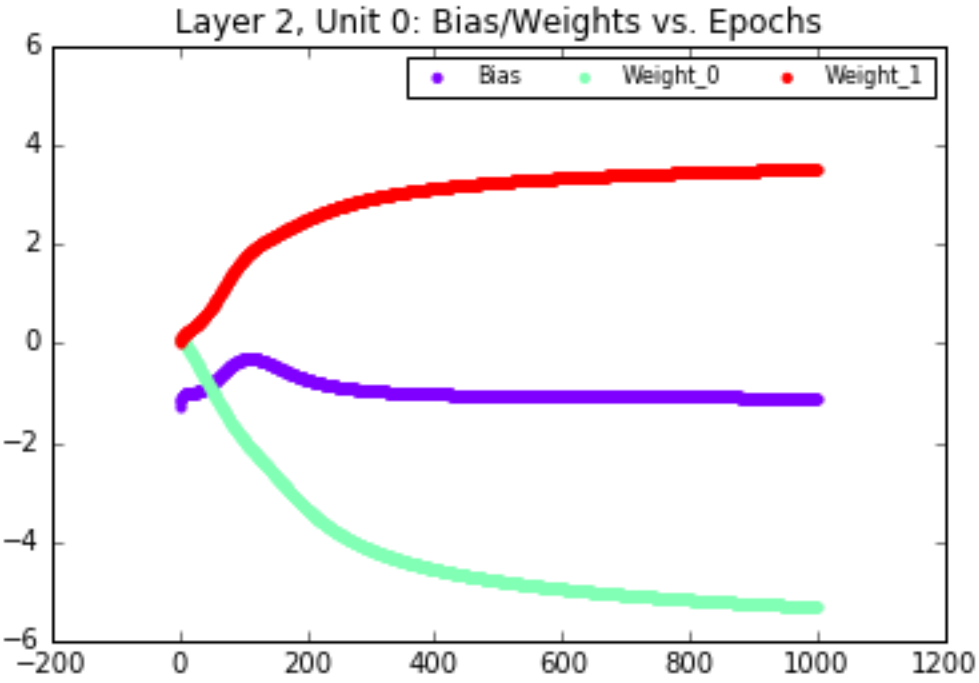

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

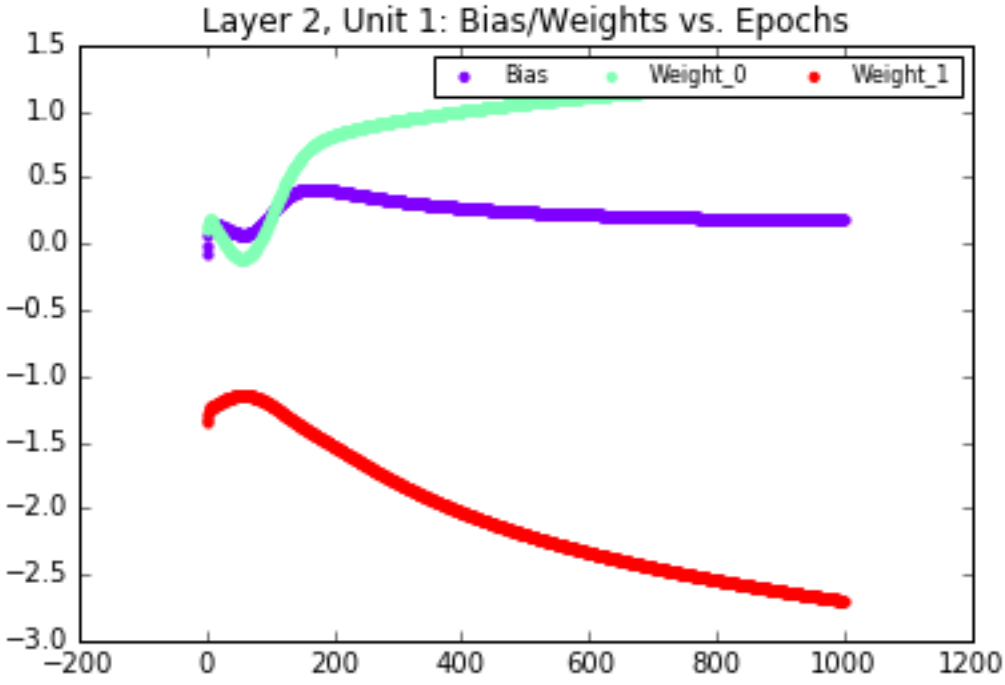

Cost vs. Epochs
C
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

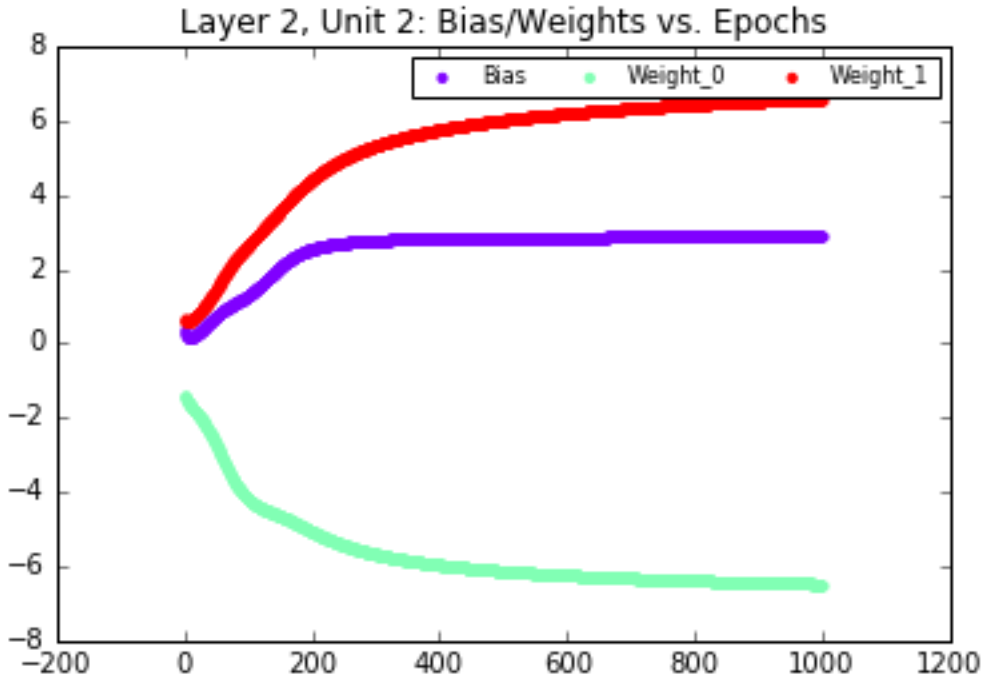
2 input units, 2 hidden layers, first of size 2 second of size 4, 4 output units

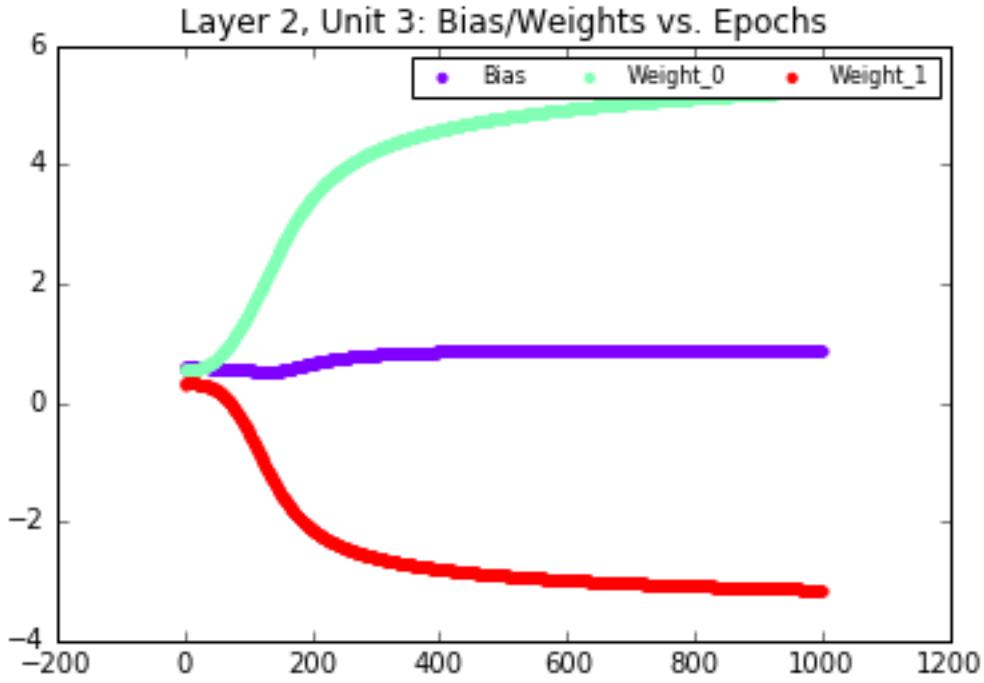


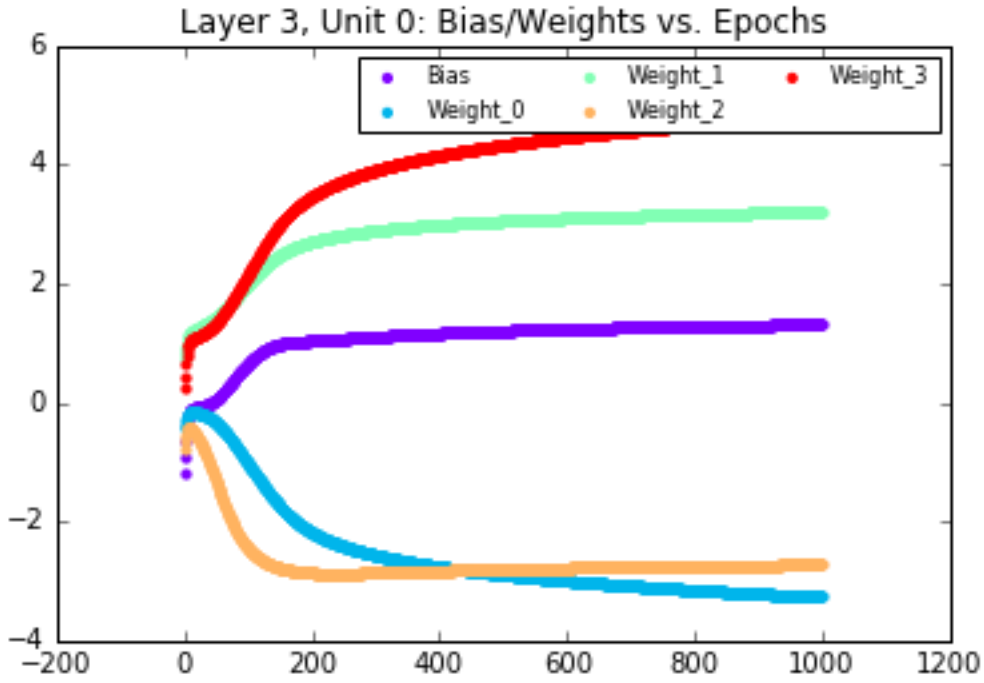


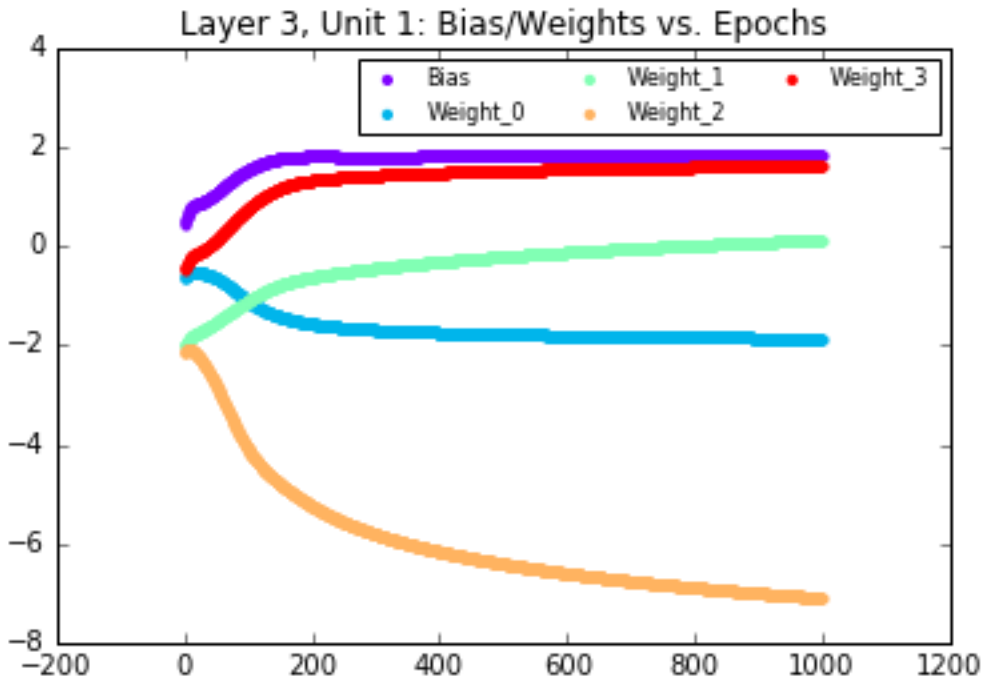

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

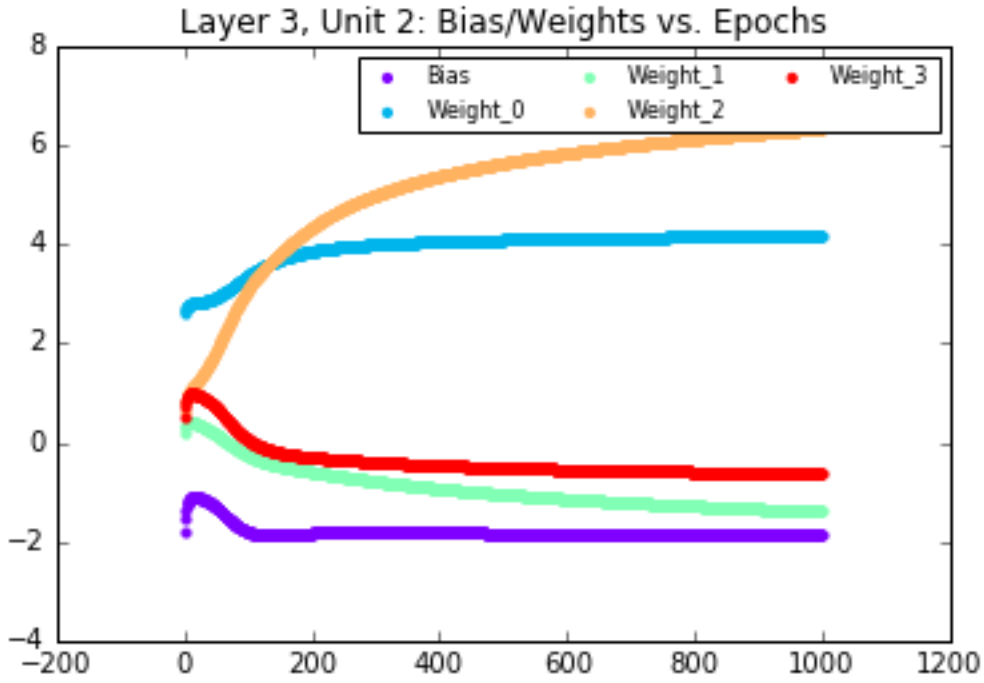

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−200
0
200
400
600
800
1000
1200

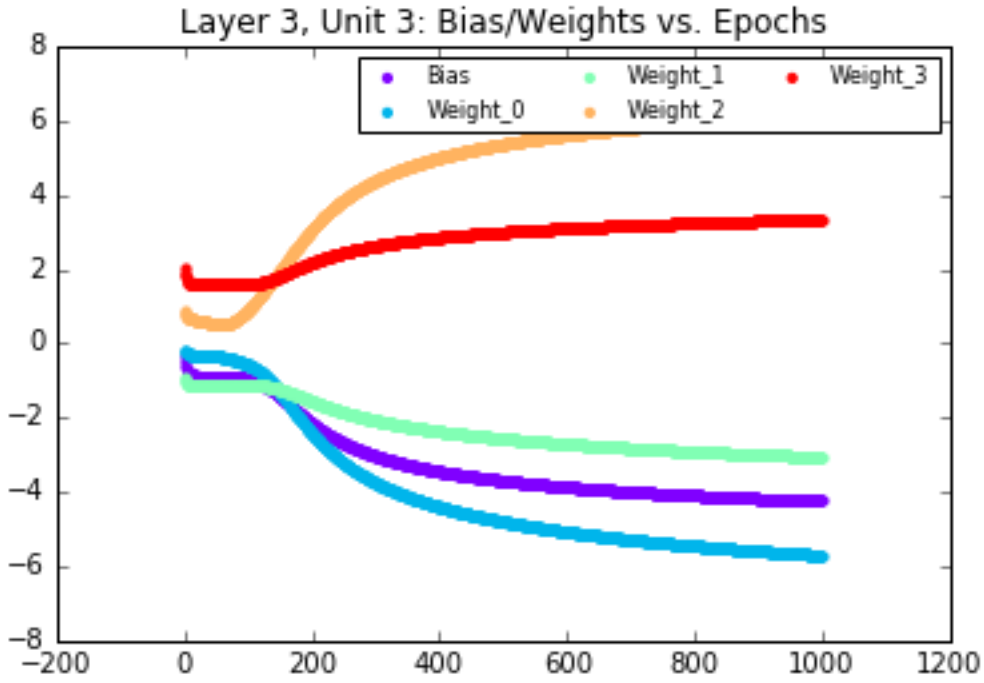

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

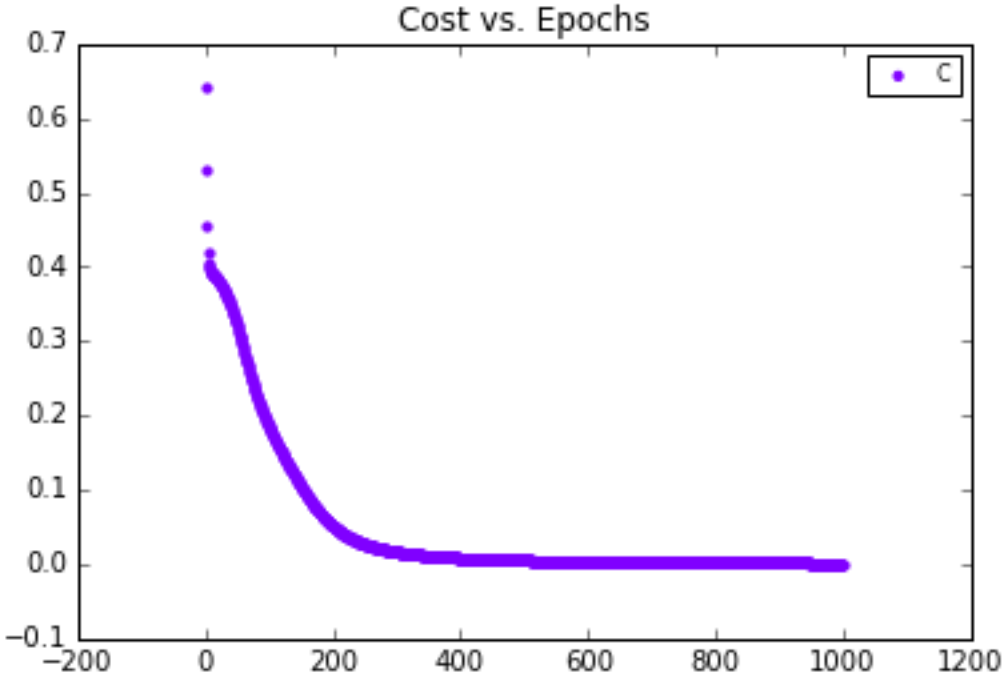

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

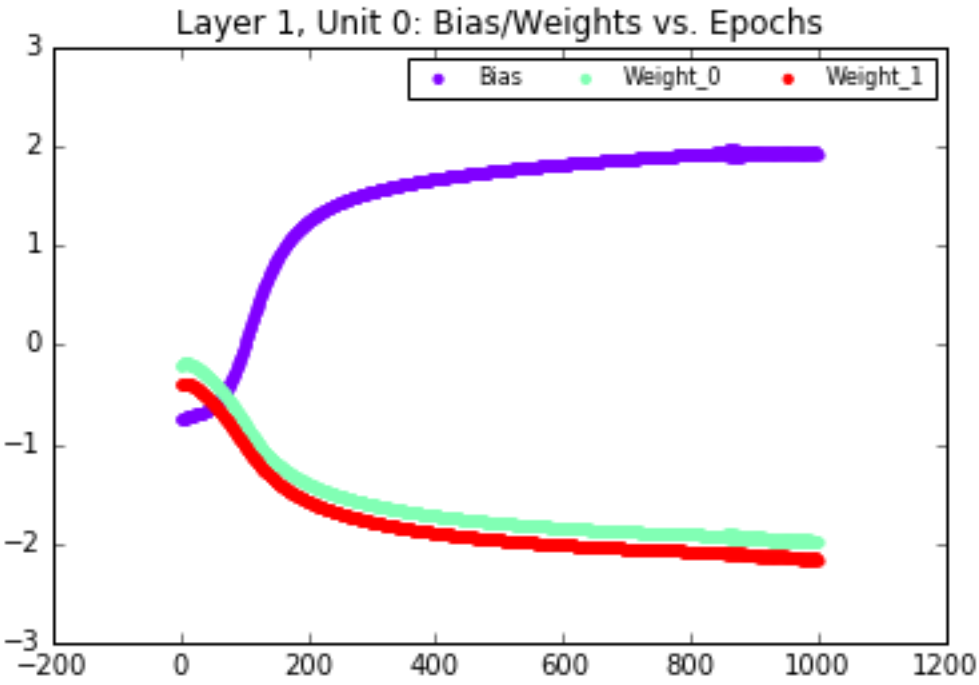

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

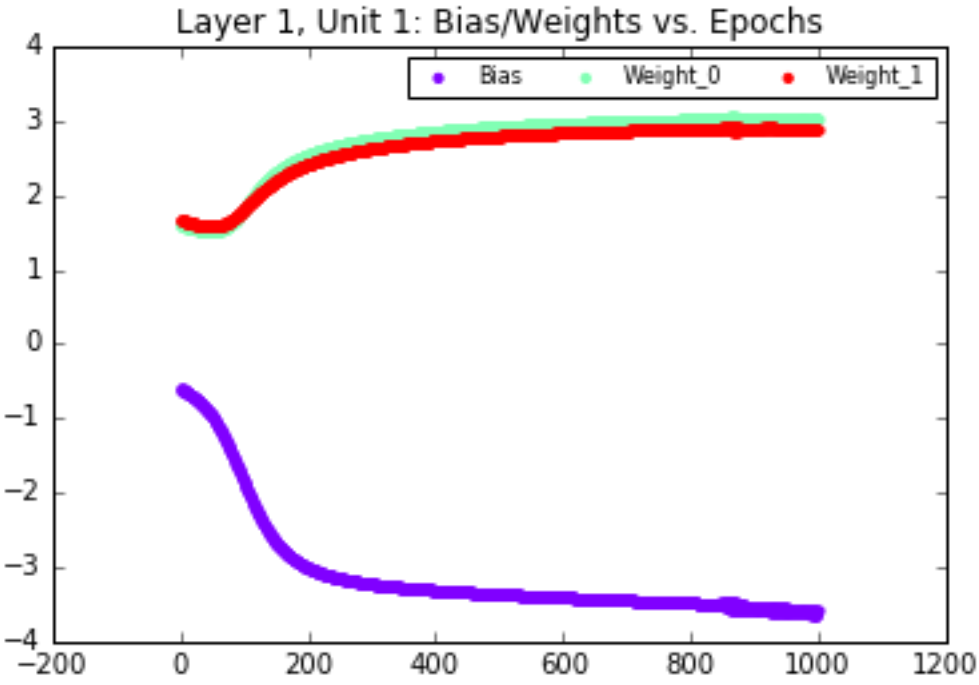

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

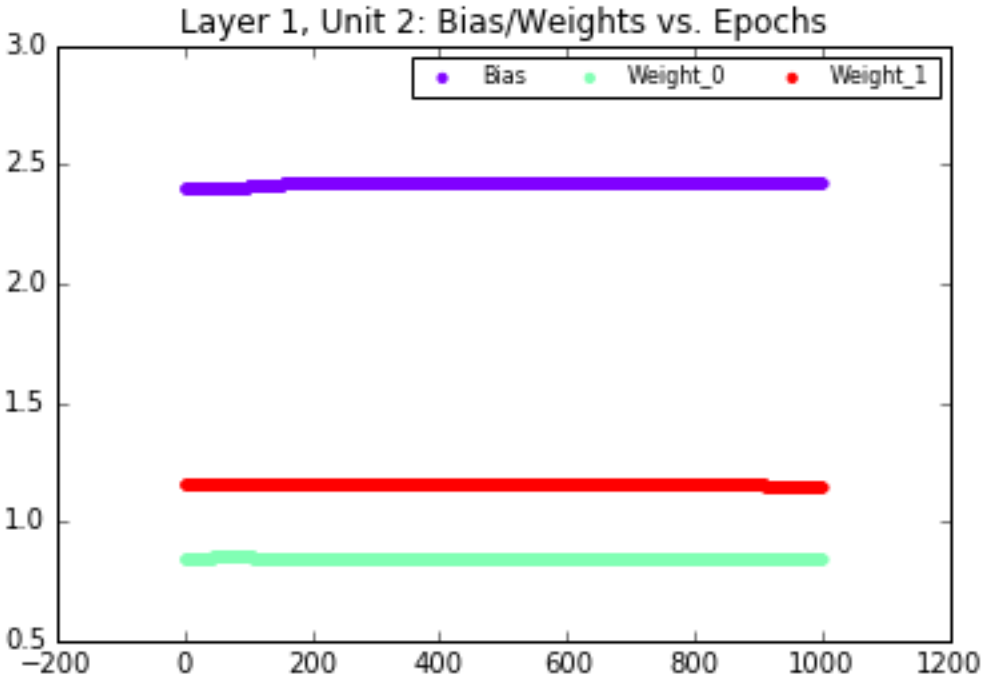

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

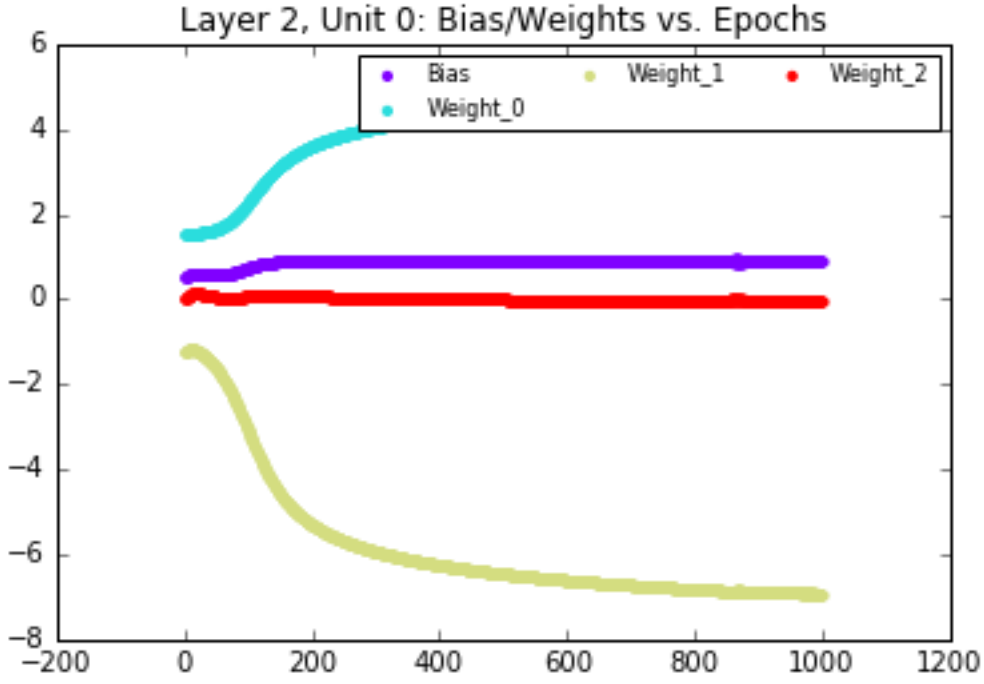

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

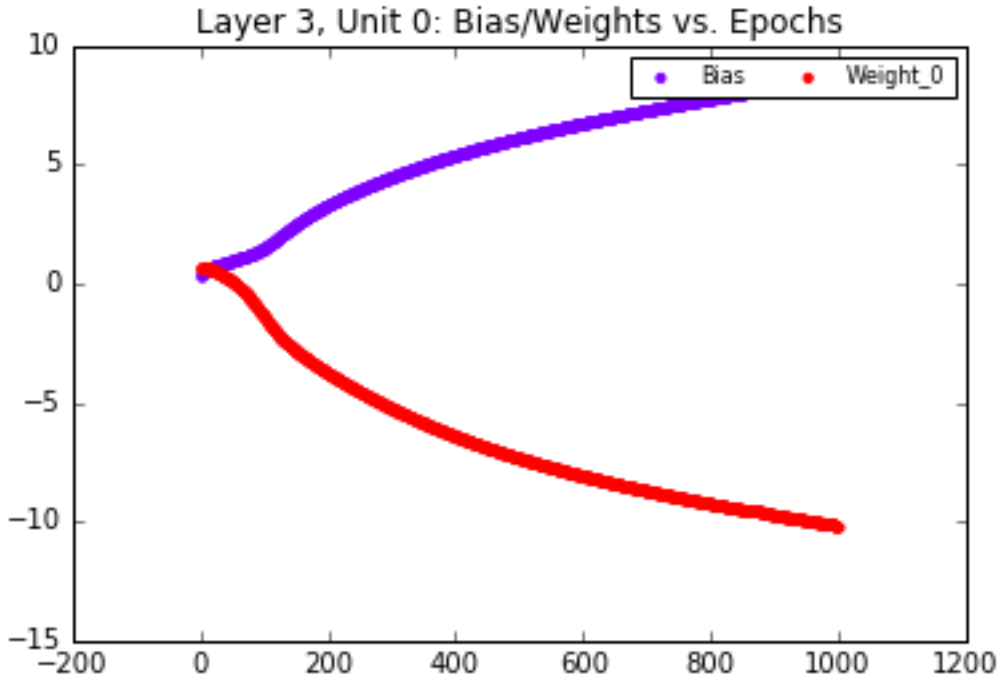
2 input units, 2 hidden layers, first of size 3, second of size 1, 4 output units

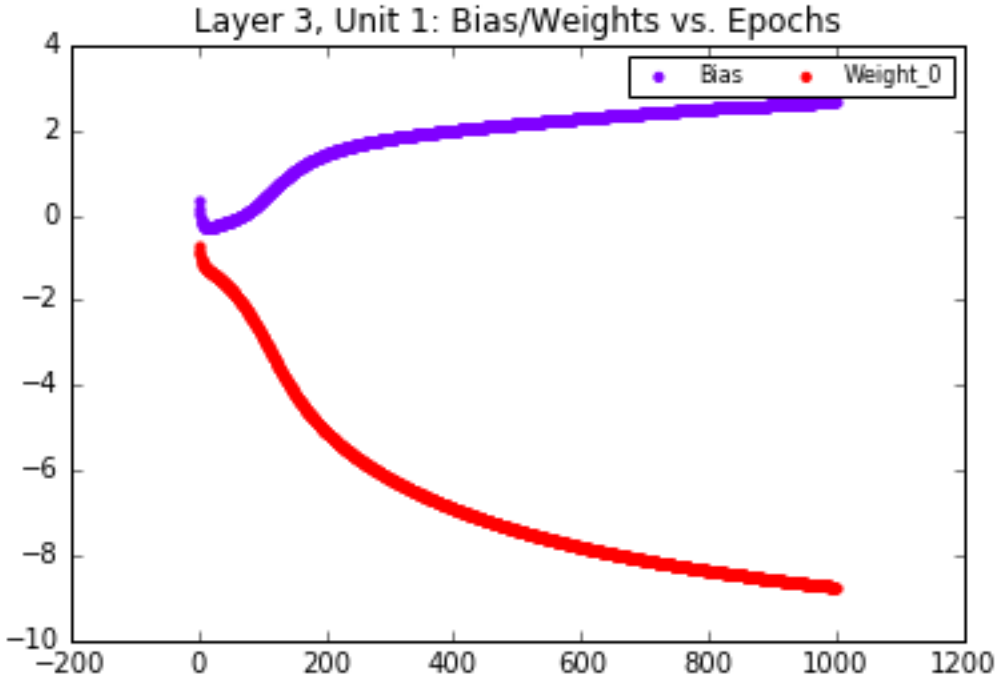


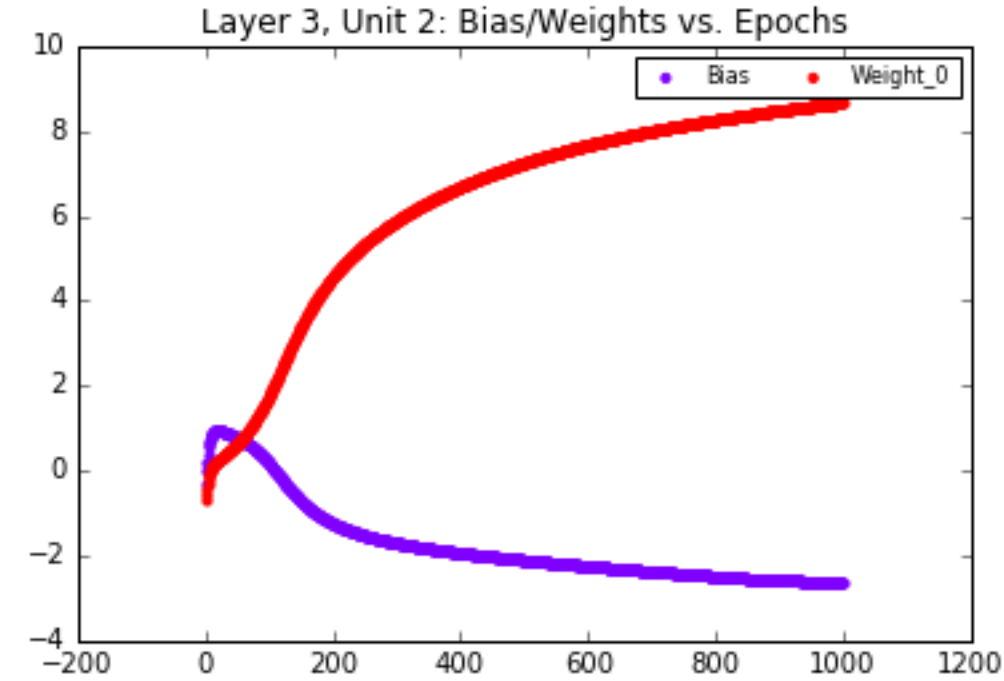


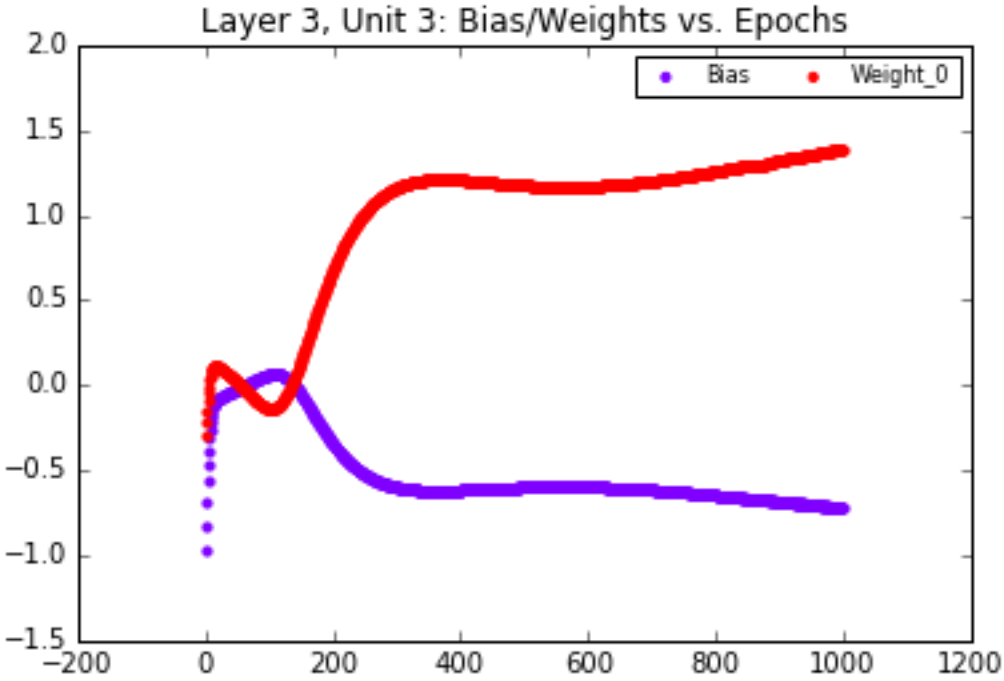

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
3.0
2.5
2.0
1.5
1.0
0.5
−200
0
200
400
600
800
1000
1200

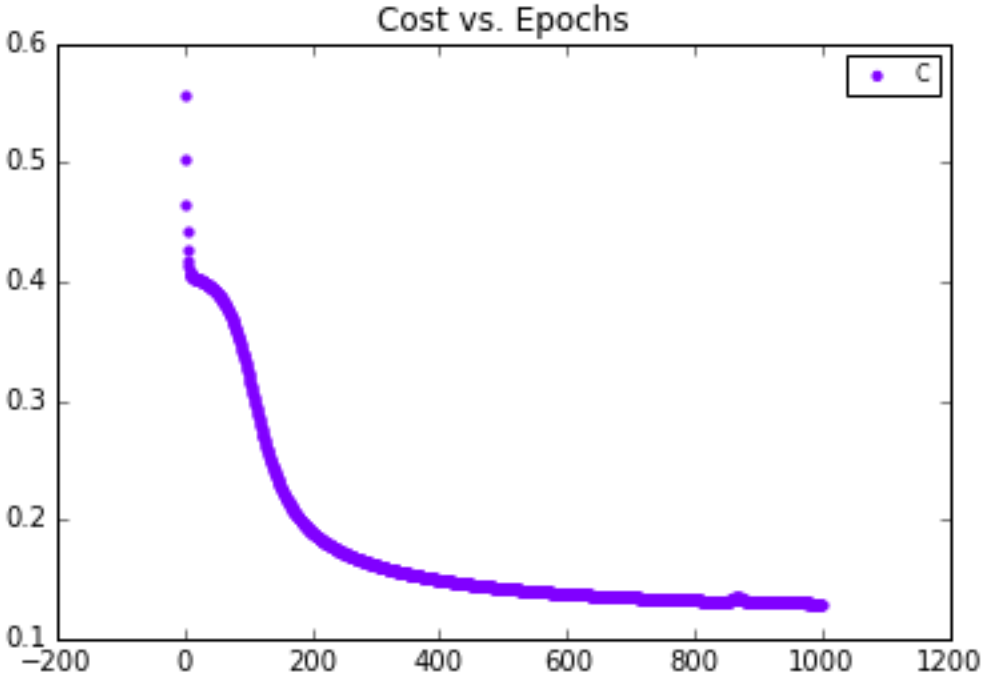

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
10
5
0
−5
−10
−15
−200
0
200
400
600
800
1000
1200

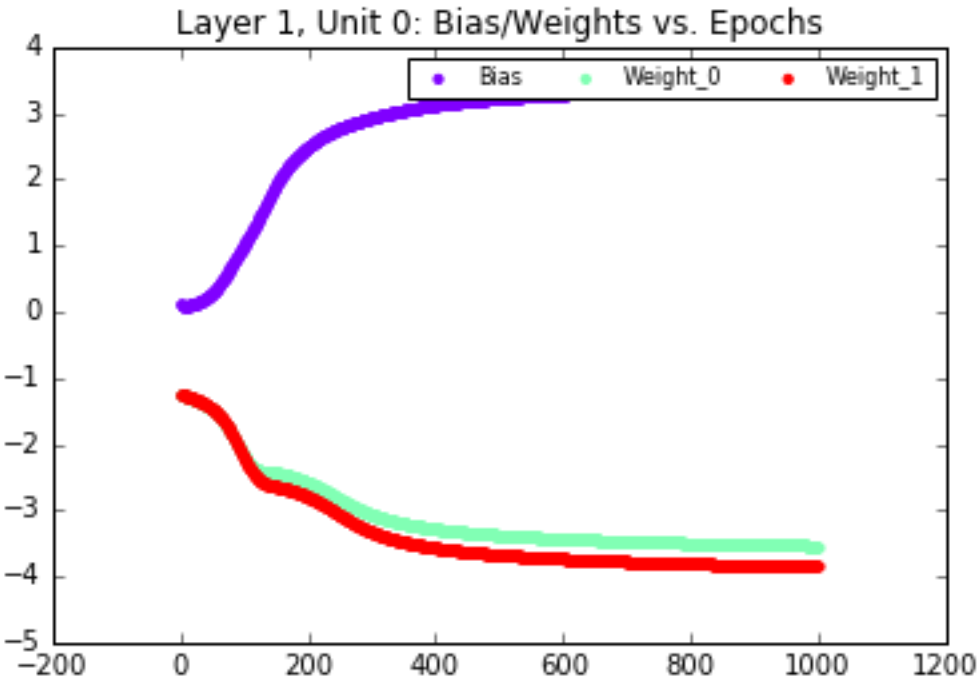

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

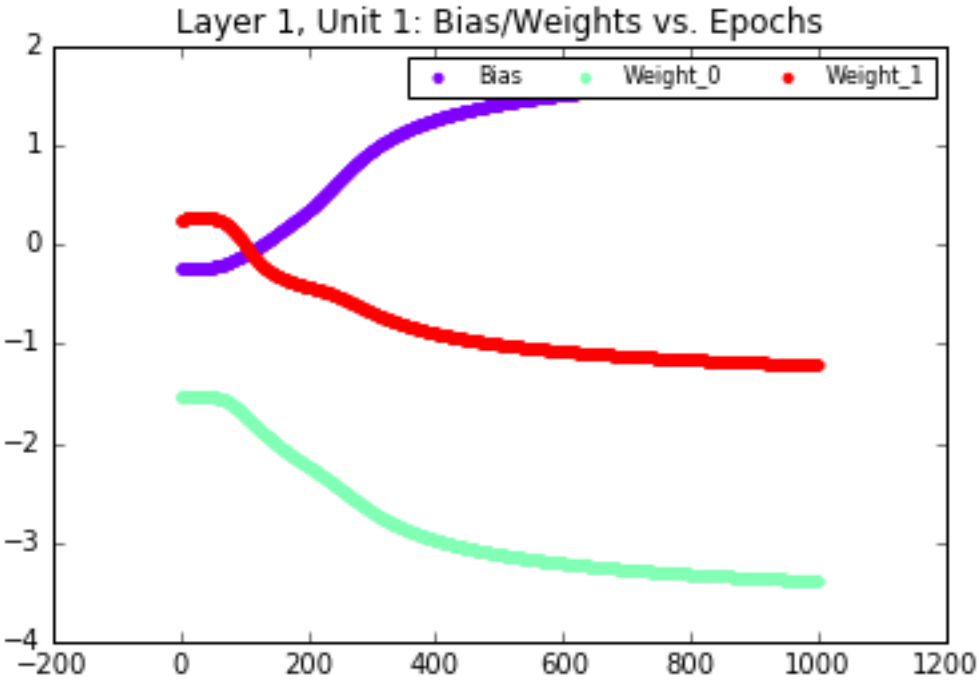

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
10
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

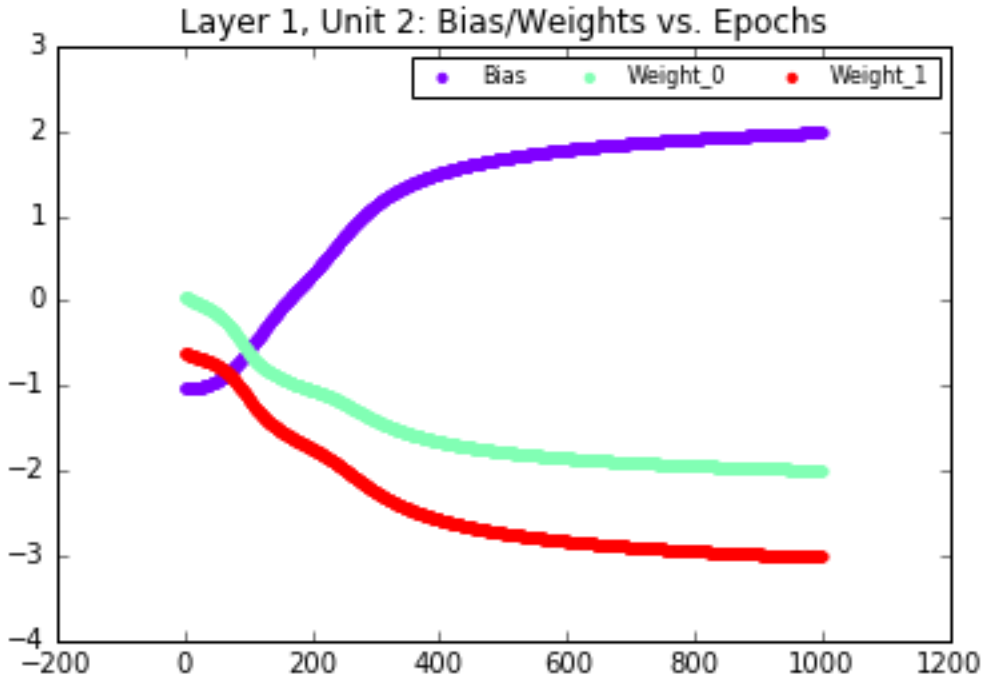

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−200
0
200
400
600
800
1000
1200

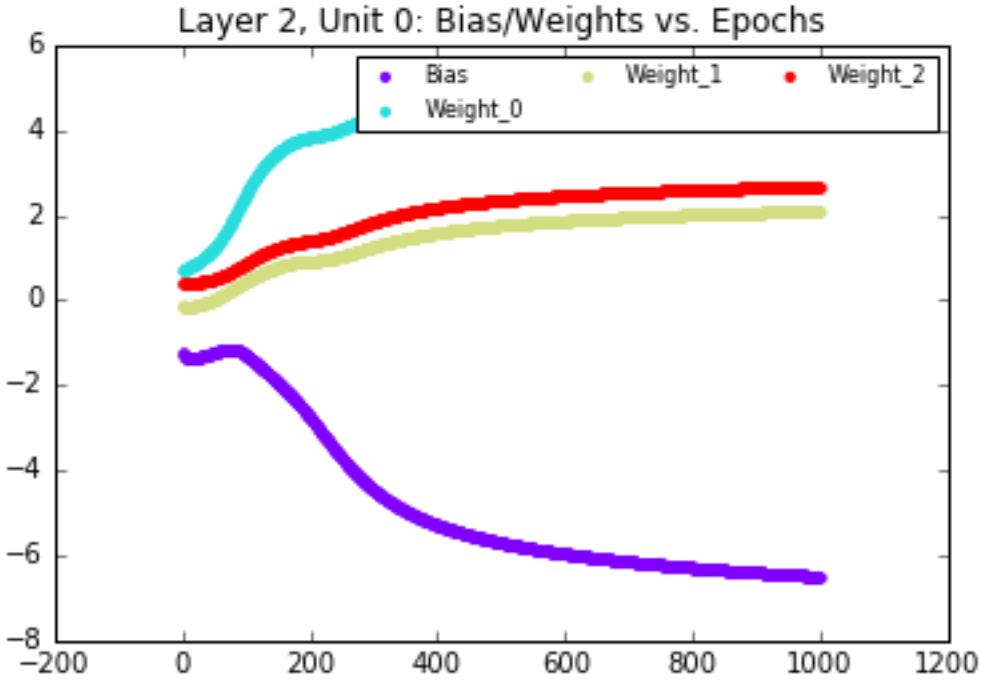

Cost vs. Epochs
C
0.6
0.5
0.4
0.3
0.2
0.1
−200
0
200
400
600
800
1000
1200

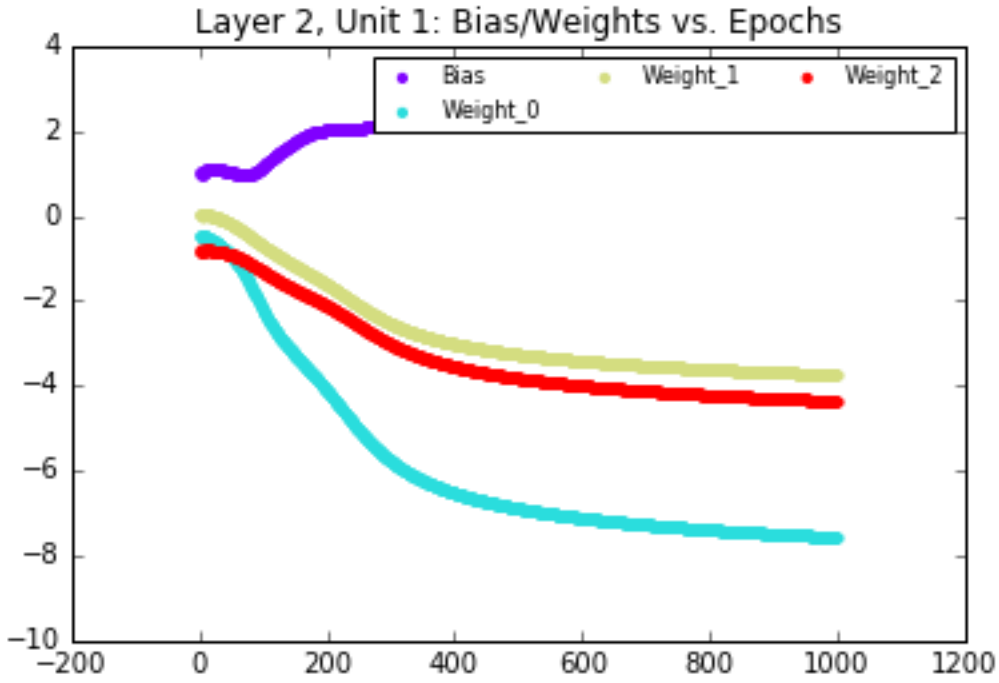
2 input units, 2 hidden layers, first of size 3, second of size 2, 4 output units

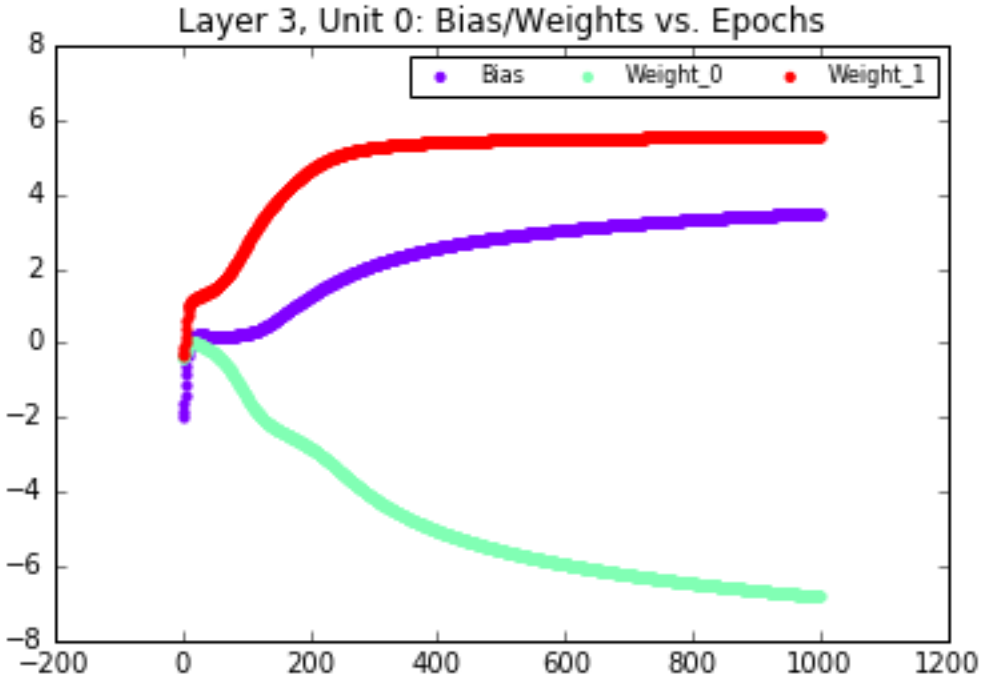


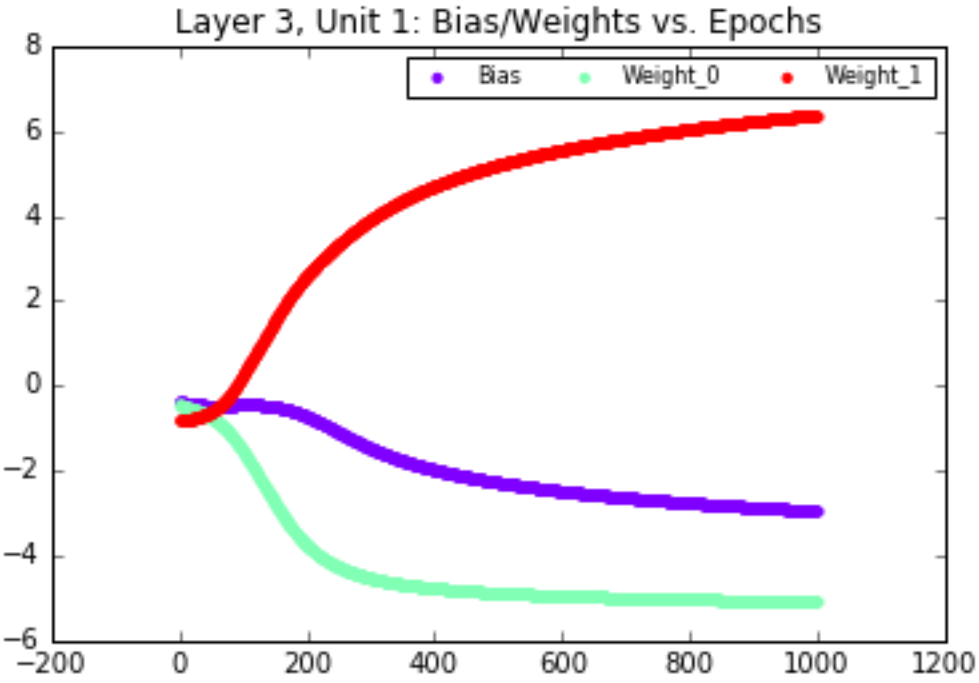


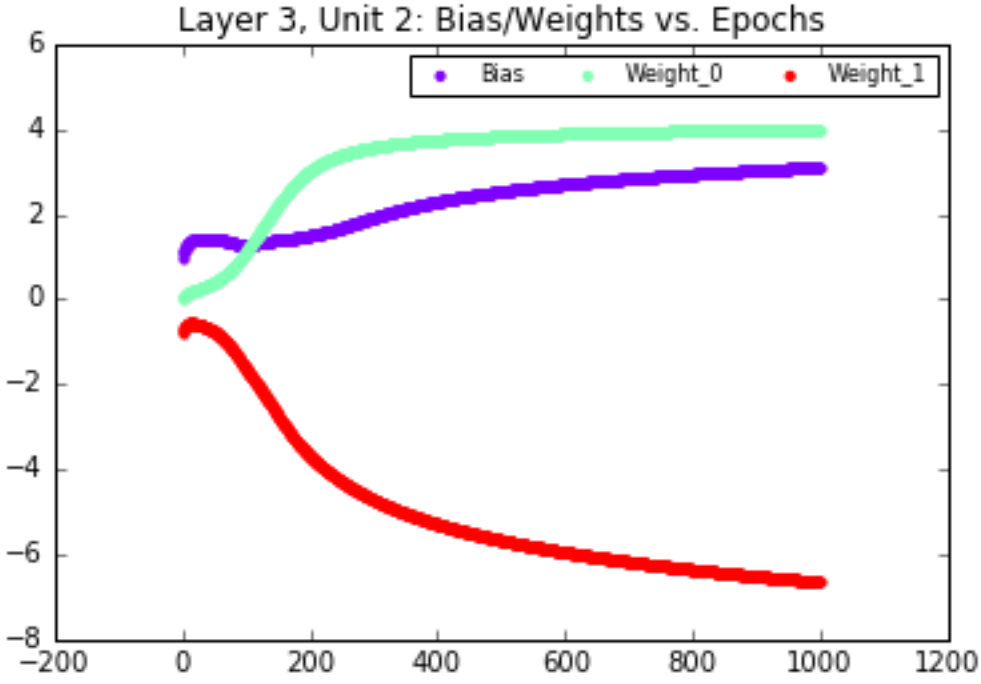

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
3
2
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

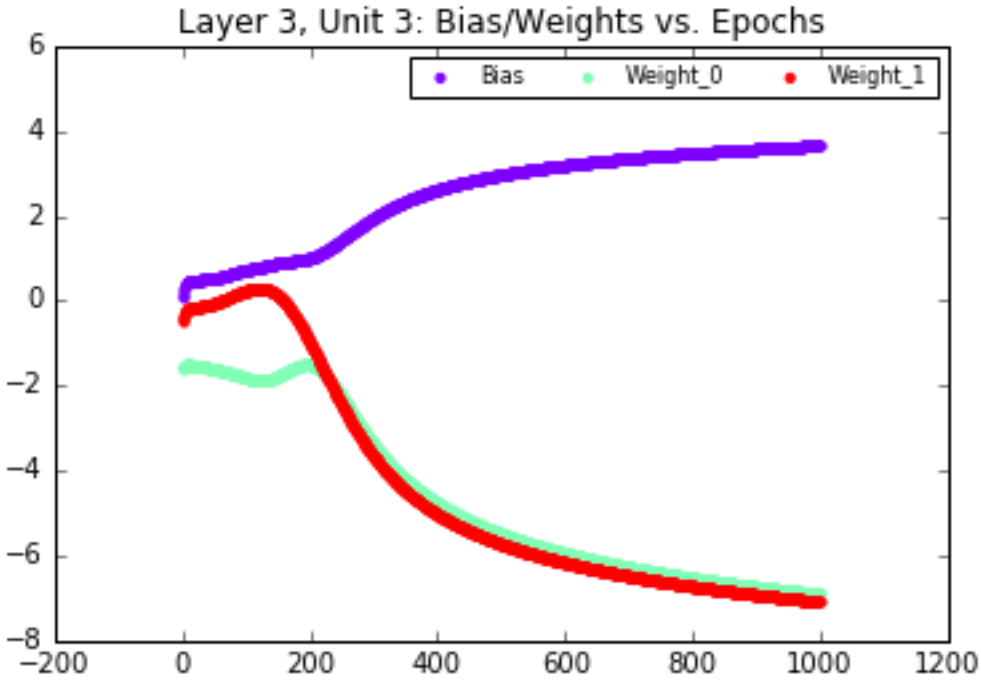

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

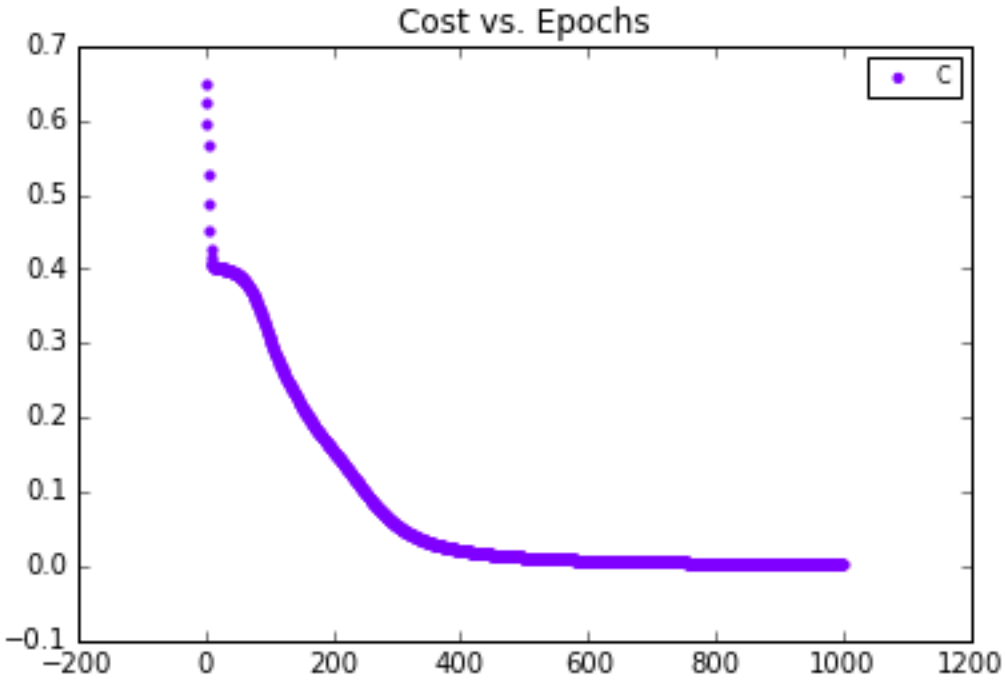

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

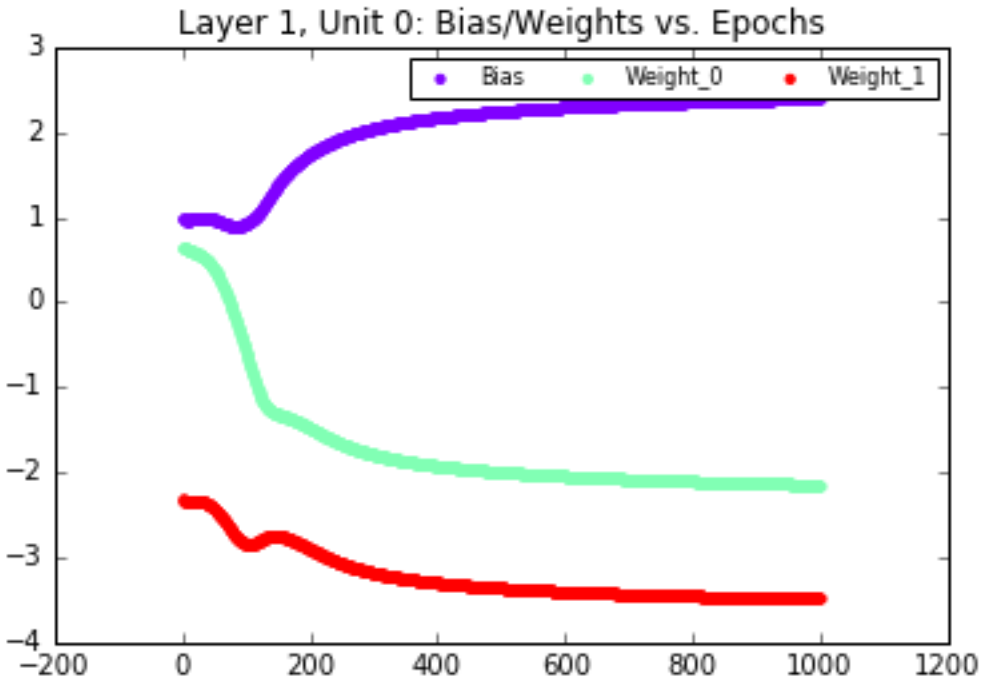

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

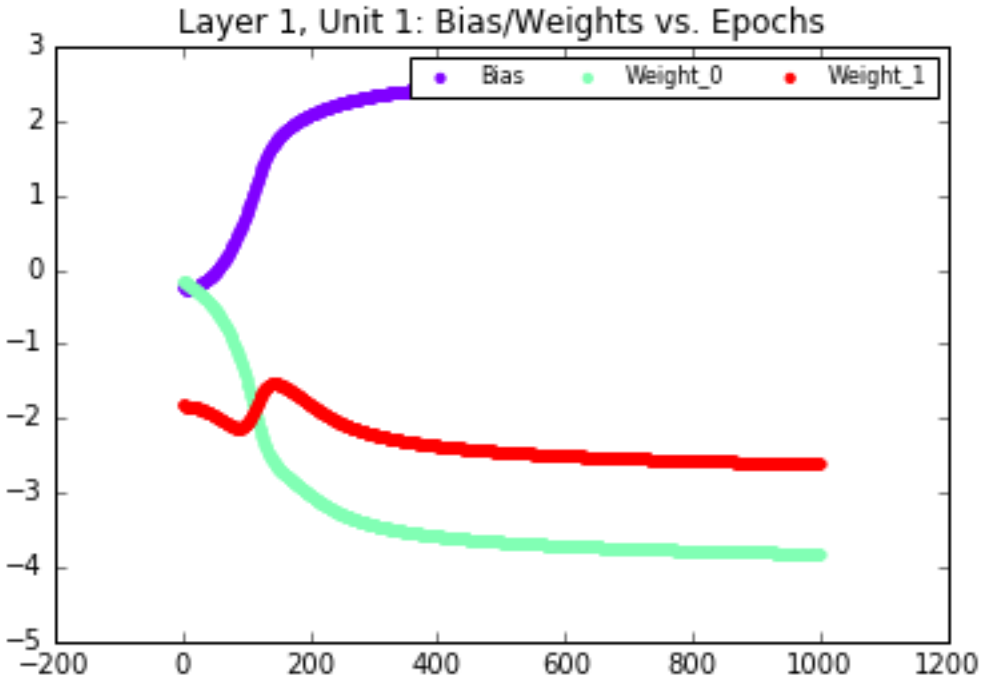

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

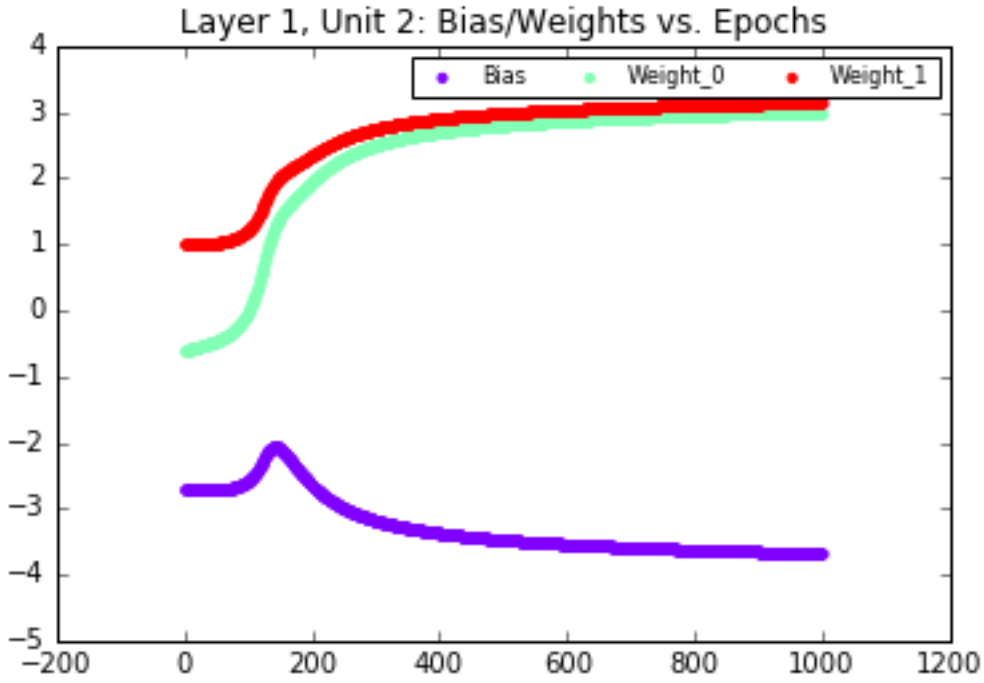

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

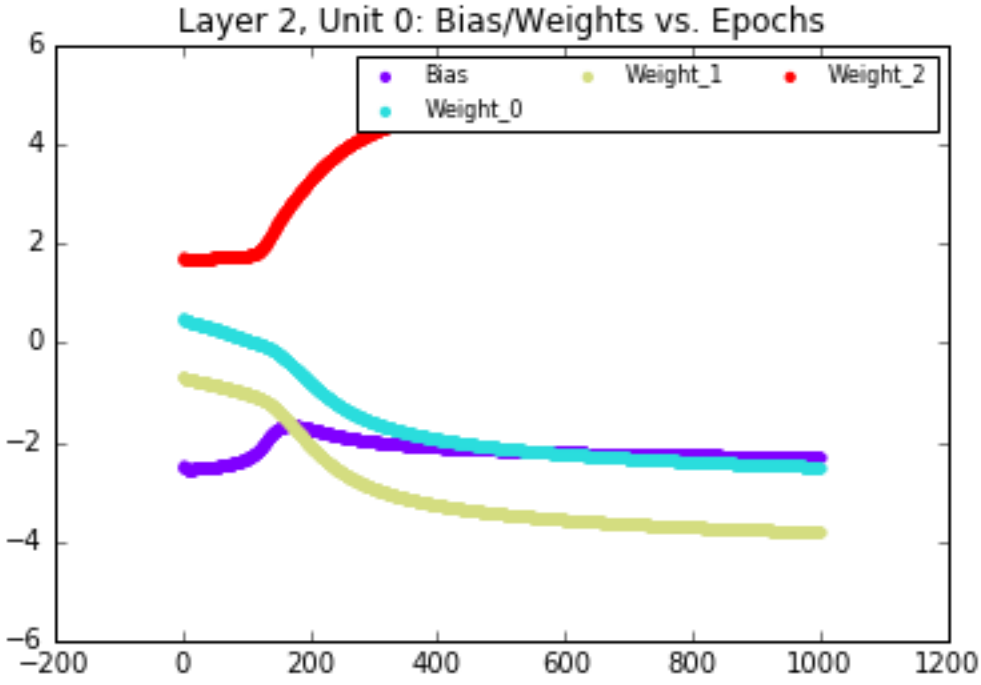

Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

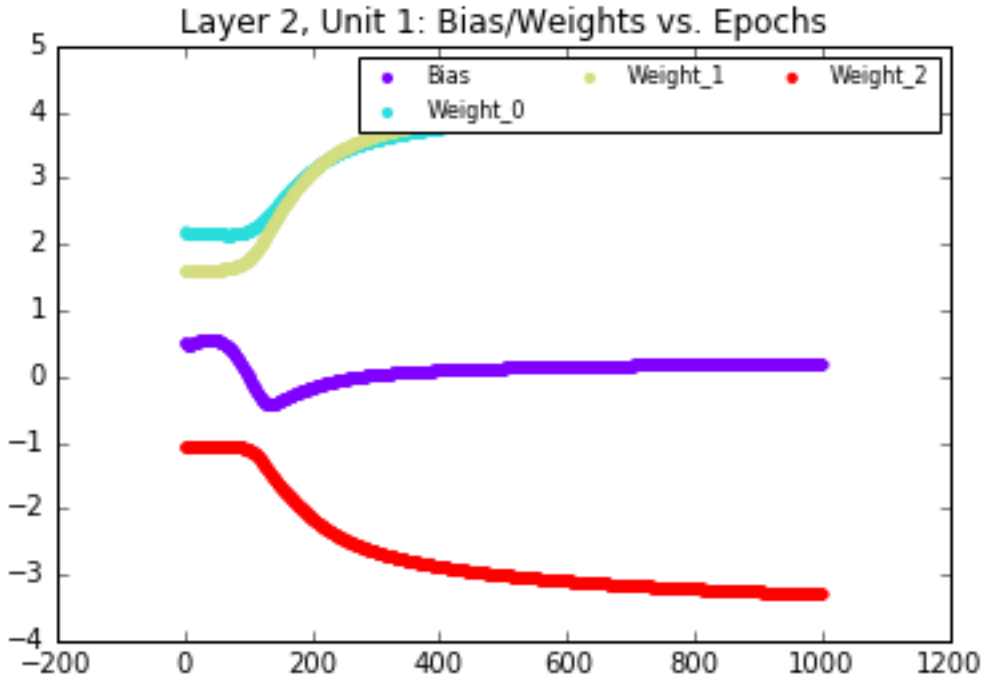
2 input units, 2 hidden layers, first of size 3, second of size 3, 4 output units

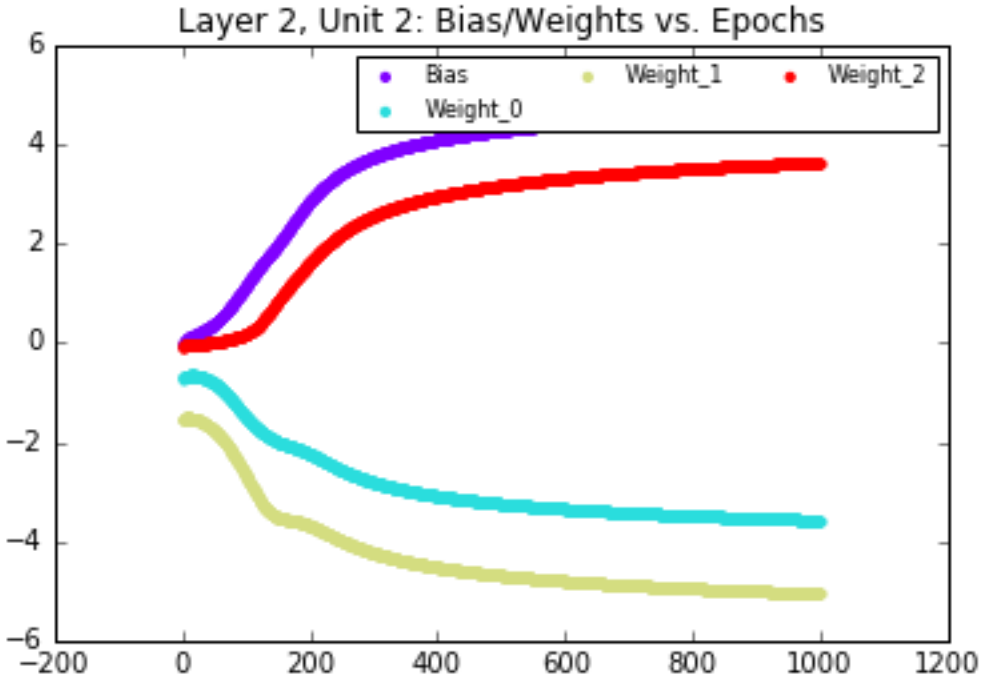


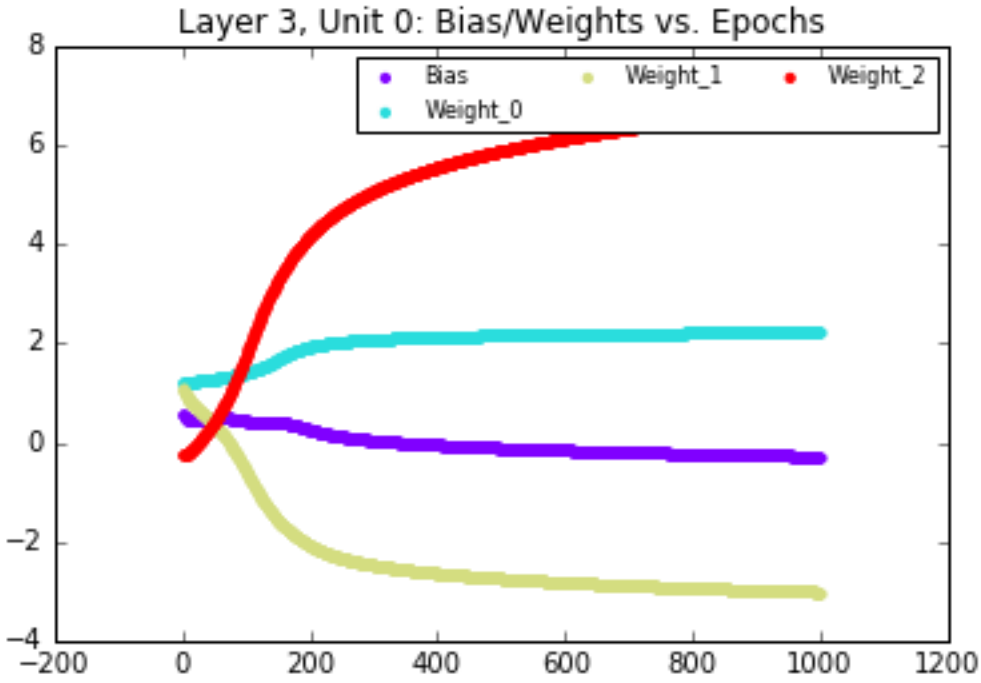


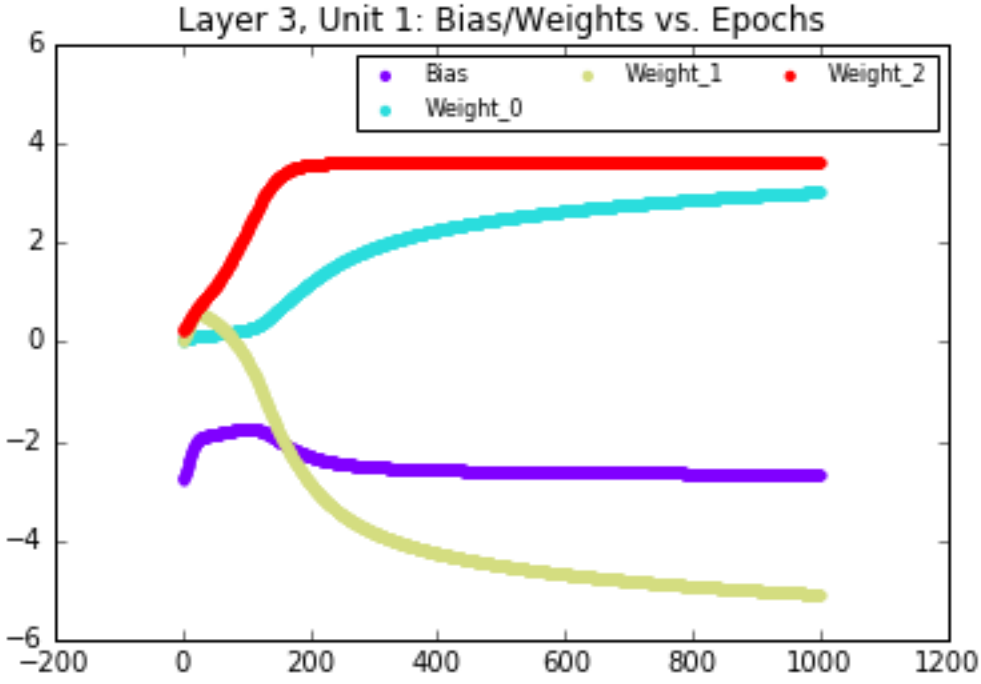

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
3
2
1
0
−1
−2
−3
−4
−5
−200
0
200
400
600
800
1000
1200

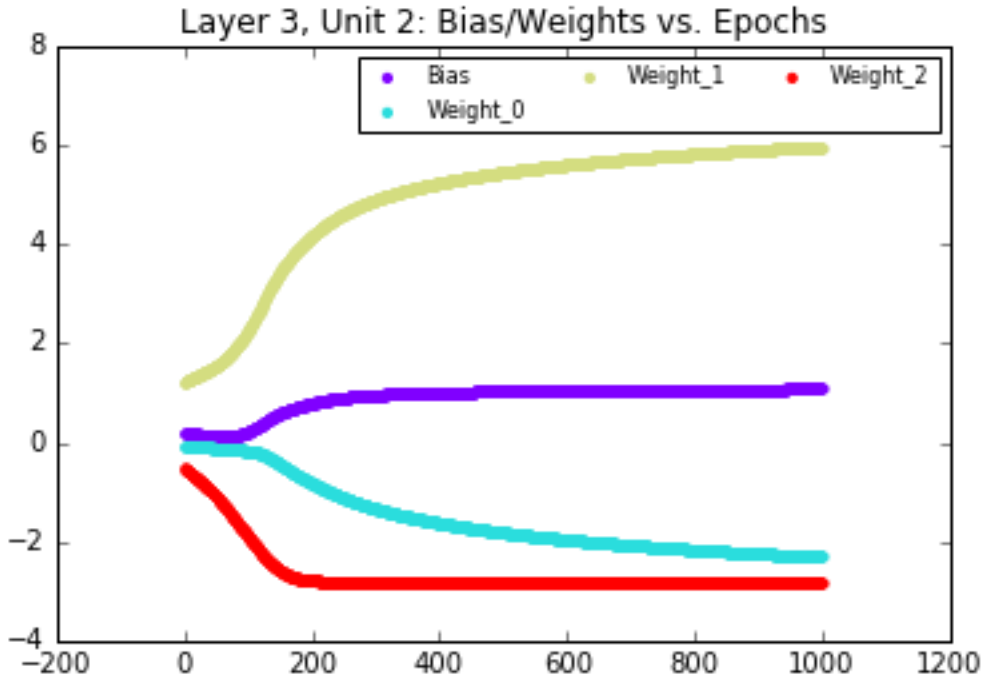

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

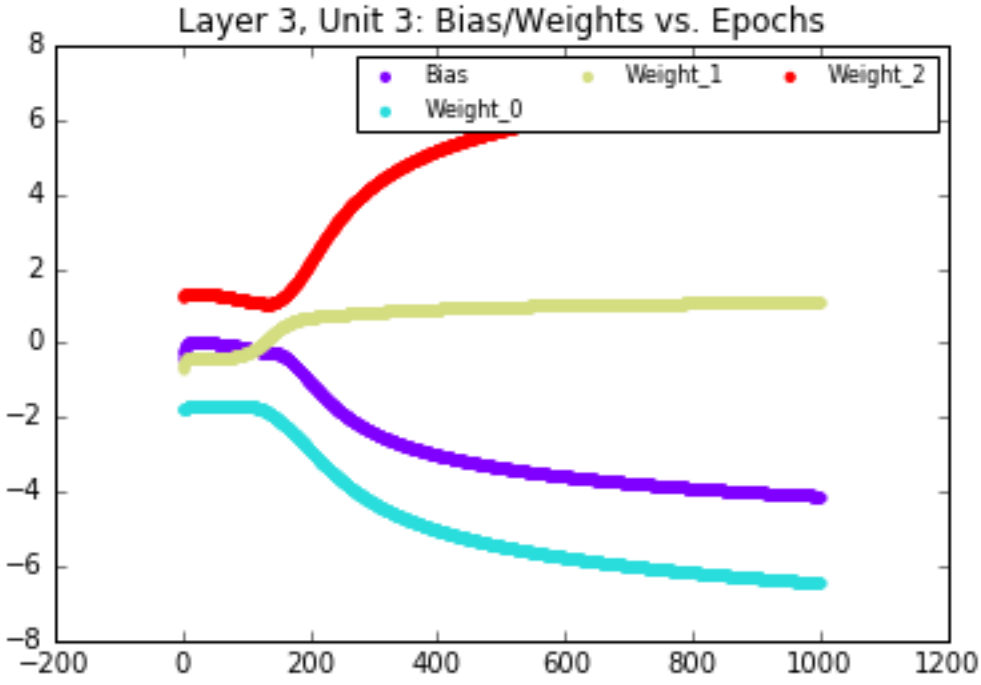

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
5
4
3
2
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

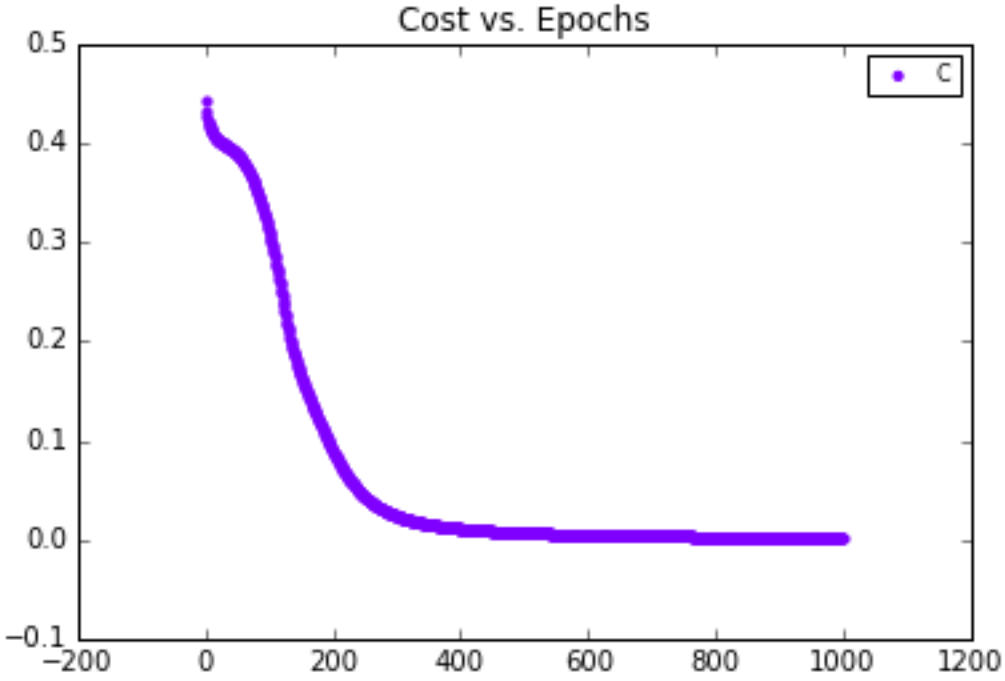

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

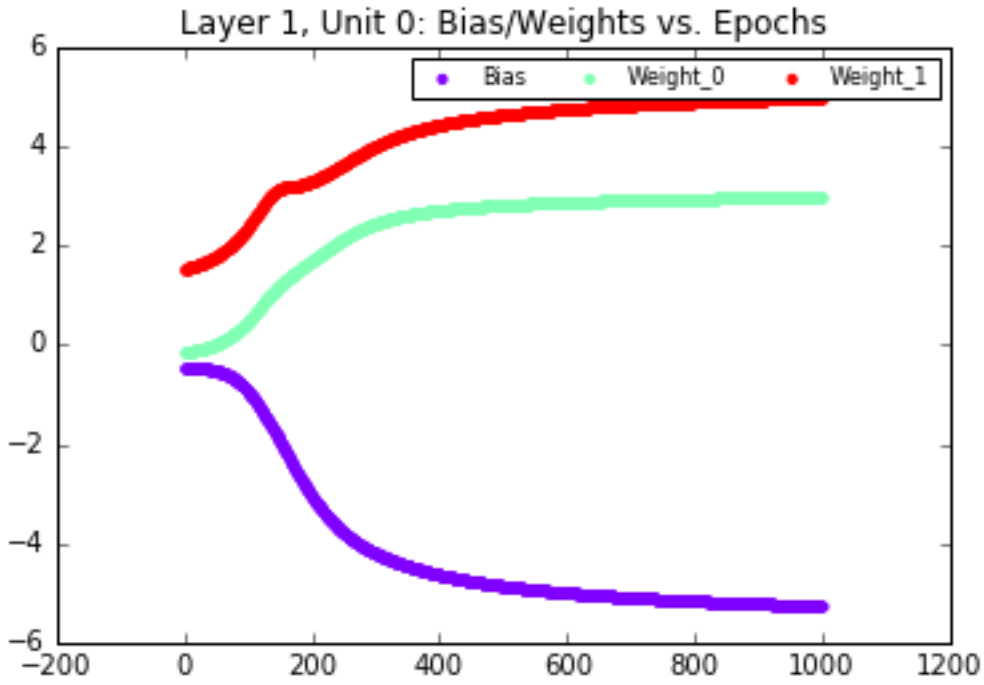

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

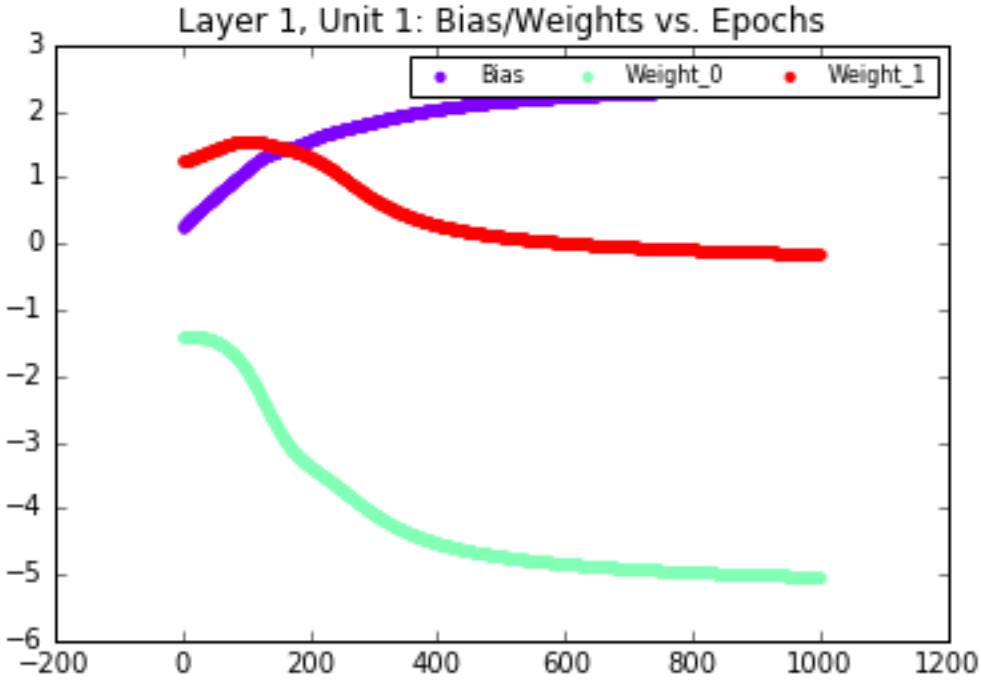

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

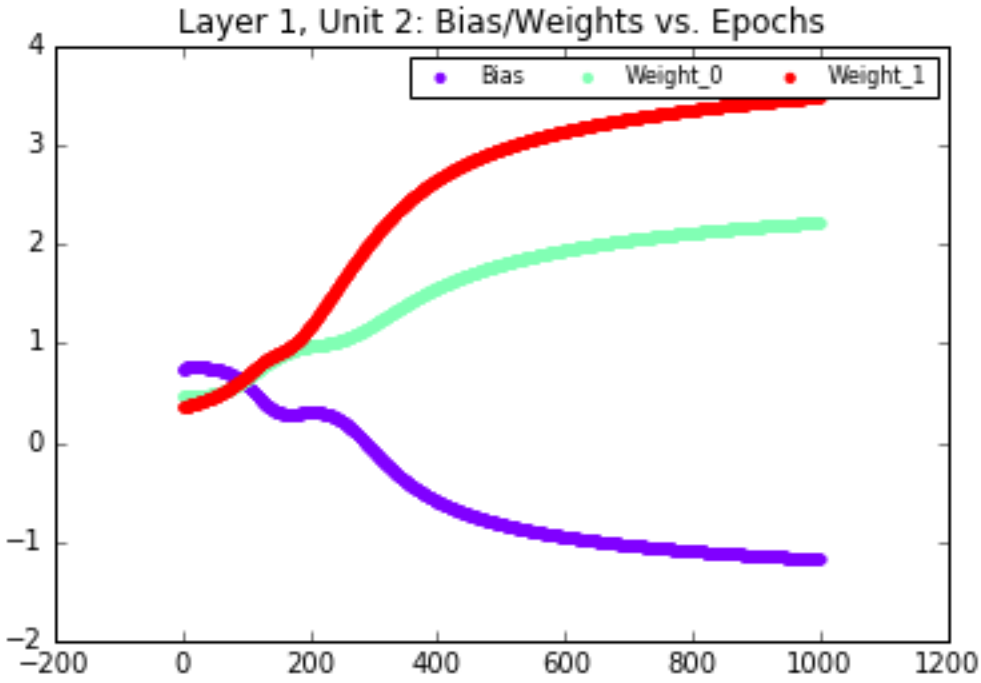

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

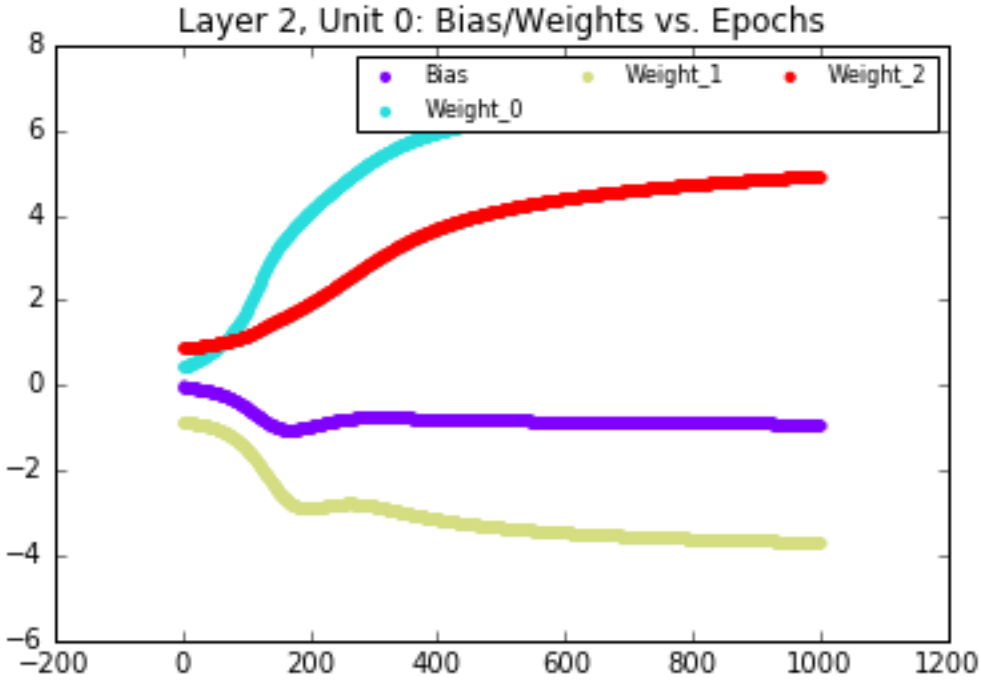

Cost vs. Epochs
C
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

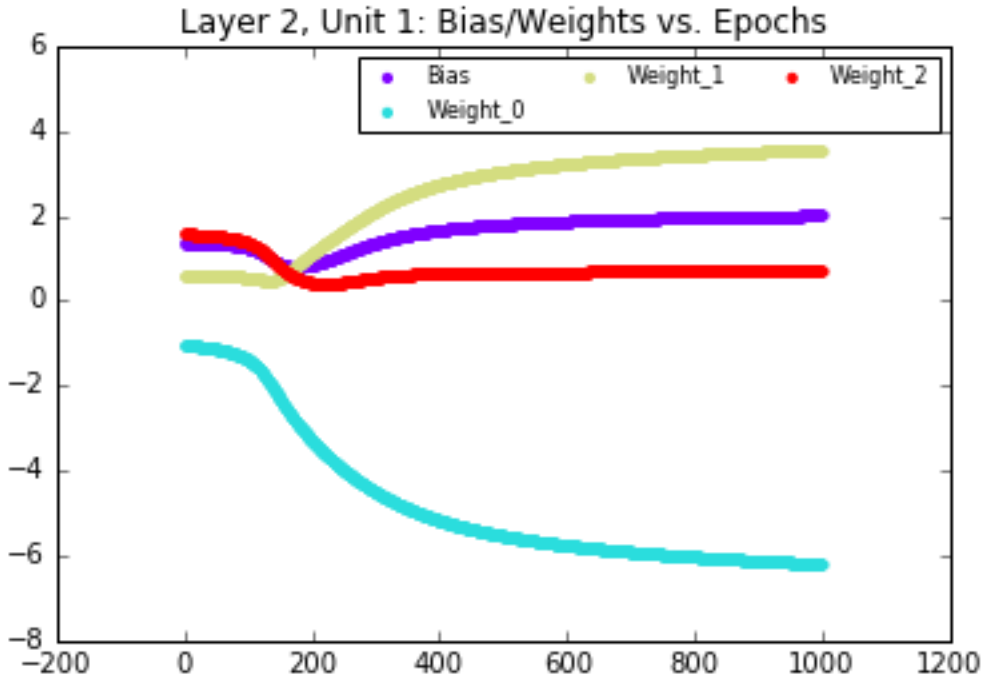
2 input units, 2 hidden layers, first of size 3, second of size 4, 4 output units

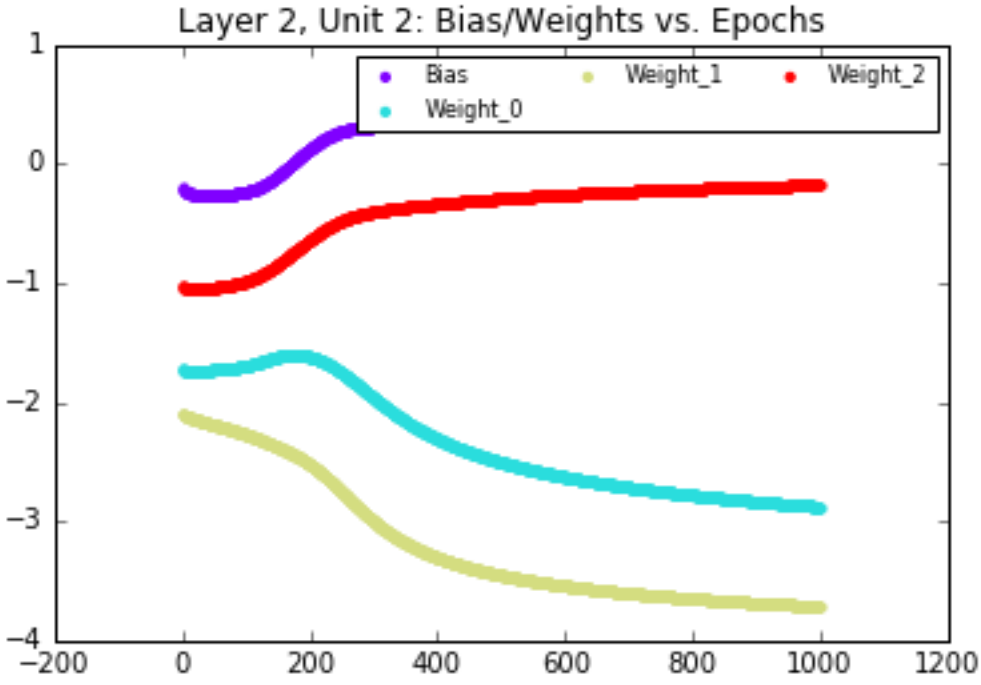


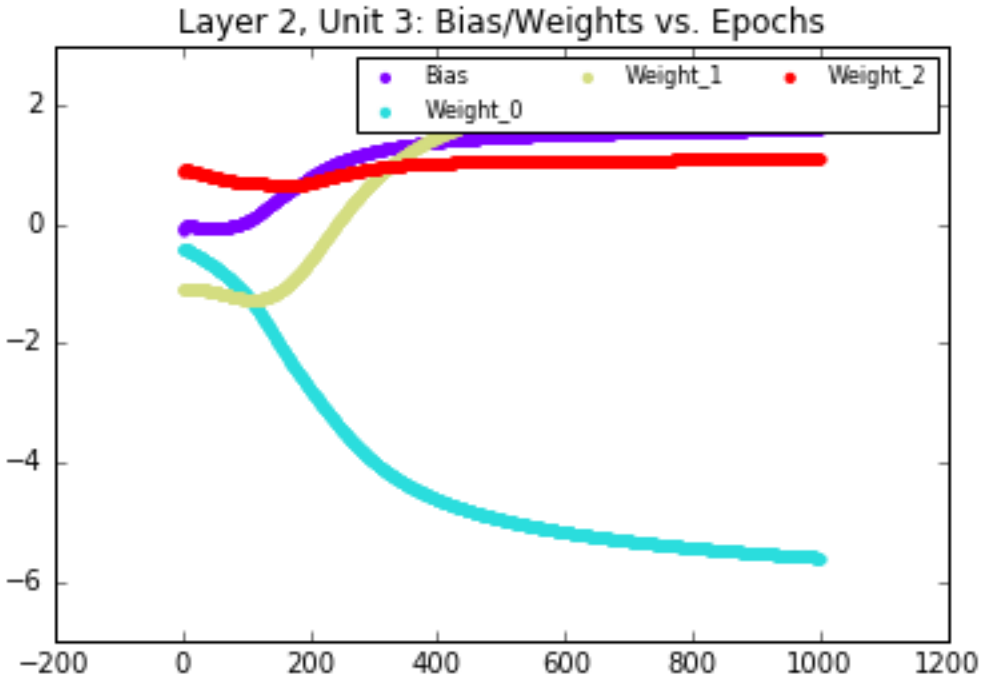


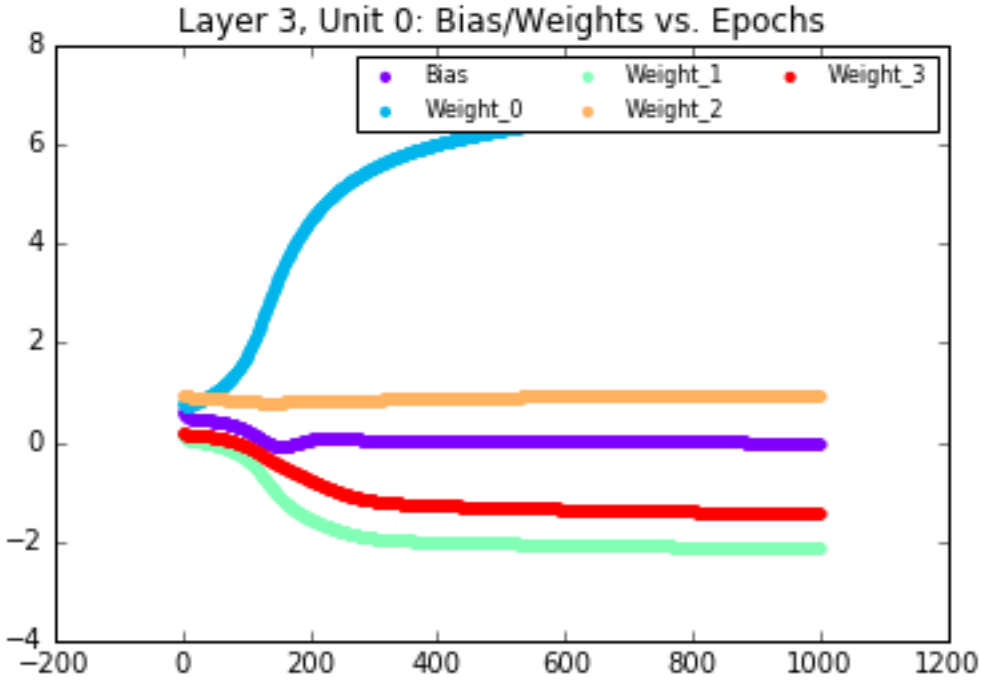

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
3
2
1
0
−1
−2
−200
0
200
400
600
800
1000
1200

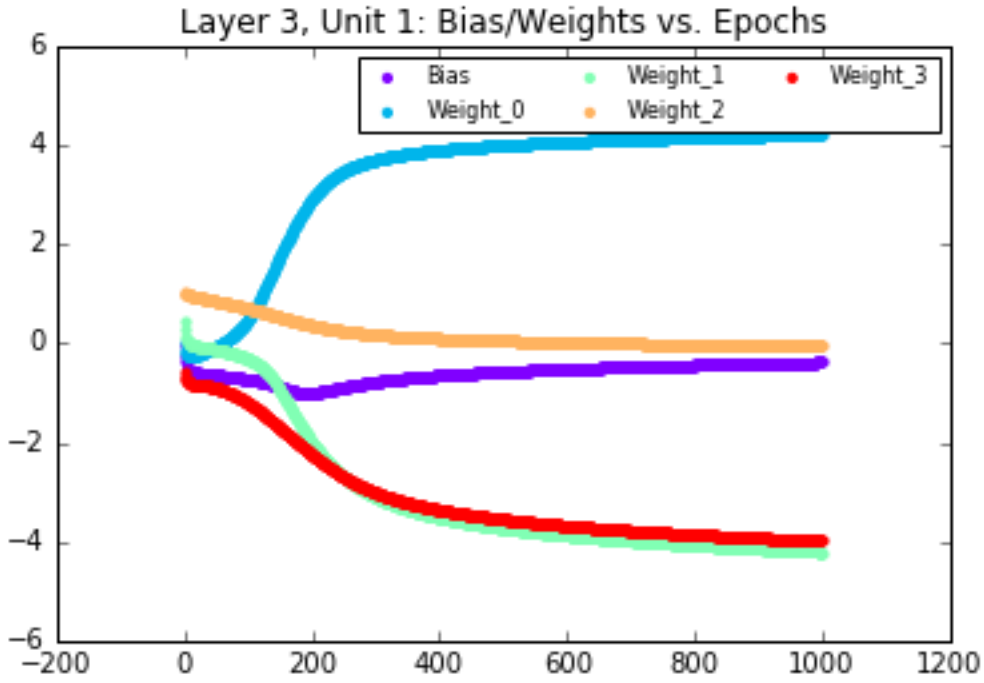

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

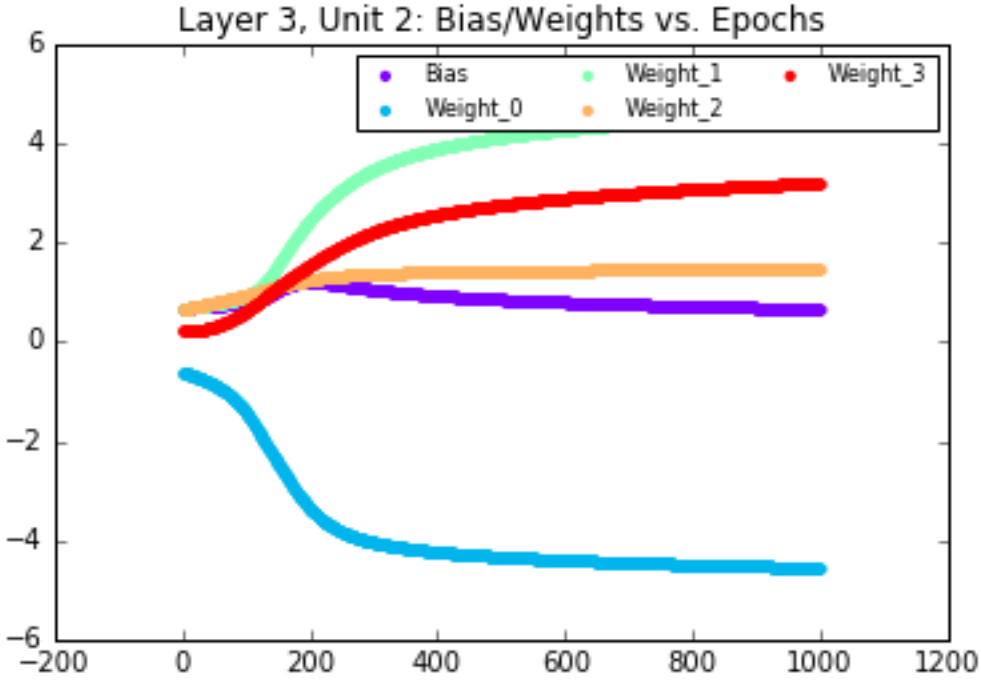

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

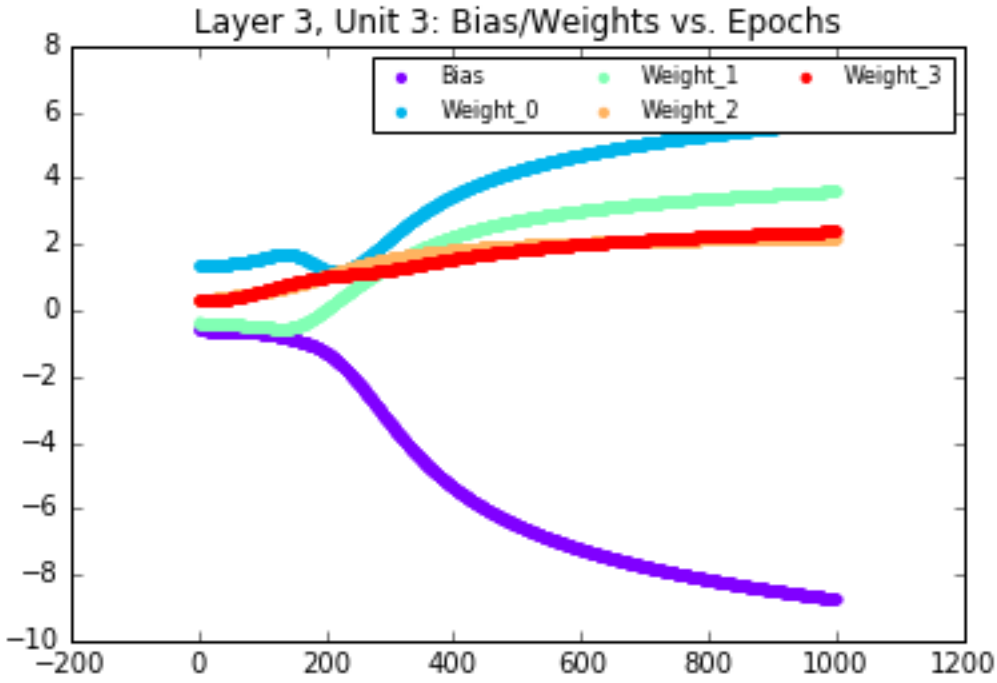

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

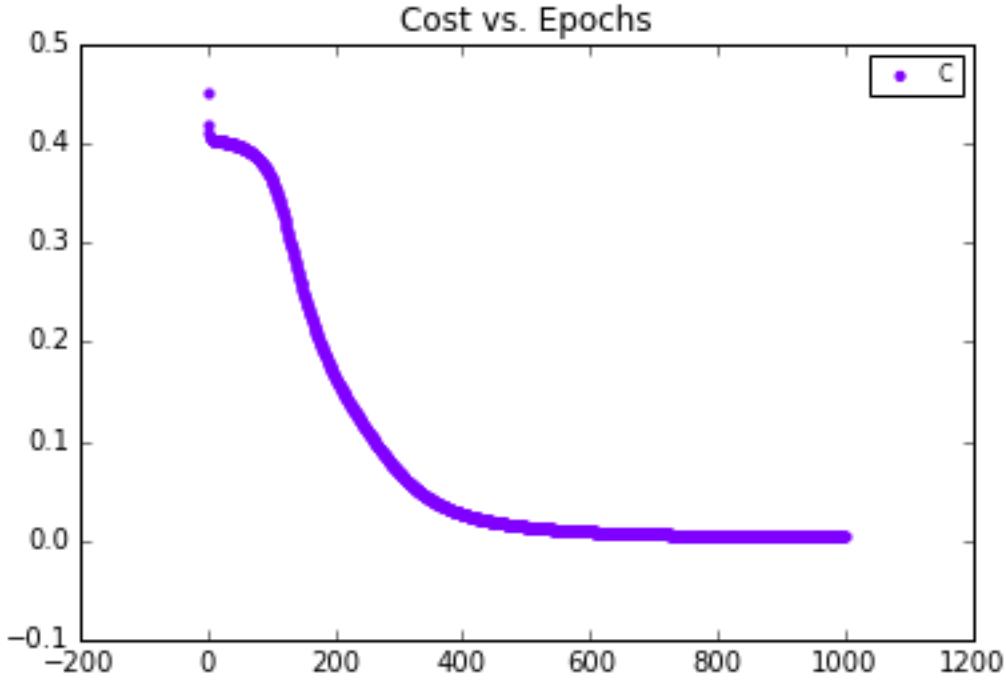

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

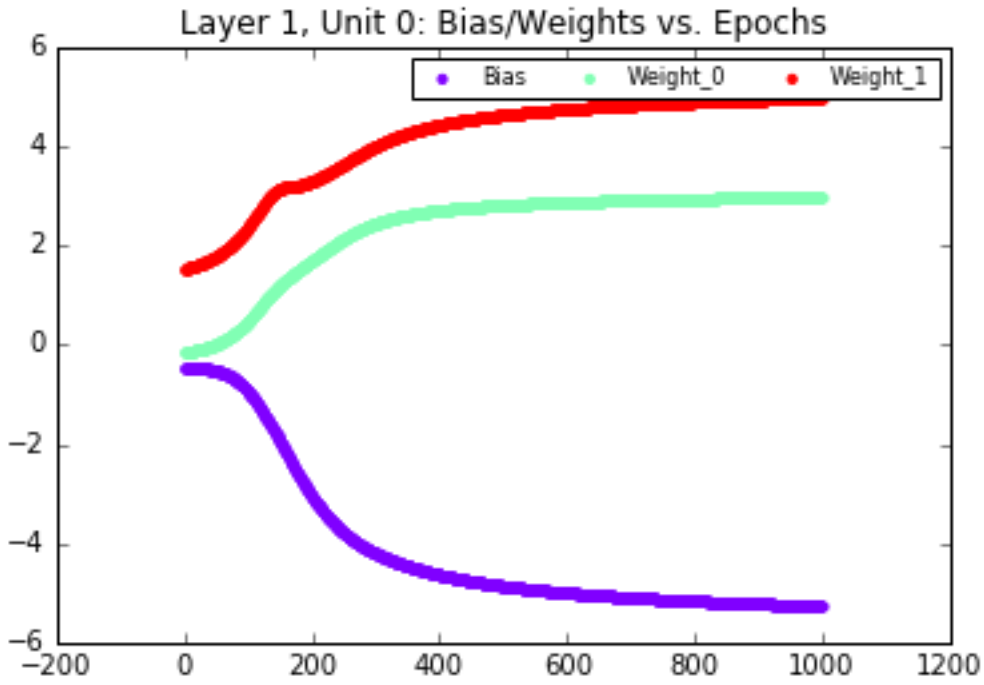

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

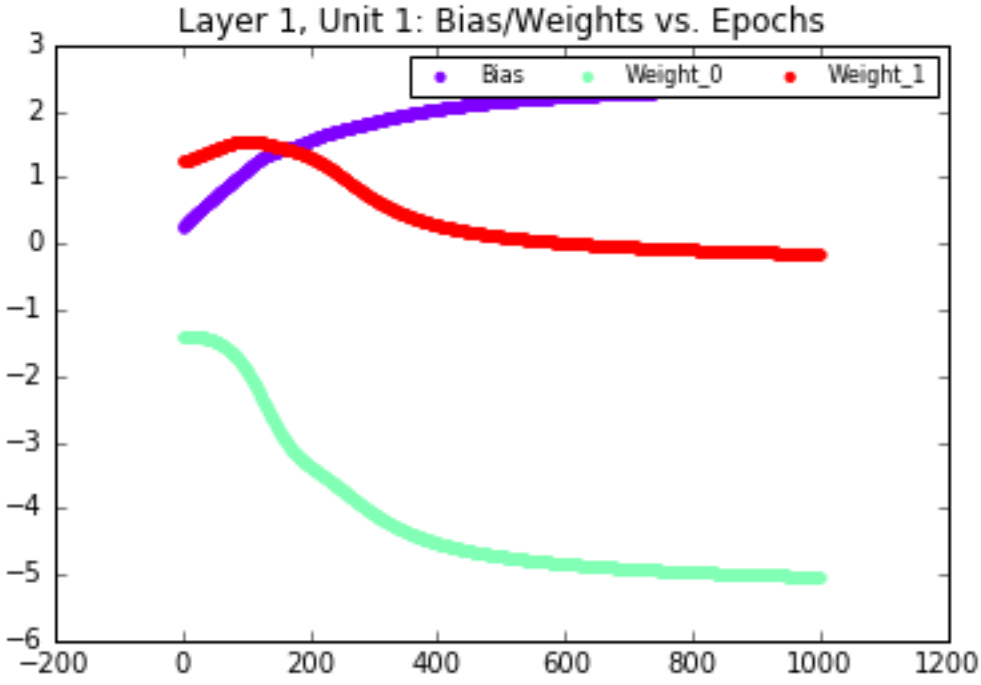

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

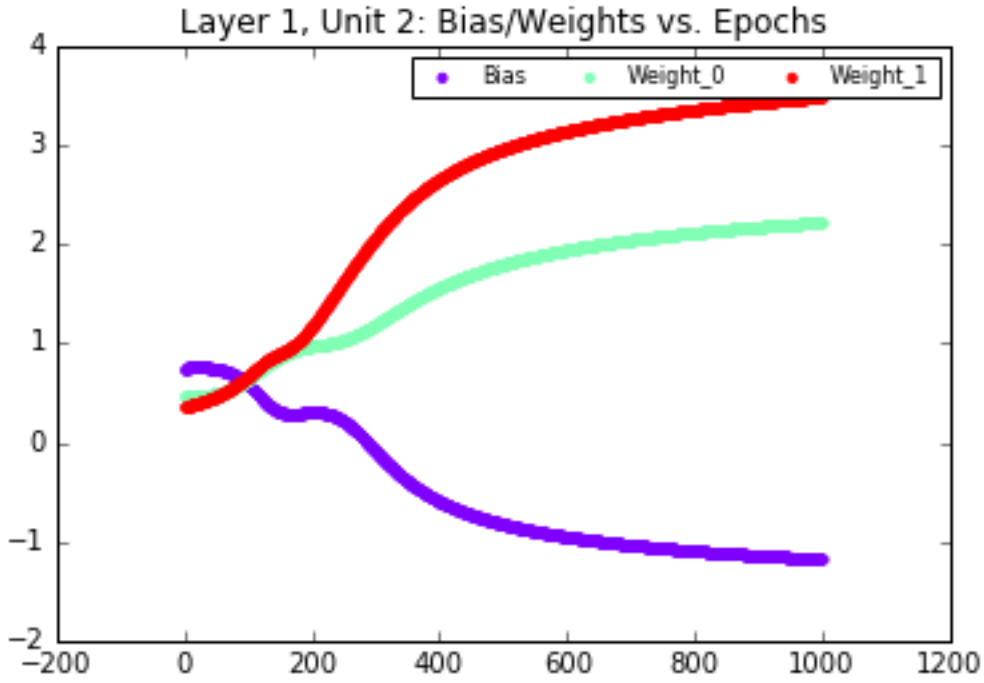

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

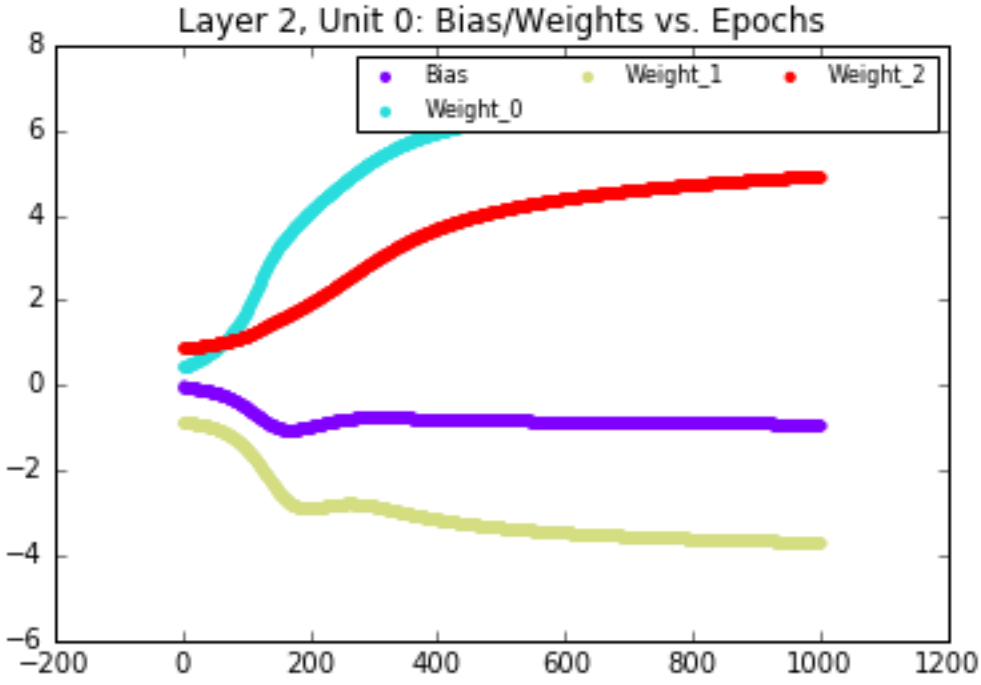

Cost vs. Epochs
C
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

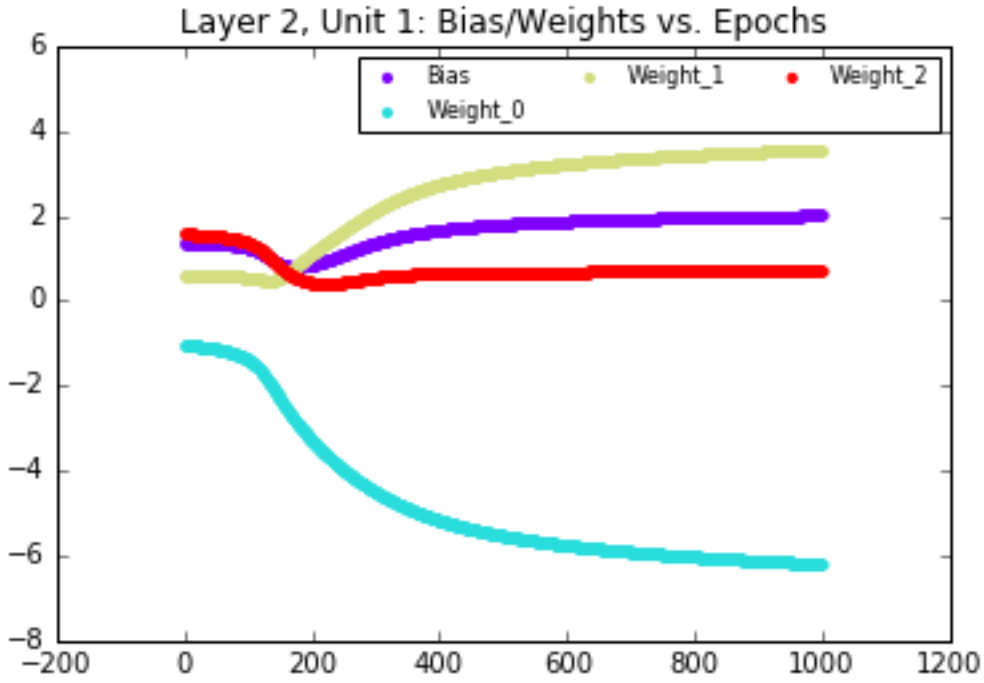
2 input units, 2 hidden layers, first of size 3, second of size 4, 4 output units

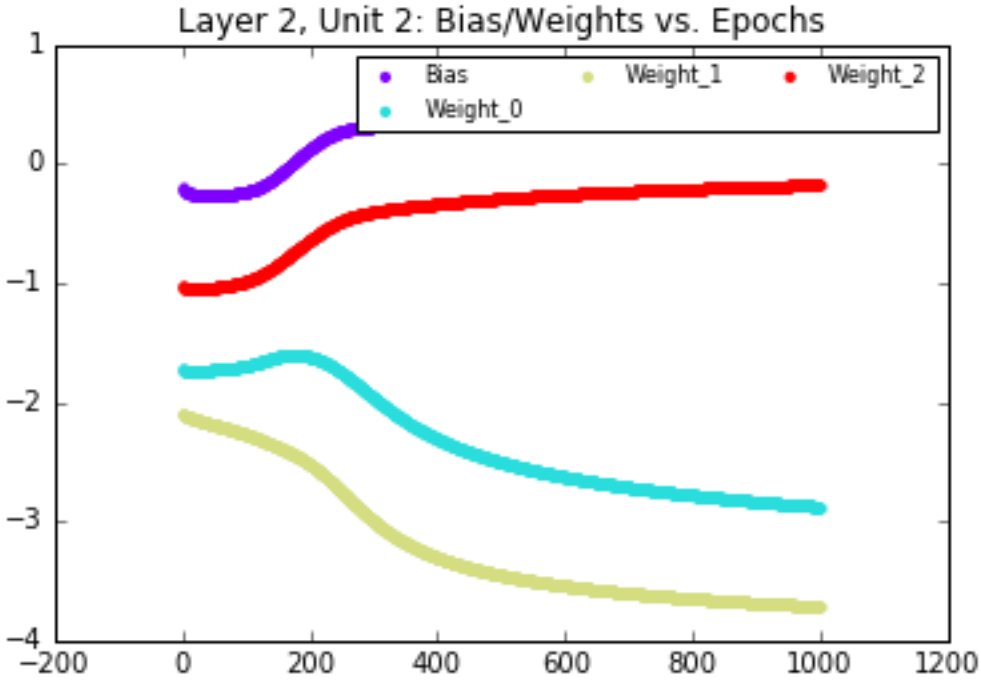


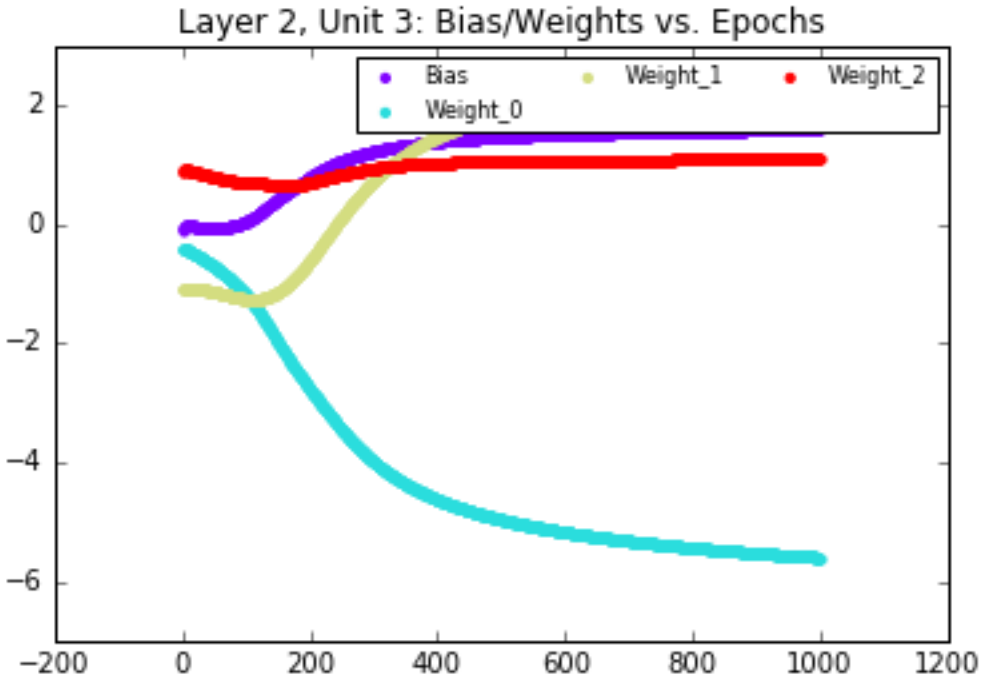


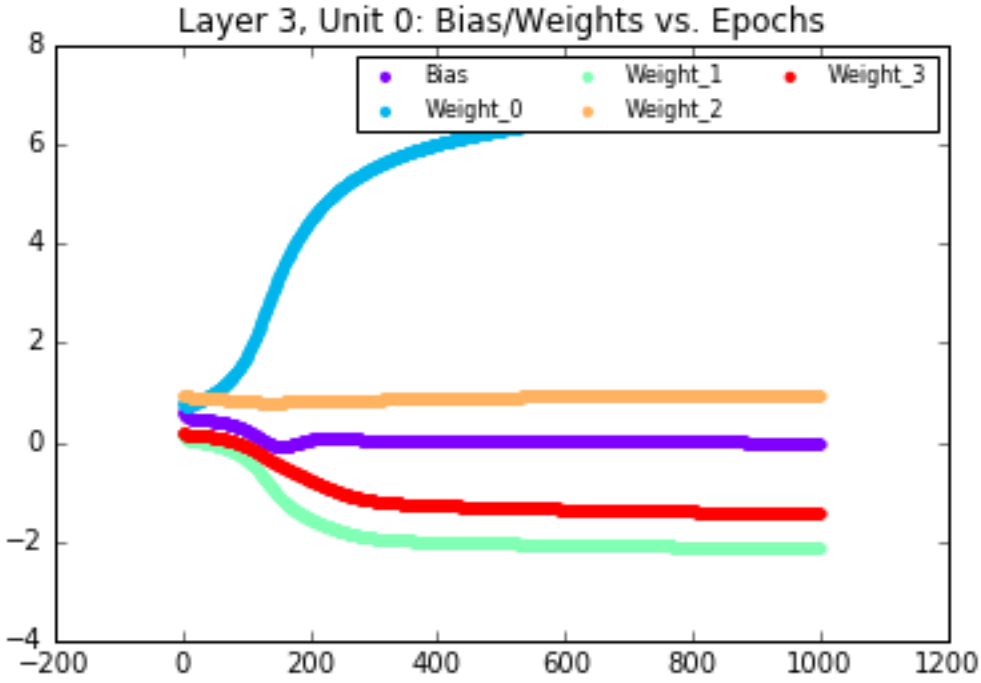

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
3
2
1
0
−1
−2
−200
0
200
400
600
800
1000
1200

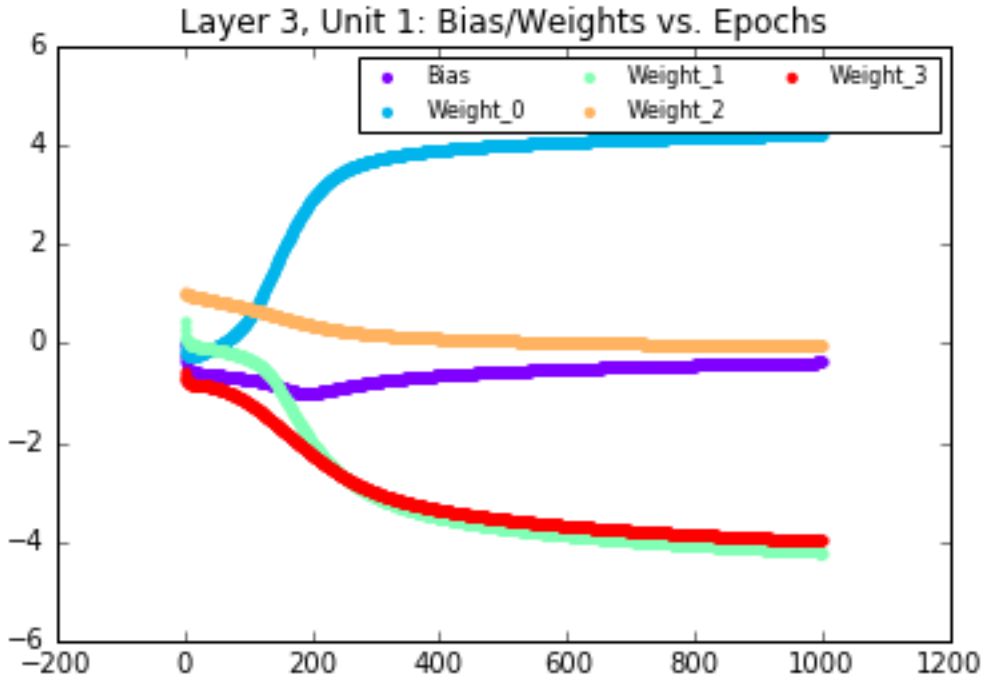

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

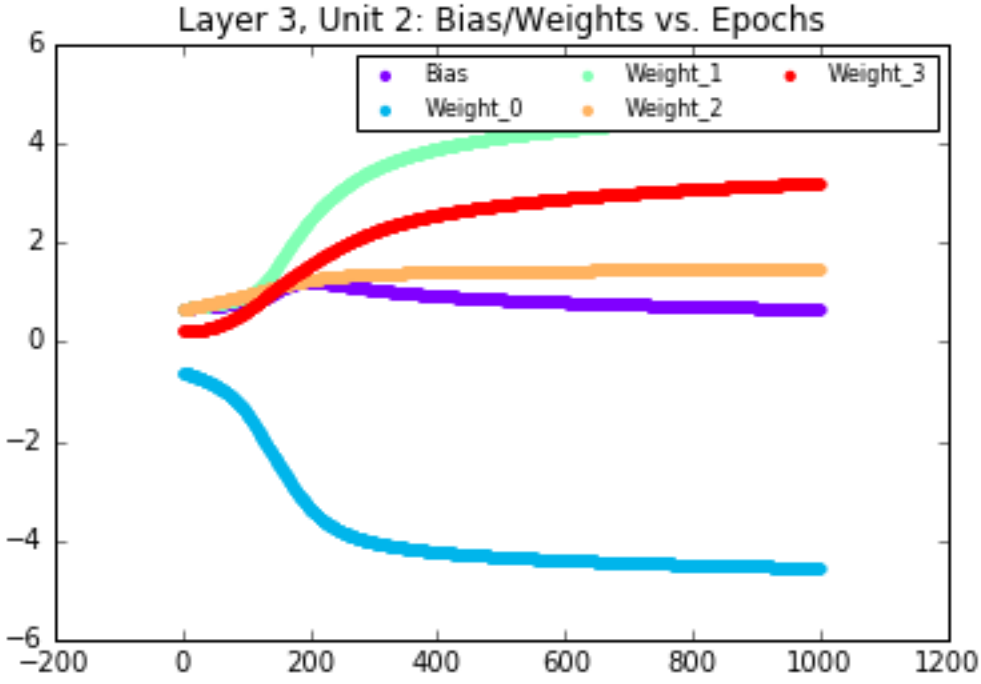

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

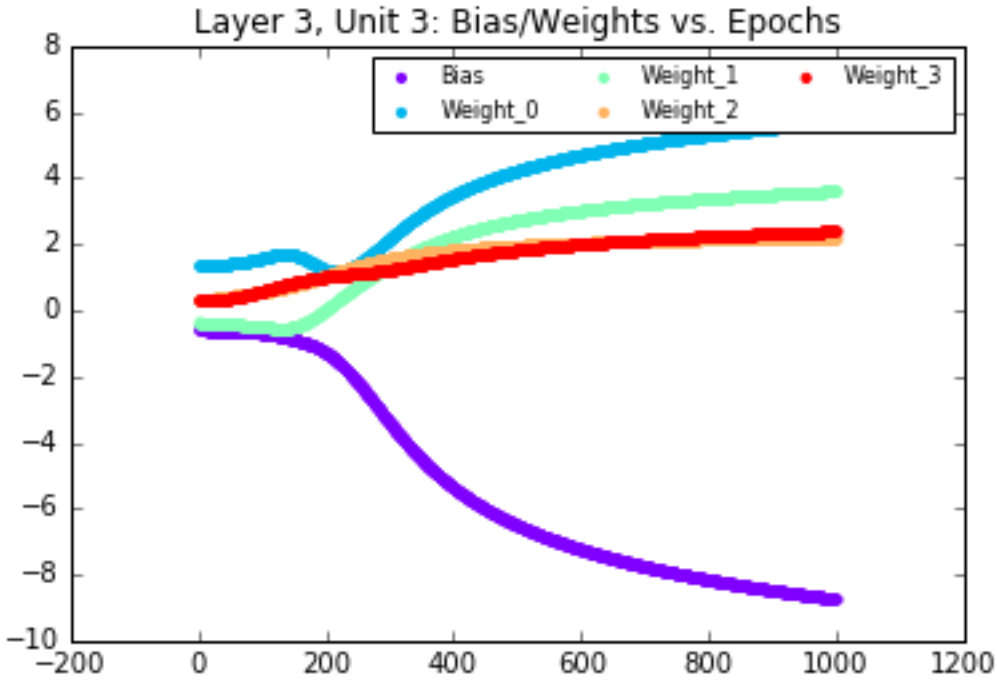

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

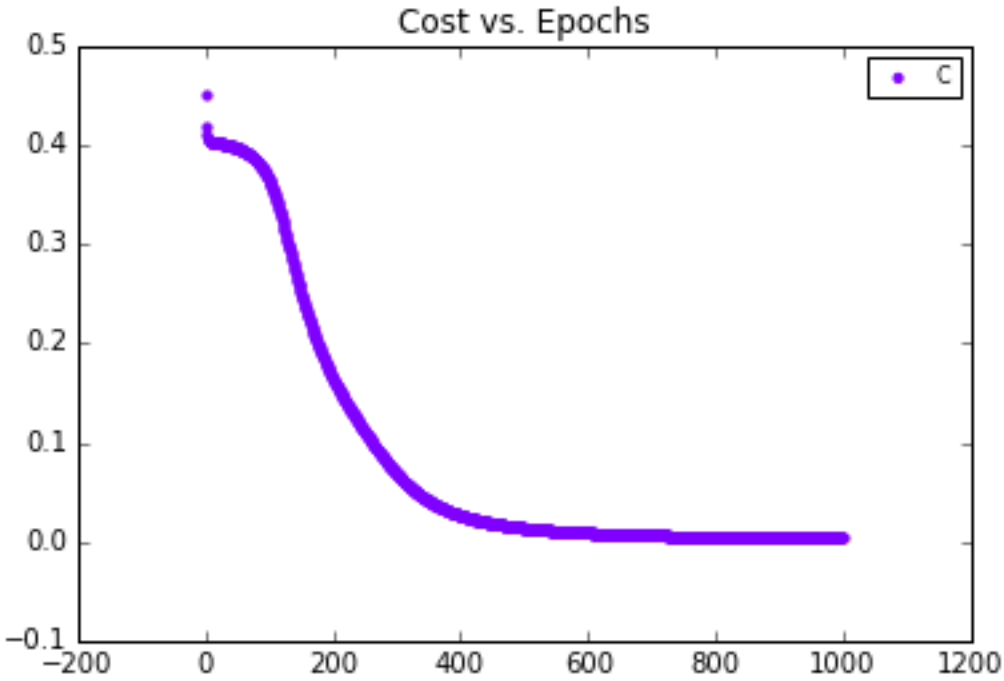

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

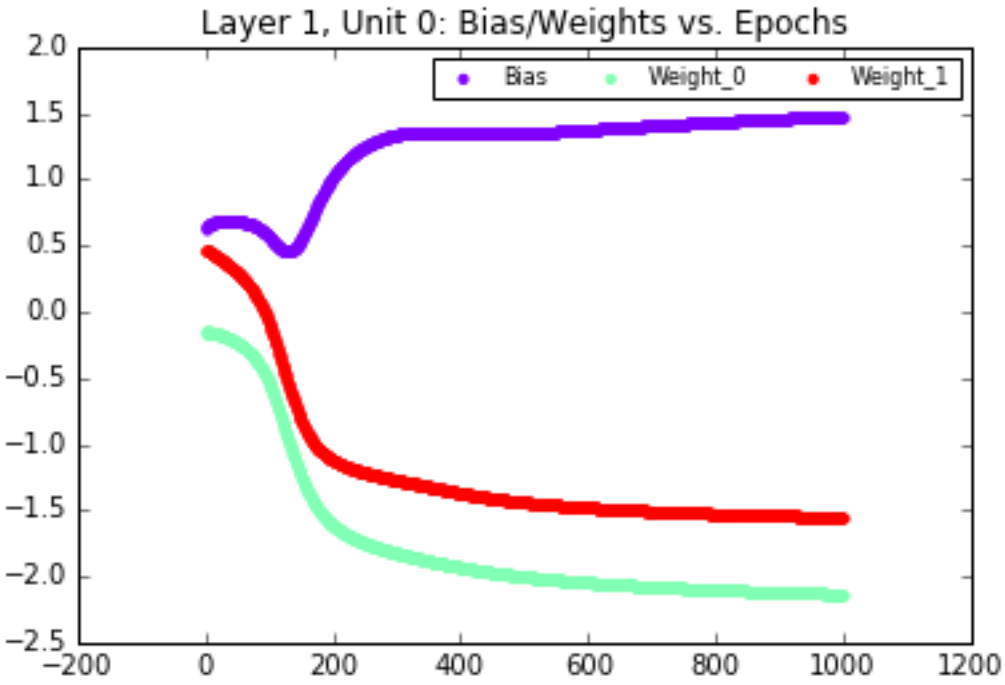

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

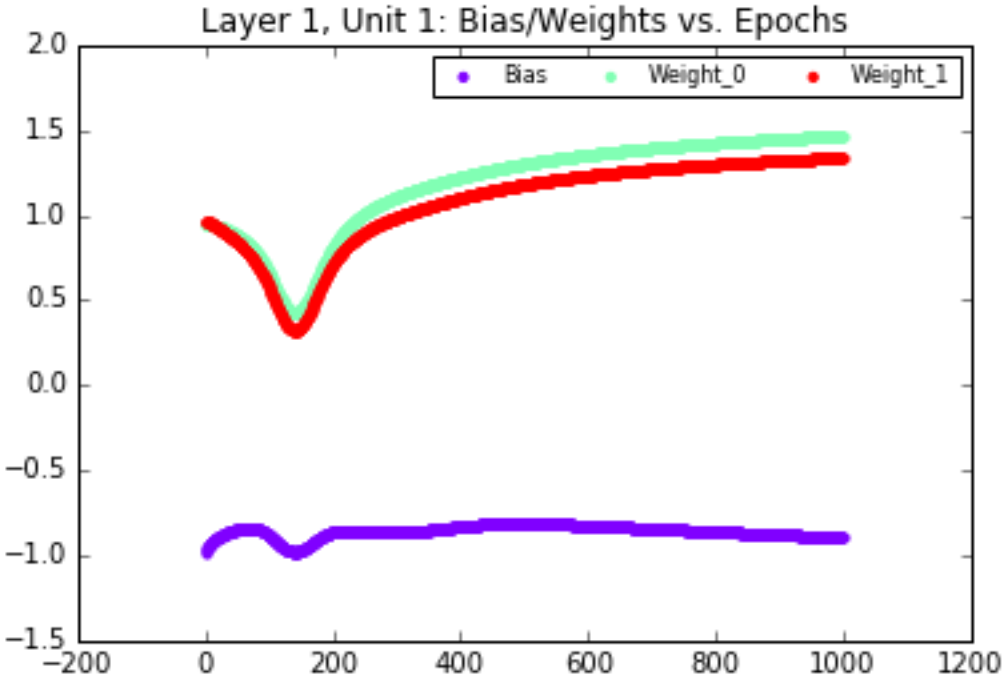

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

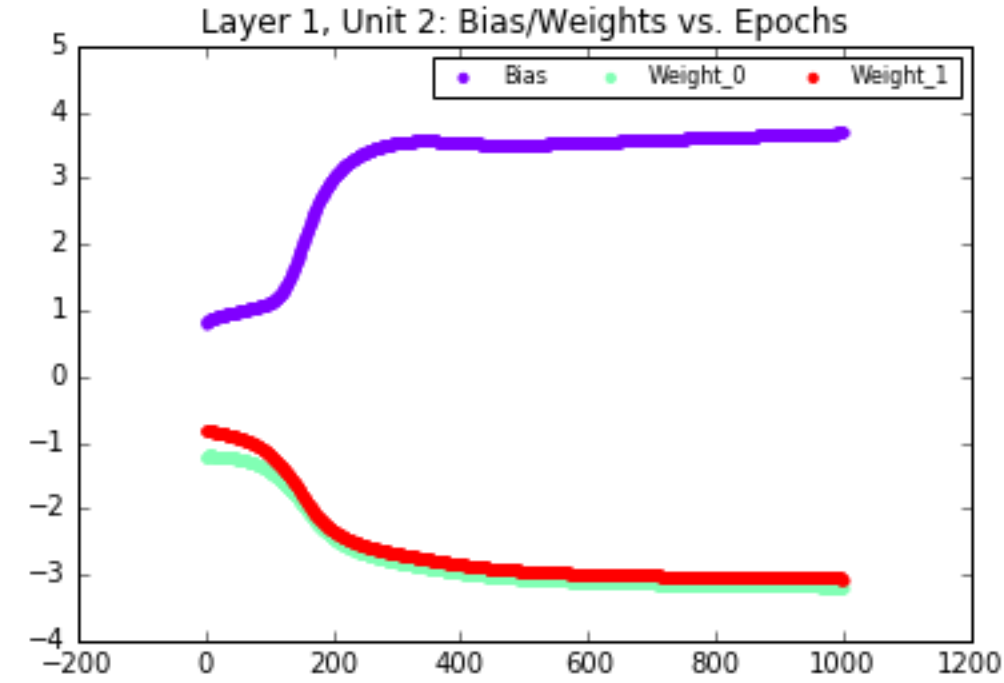

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

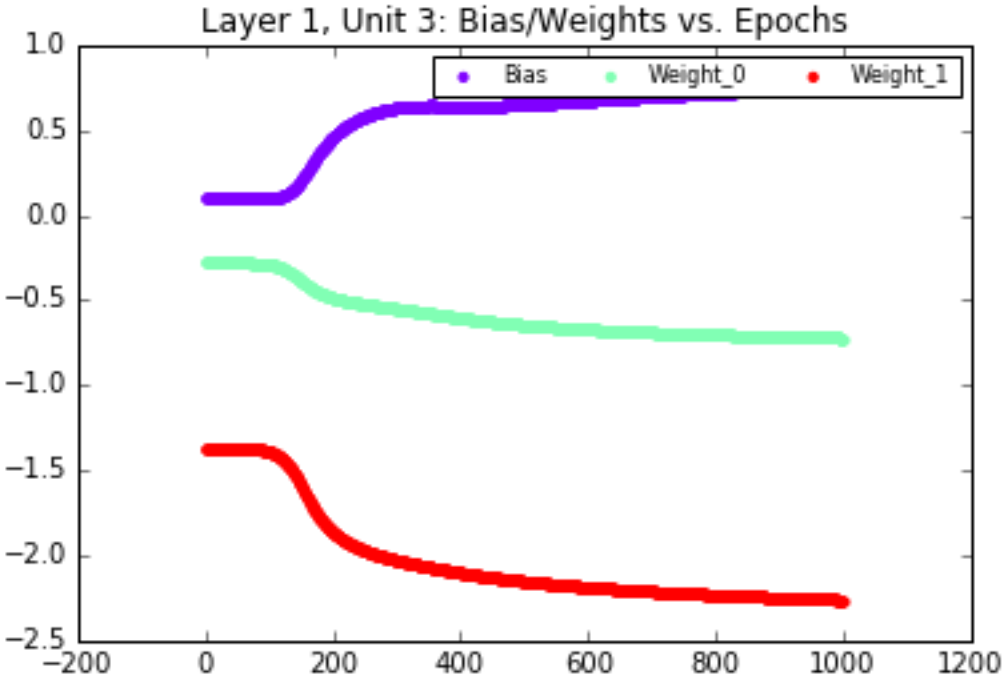

Cost vs. Epochs
C
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

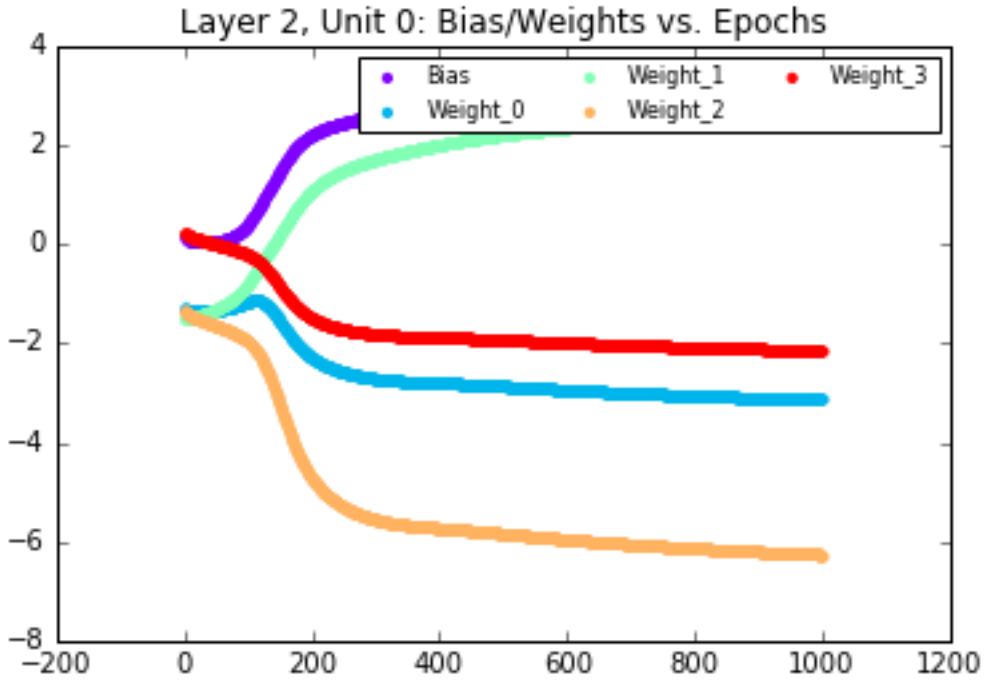
2 inputs, 2 hidden layers, first of size 4, second of size 1, 4 output units

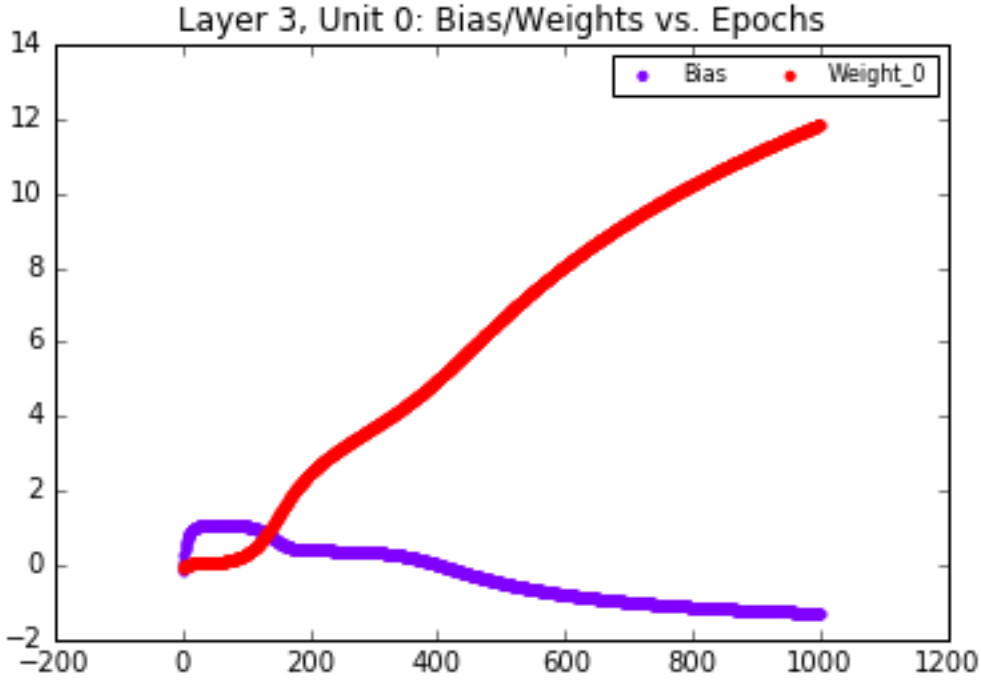


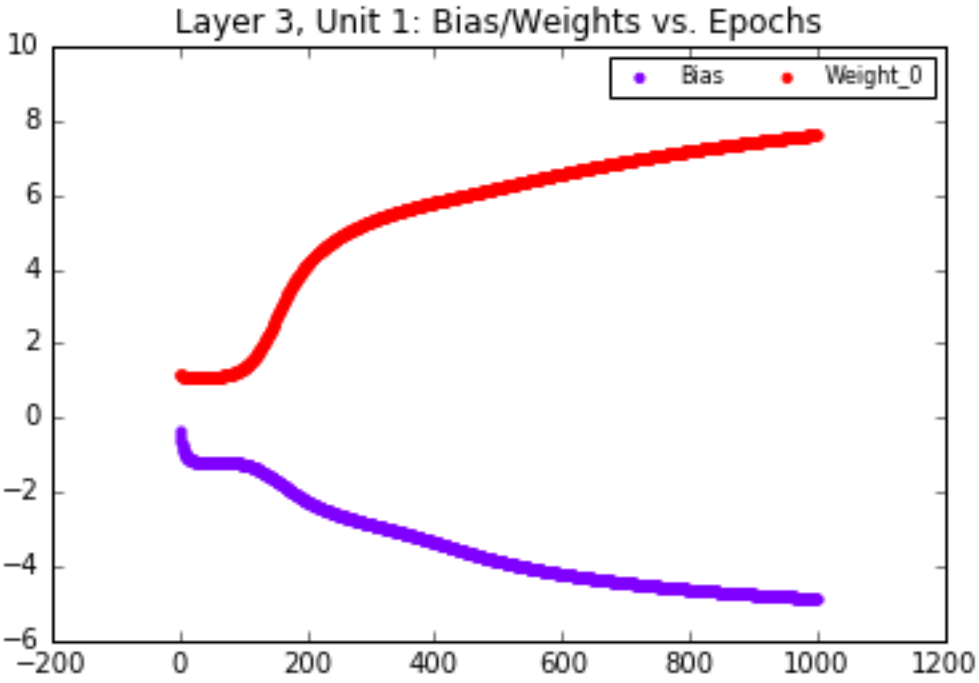


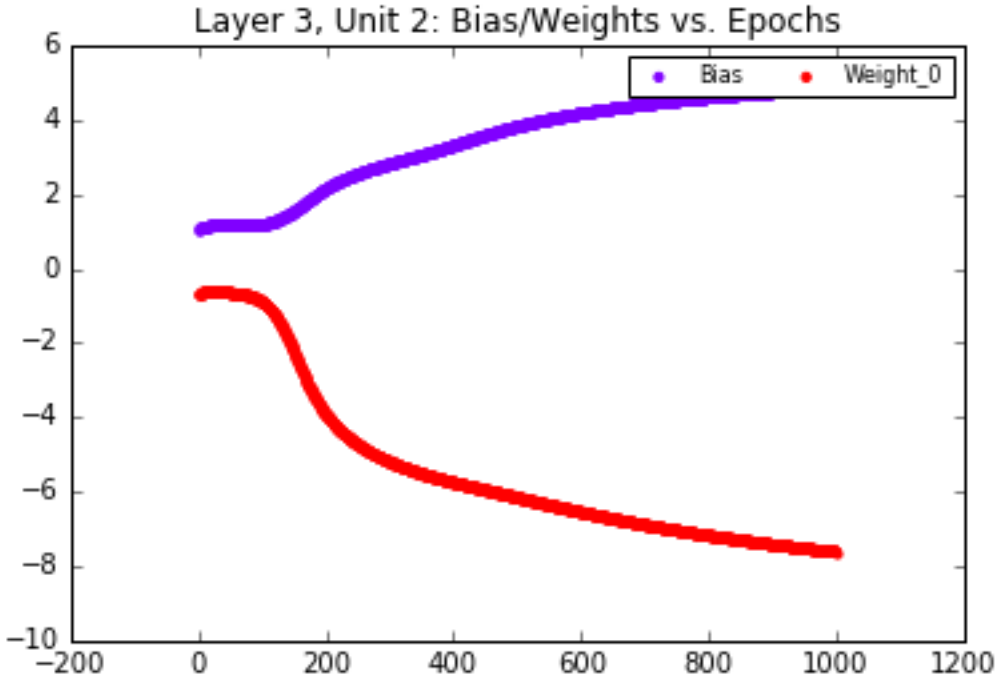

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
5
4
3
2
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

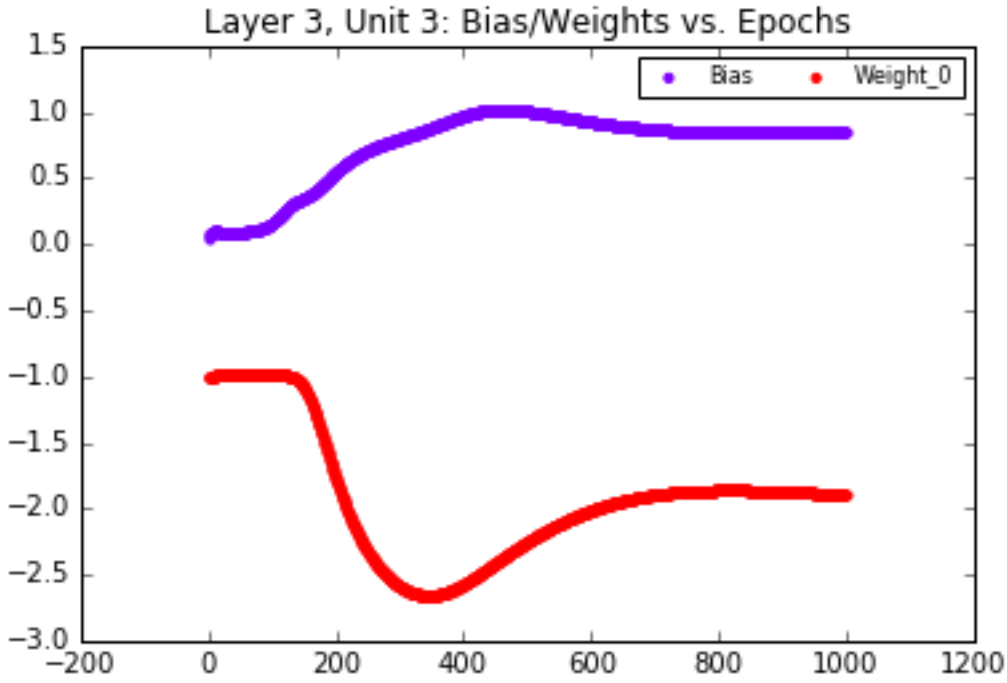

Layer 1, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−200
0
200
400
600
800
1000
1200

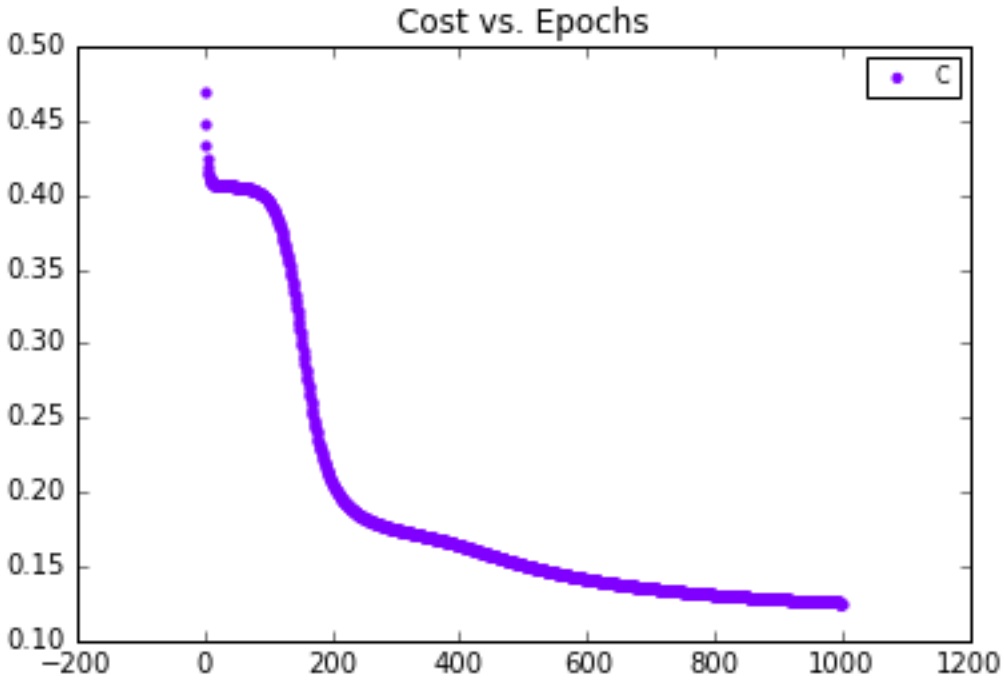

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
14
12
10
8
6
4
2
0
−2
−200
0
200
400
600
800
1000
1200

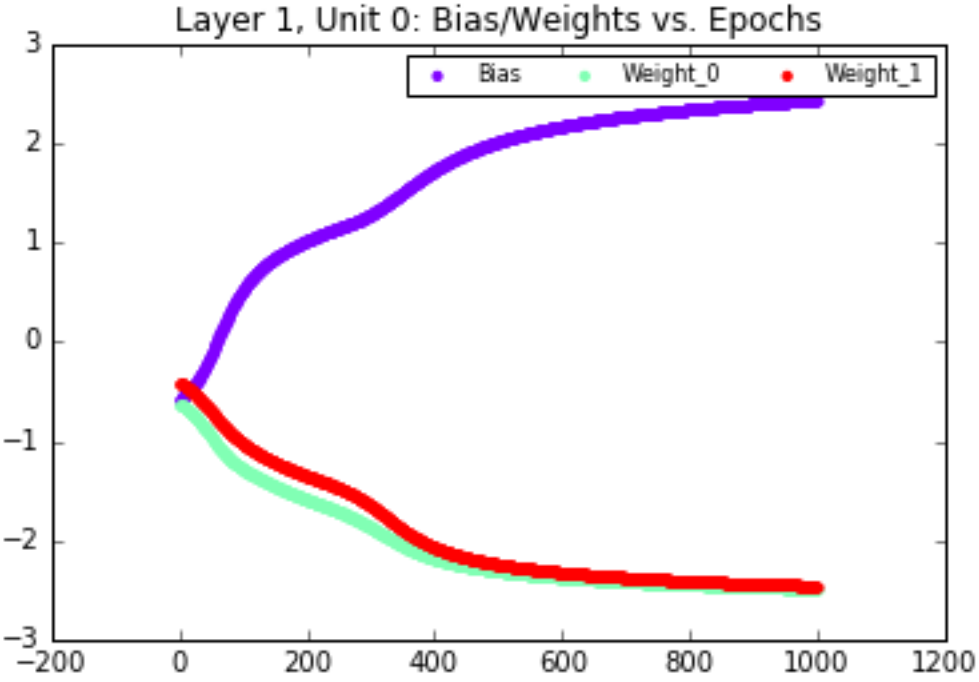

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
10
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

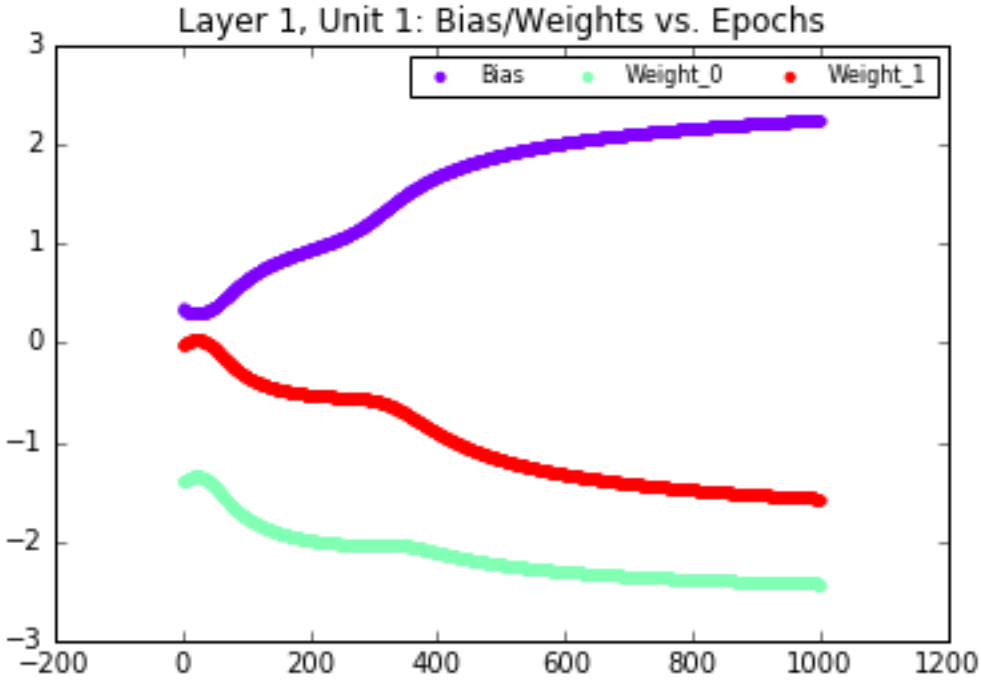

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

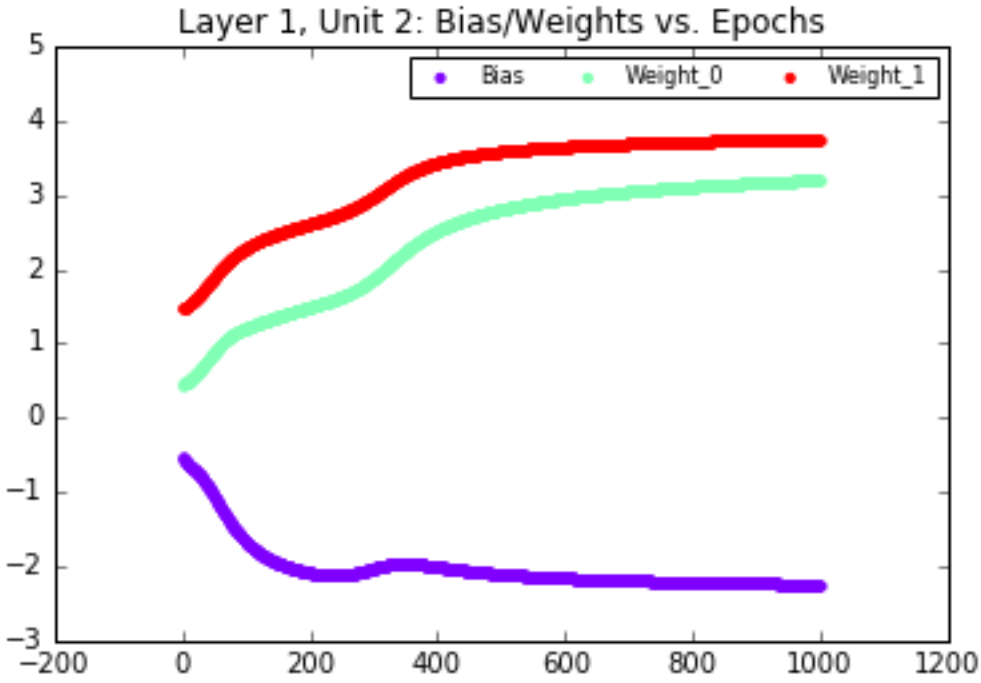

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−200
0
200
400
600
800
1000
1200

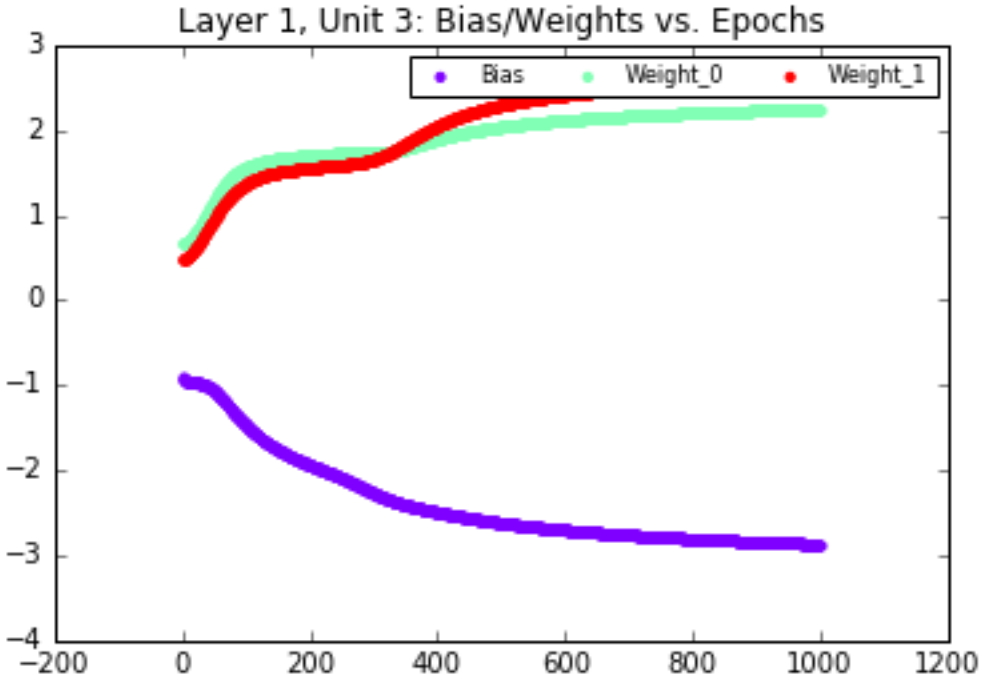

Cost vs. Epochs
C
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
−200
0
200
400
600
800
1000
1200

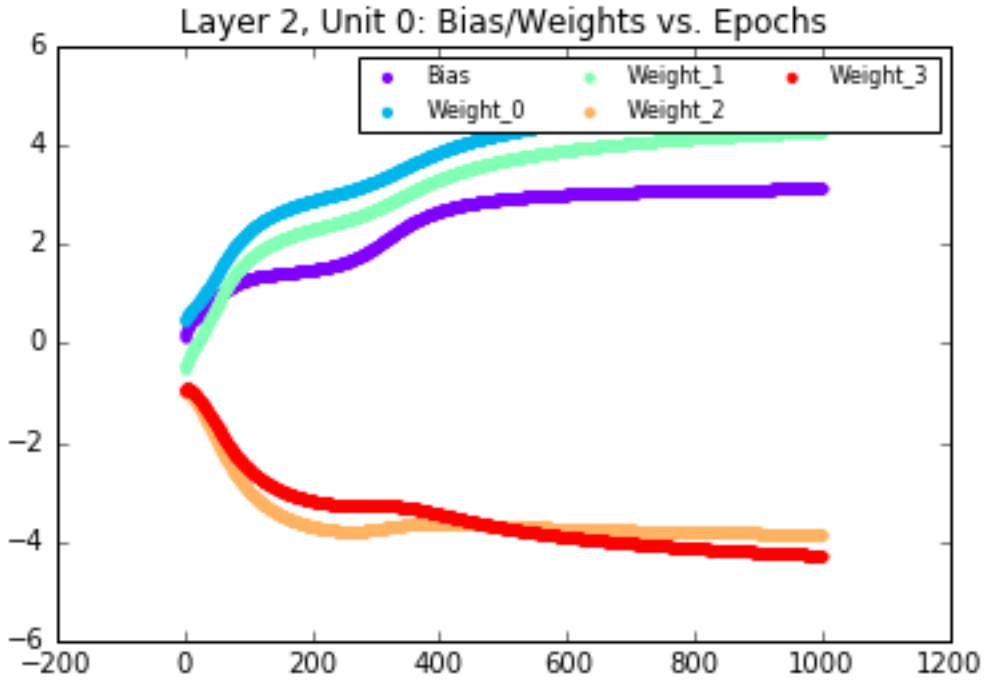
2 inputs, 2 hidden layers, first of size 4, second of size 2, 4 output units

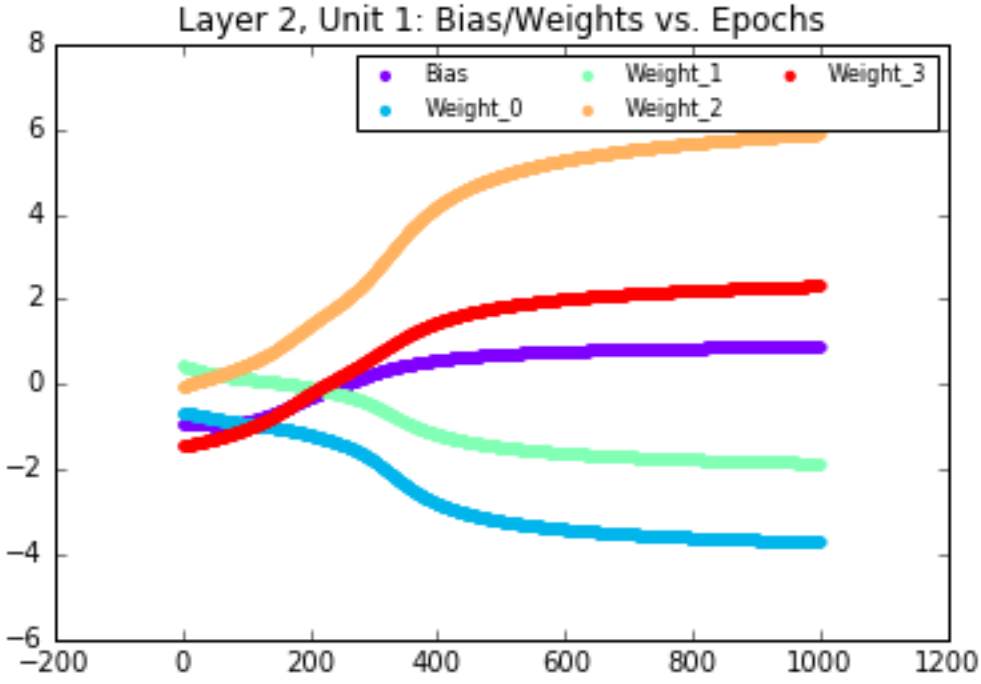


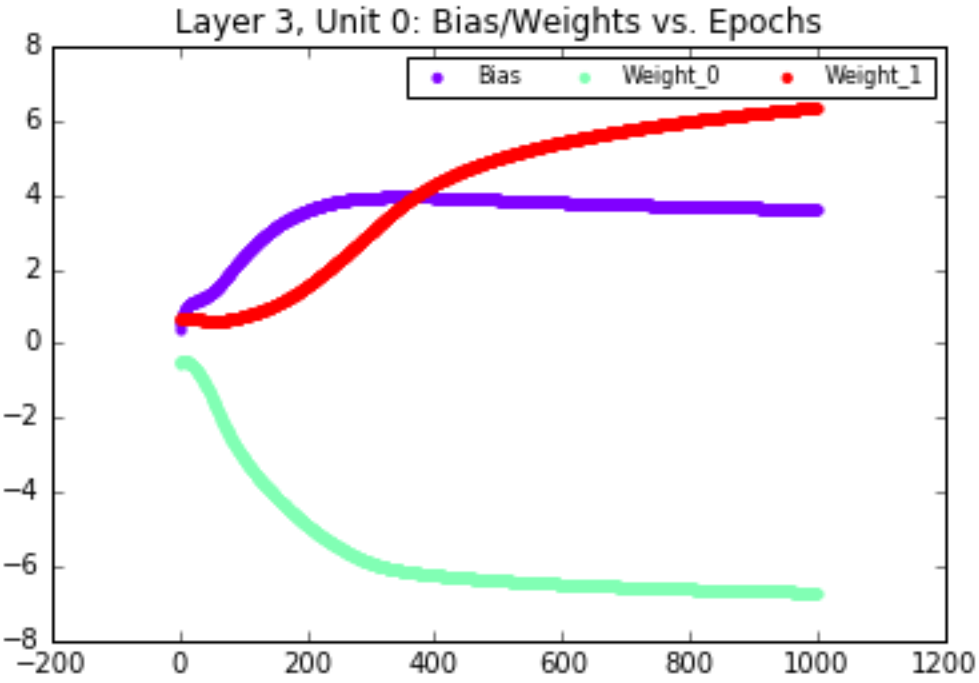


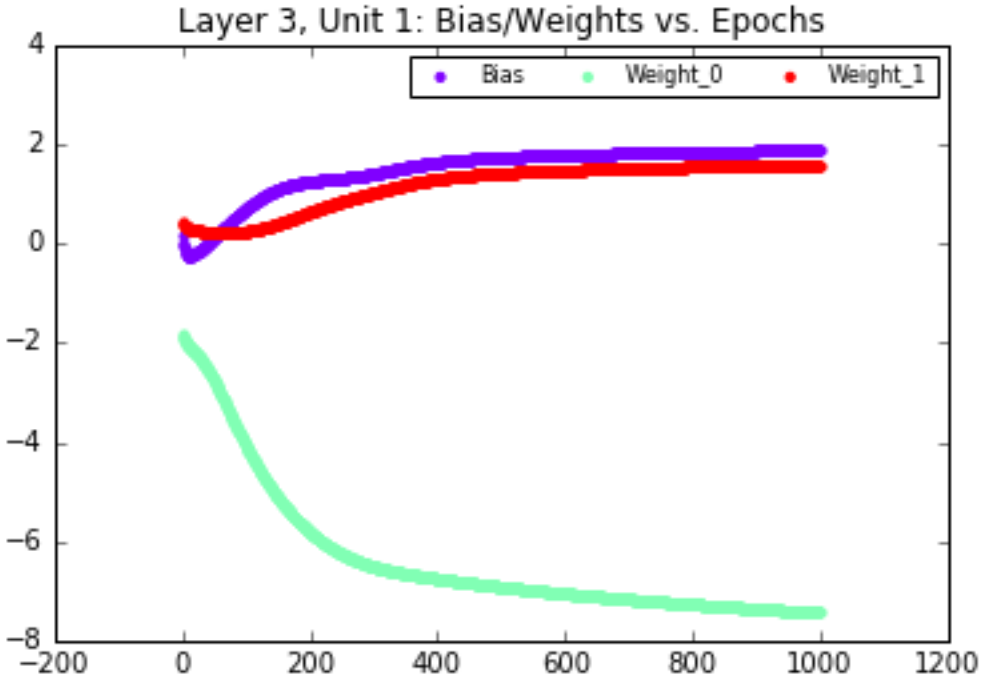

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
5
4
3
2
1
0
−1
−2
−3
−200
0
200
400
600
800
1000
1200

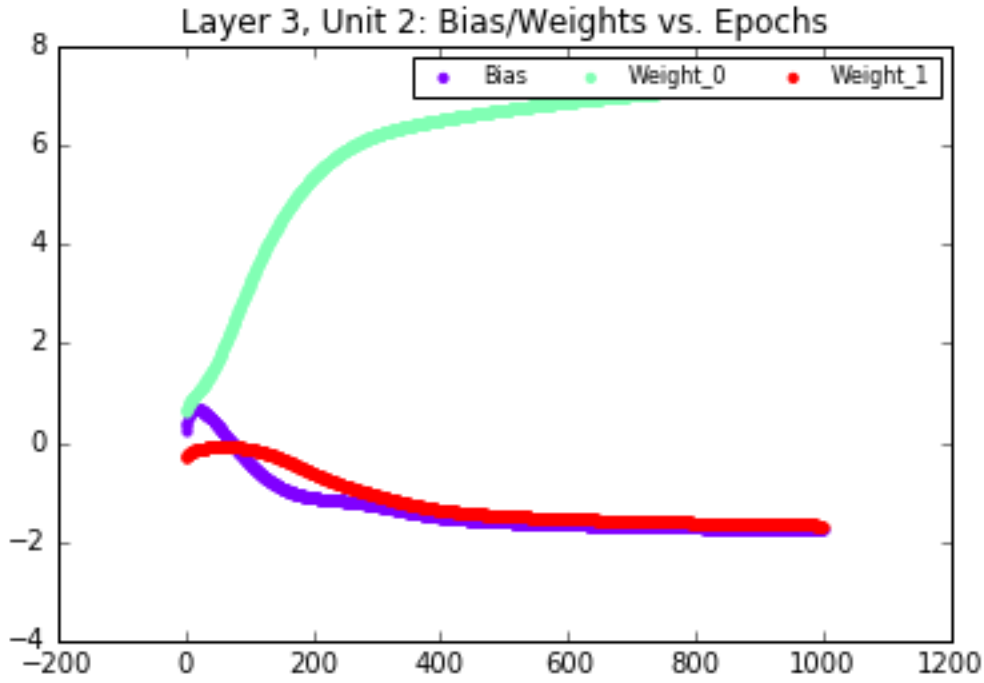

Layer 1, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
3
2
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

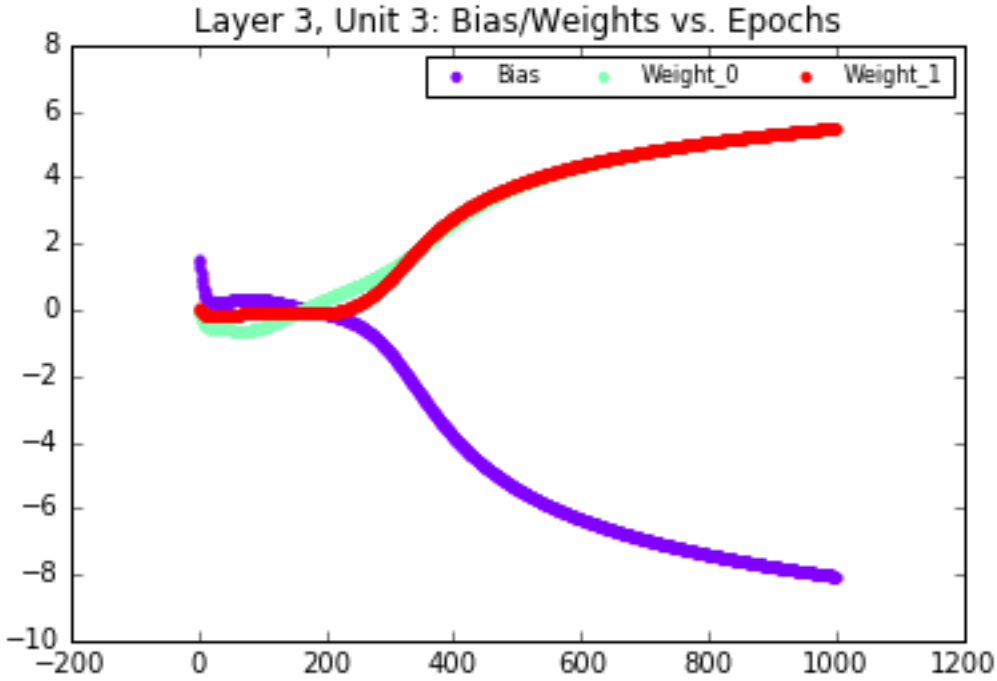

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

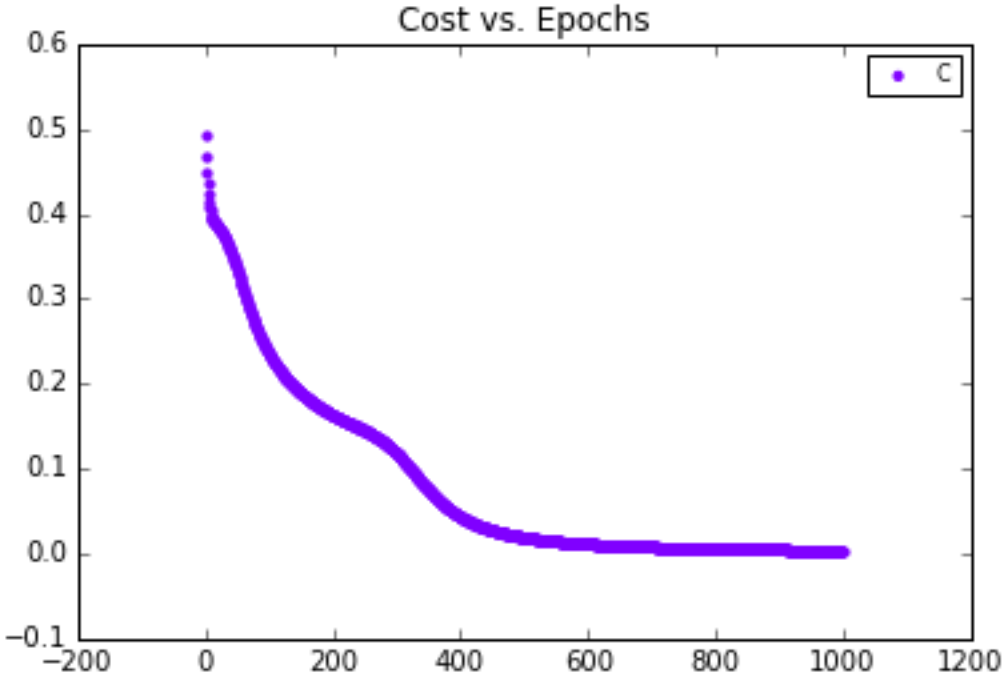

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

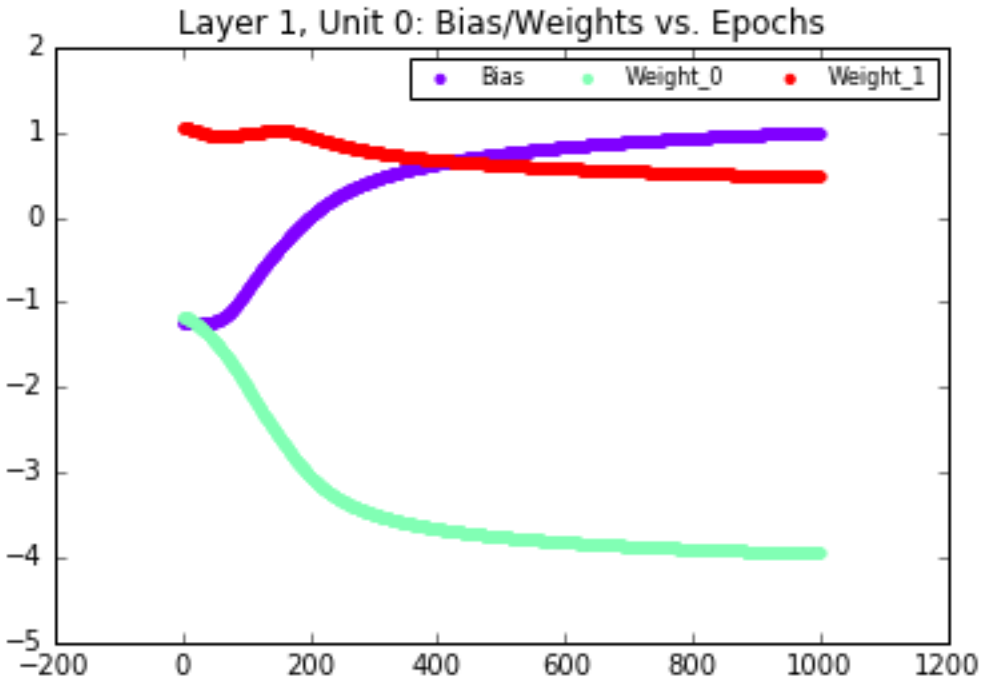

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

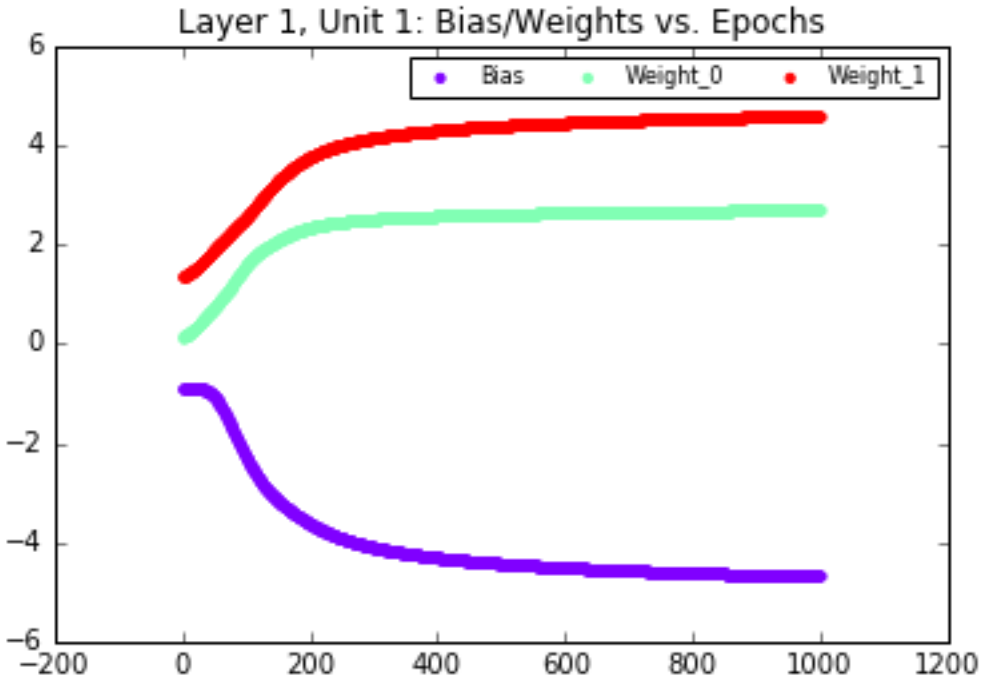

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

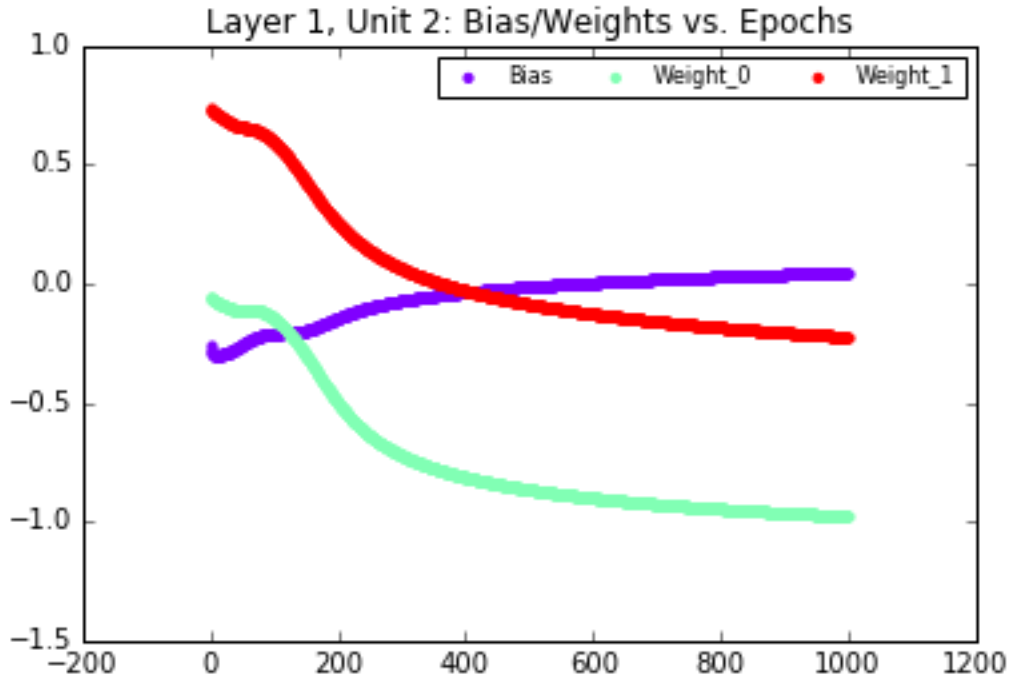

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
8
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

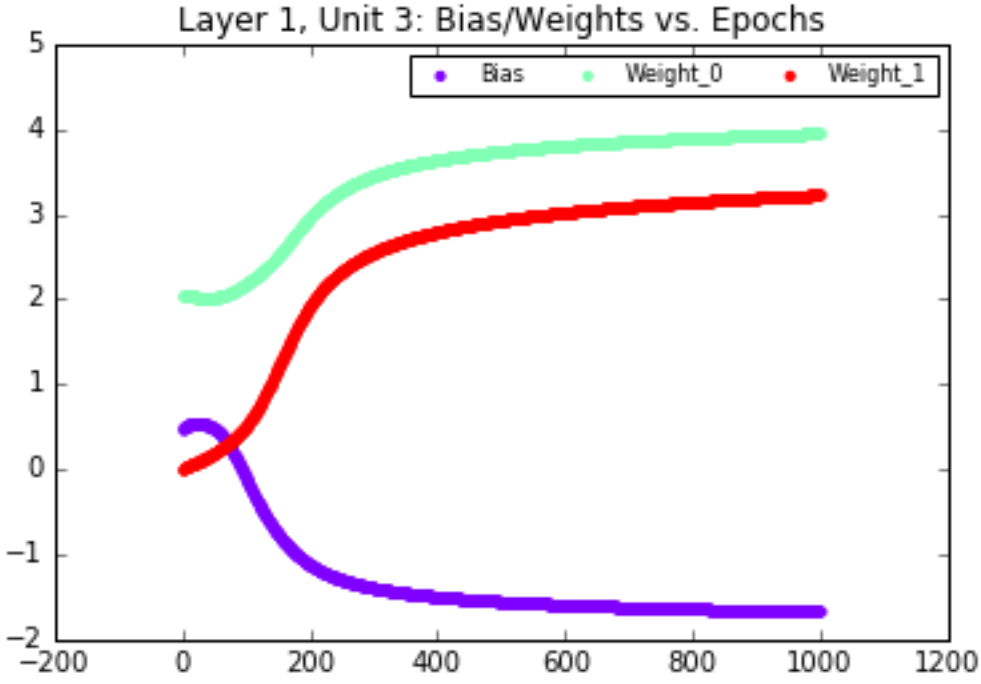

Cost vs. Epochs
C
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

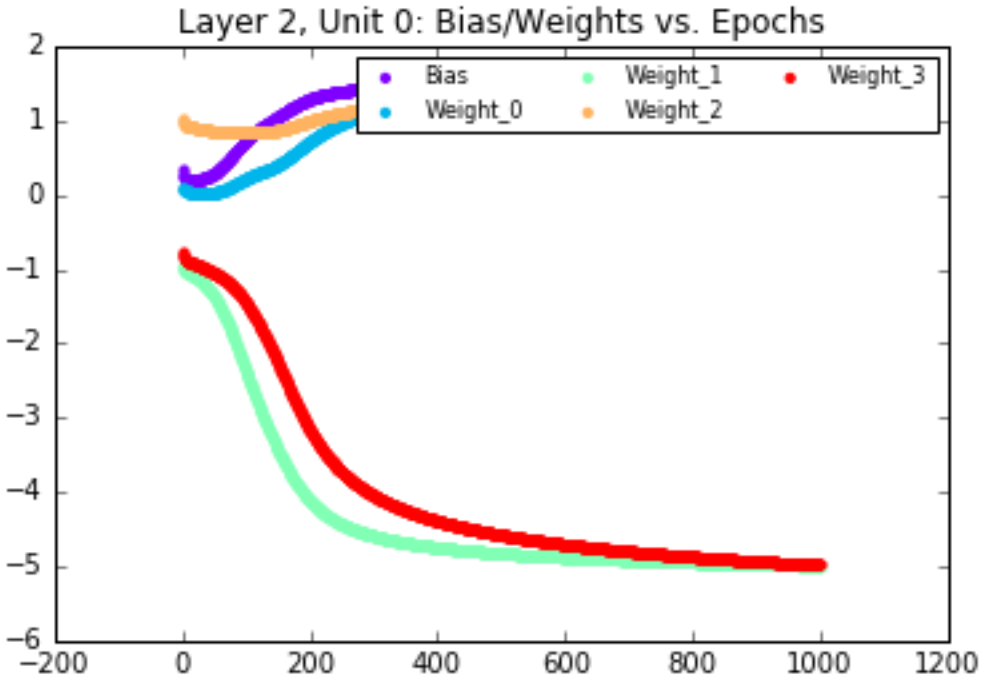
2 inputs, 2 hidden layers, first of size 4, second of size 3, 4 output units

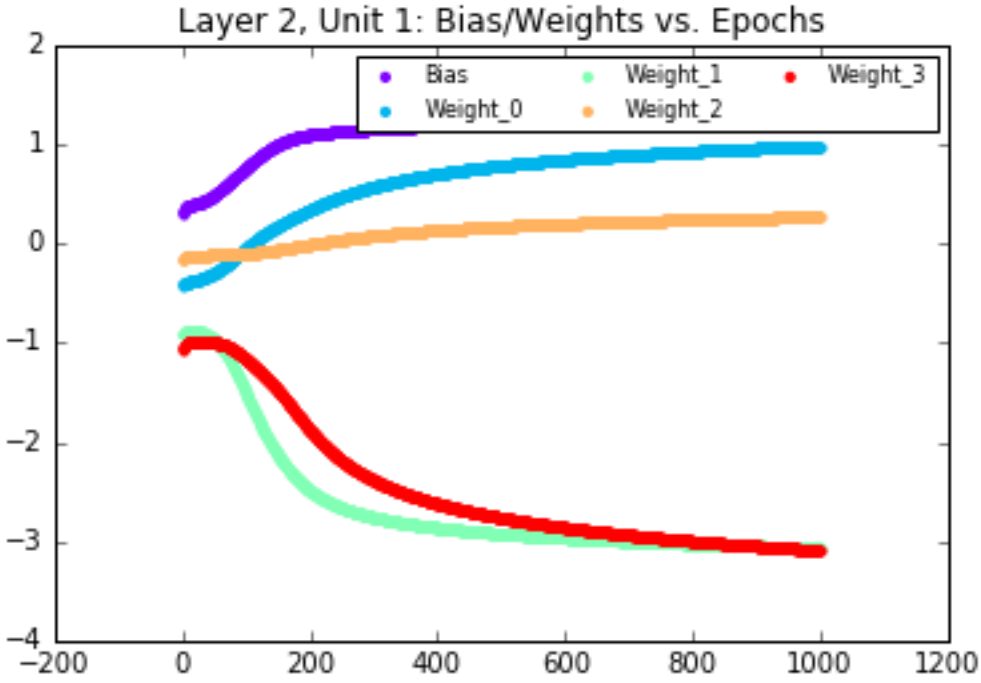


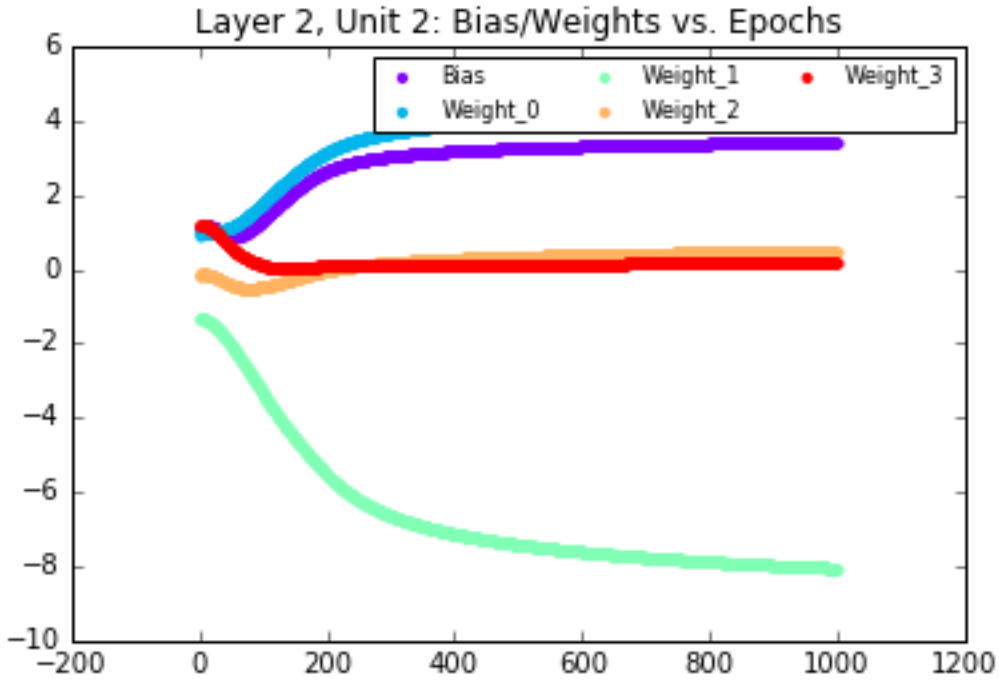


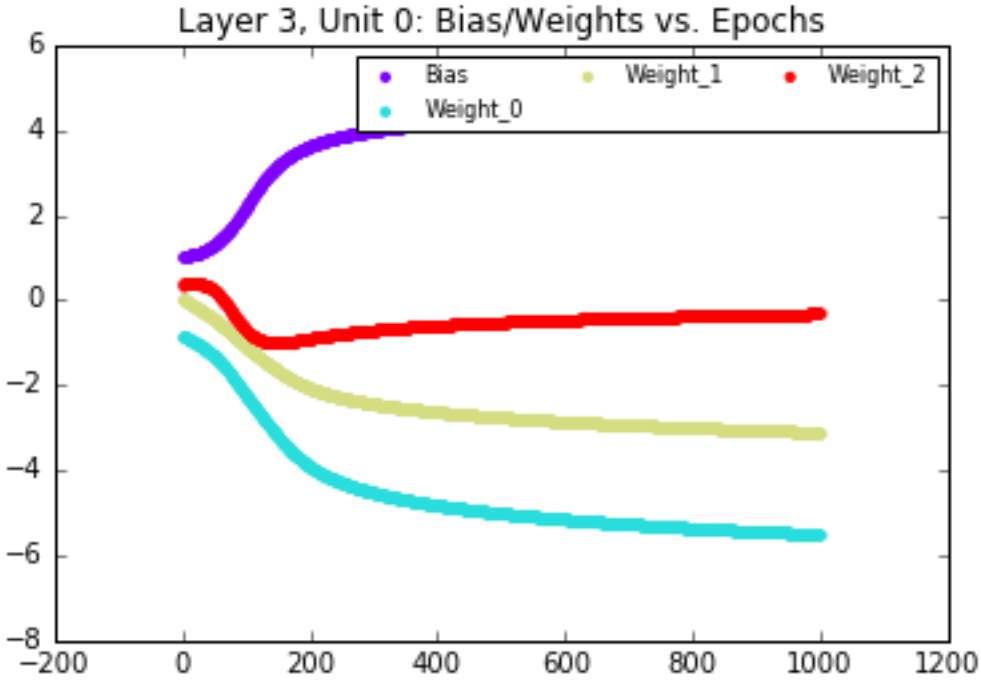

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
1.0
0.5
0.0
−0.5
−1.0
−1.5
−200
0
200
400
600
800
1000
1200

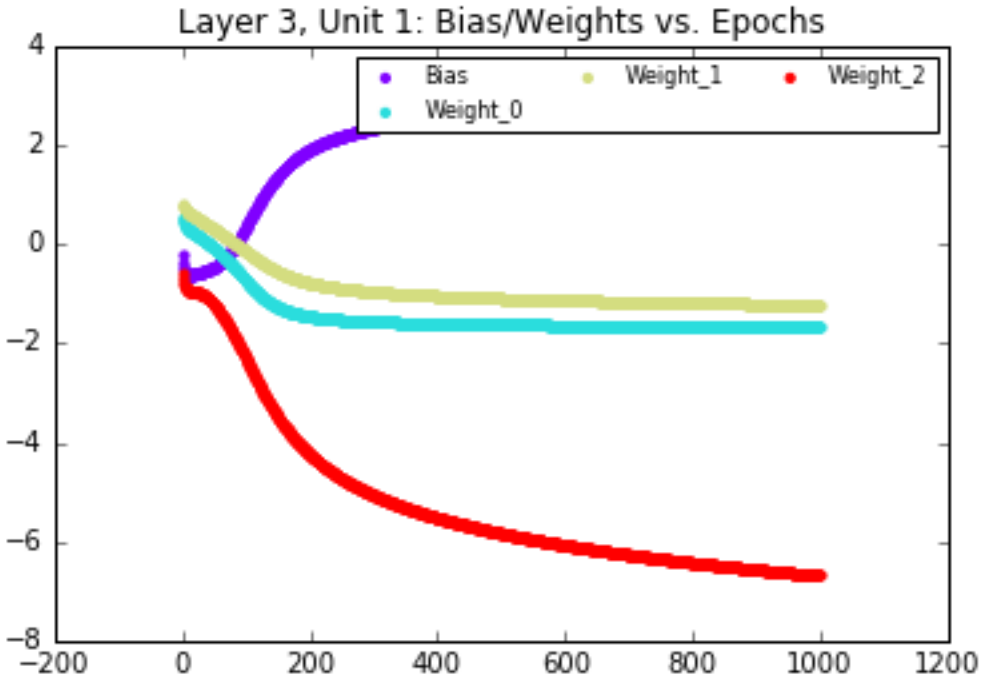

Layer 1, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
5
4
3
2
1
0
−1
−2
−200
0
200
400
600
800
1000
1200

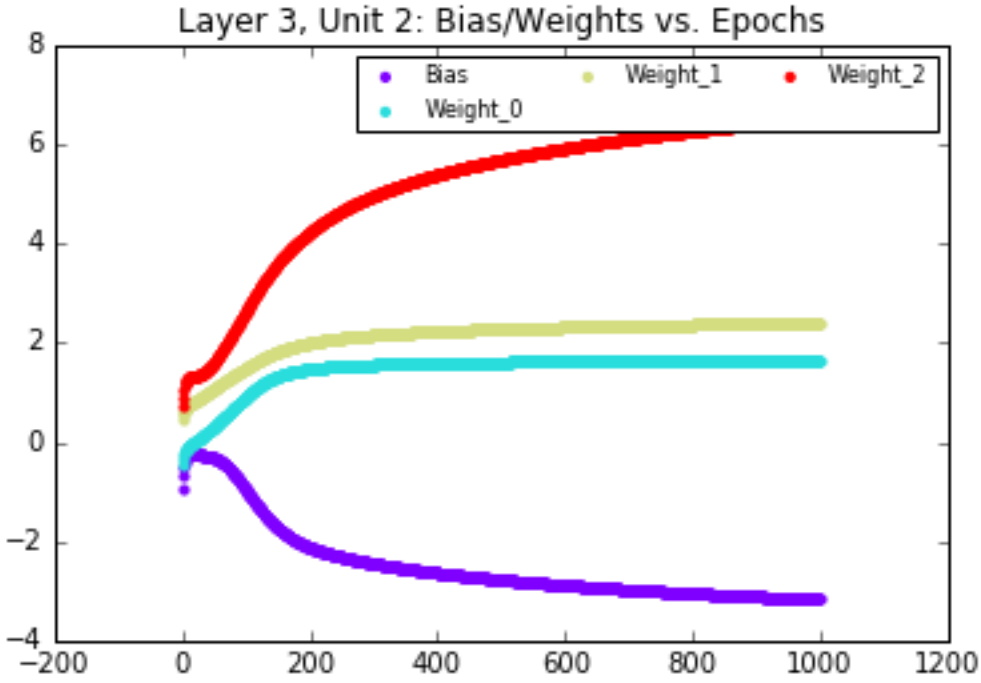

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
2
1
0
−1
−2
−3
−4
−5
−6
−200
0
200
400
600
800
1000
1200

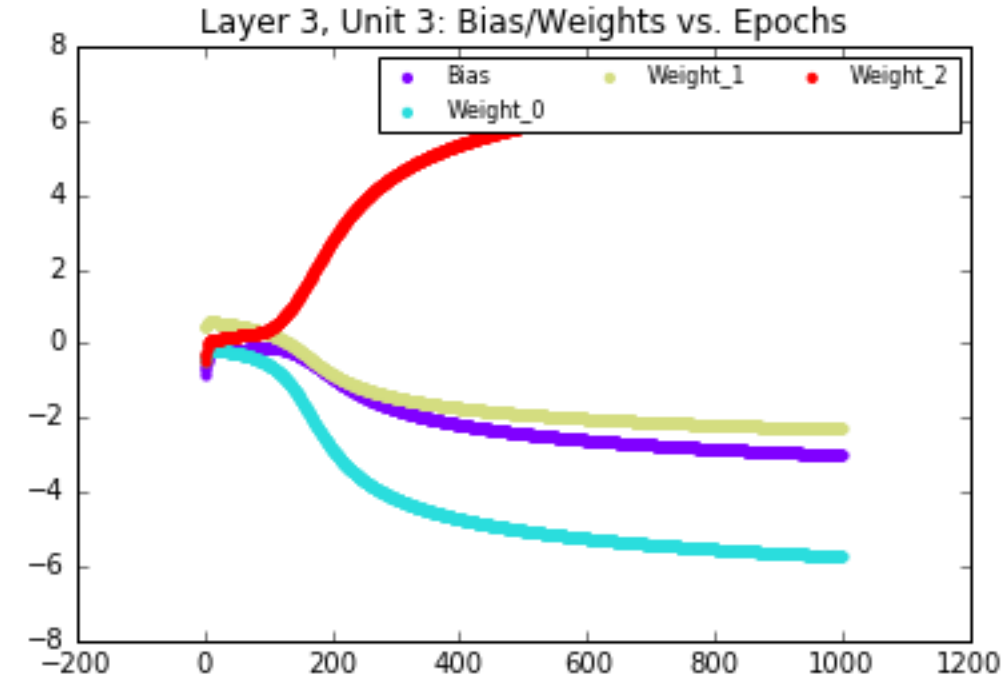

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
2
1
0
−1
−2
−3
−4
−200
0
200
400
600
800
1000
1200

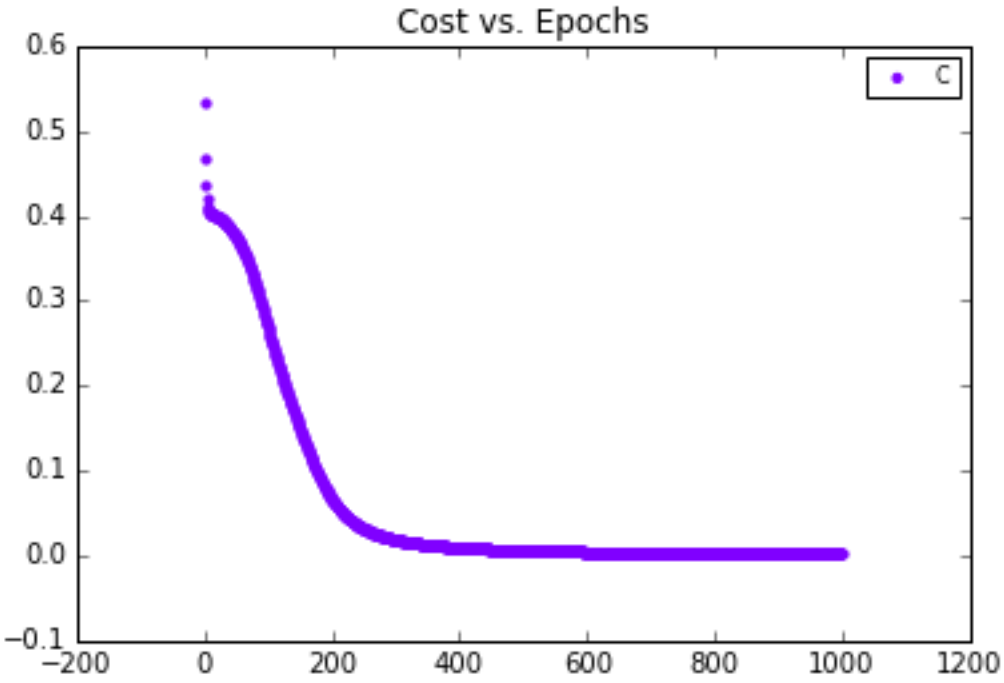

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

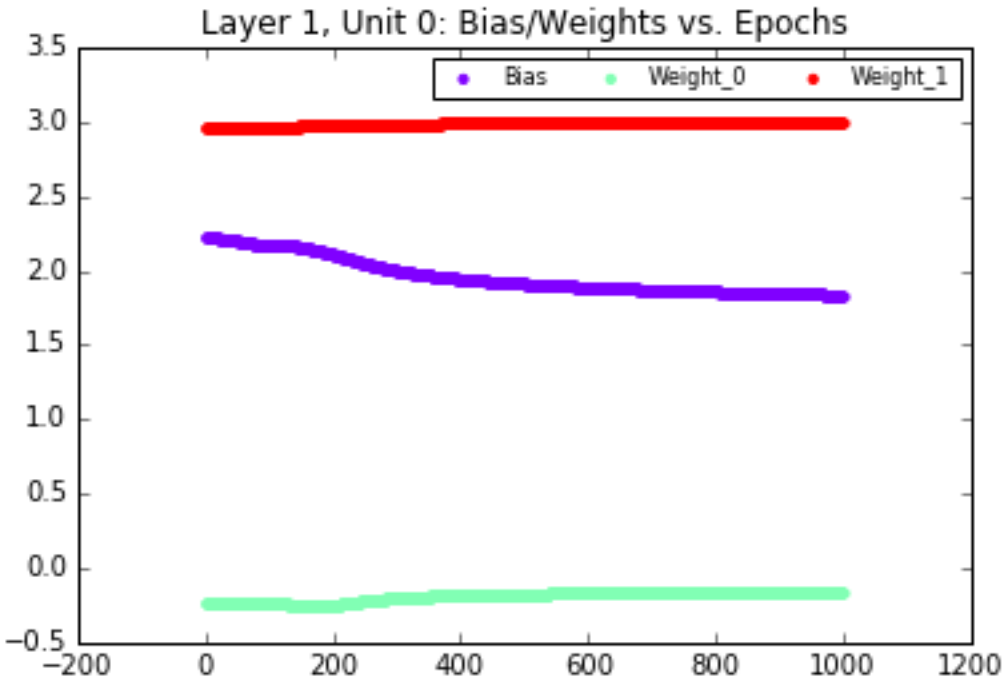

Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

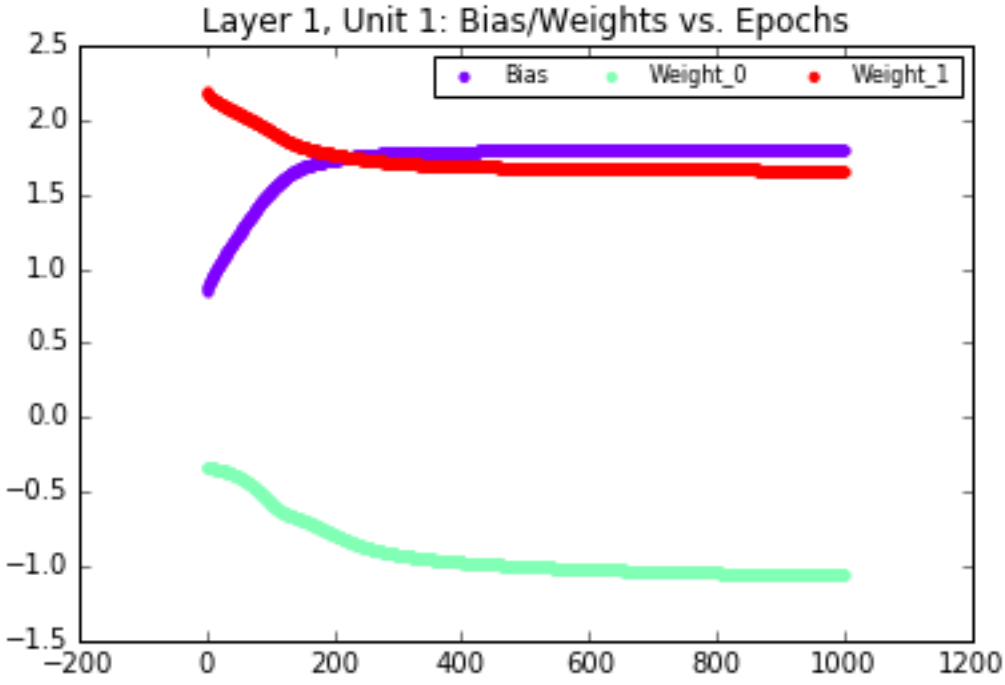

Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200

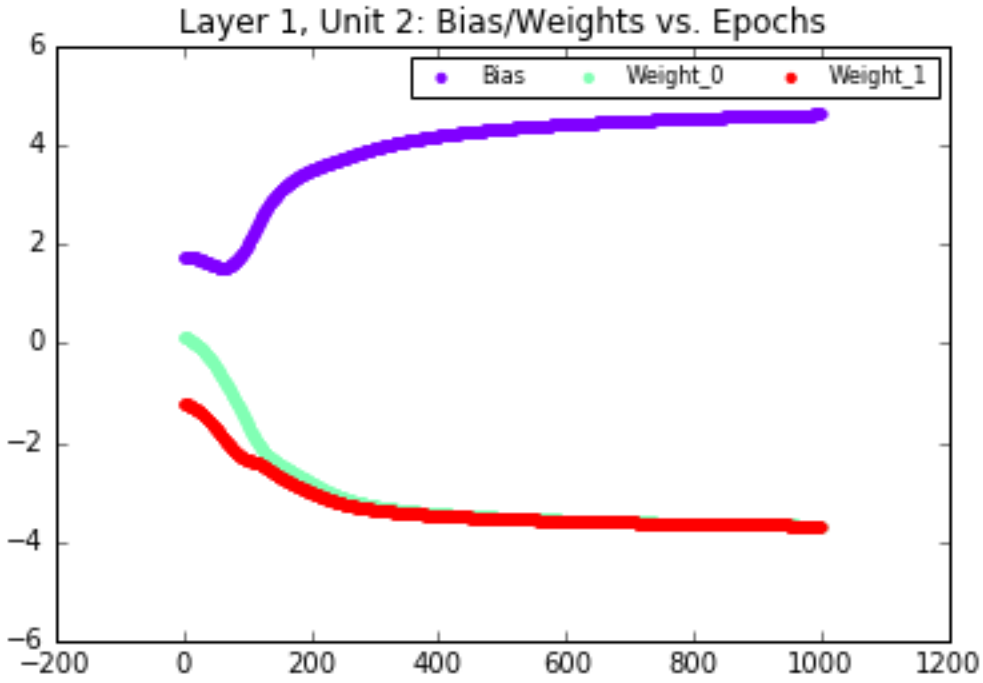

Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_1
Weight_2
Weight_0
8
6
4
2
0
−2
−4
−6
−8
−200
0
200
400
600
800
1000
1200

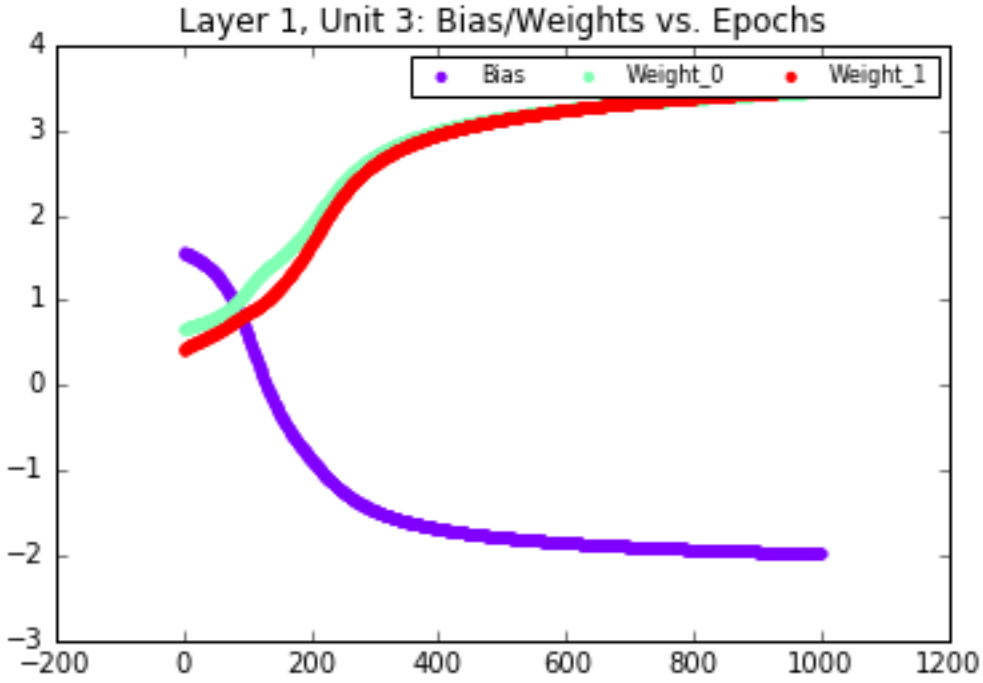

Cost vs. Epochs
C
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200

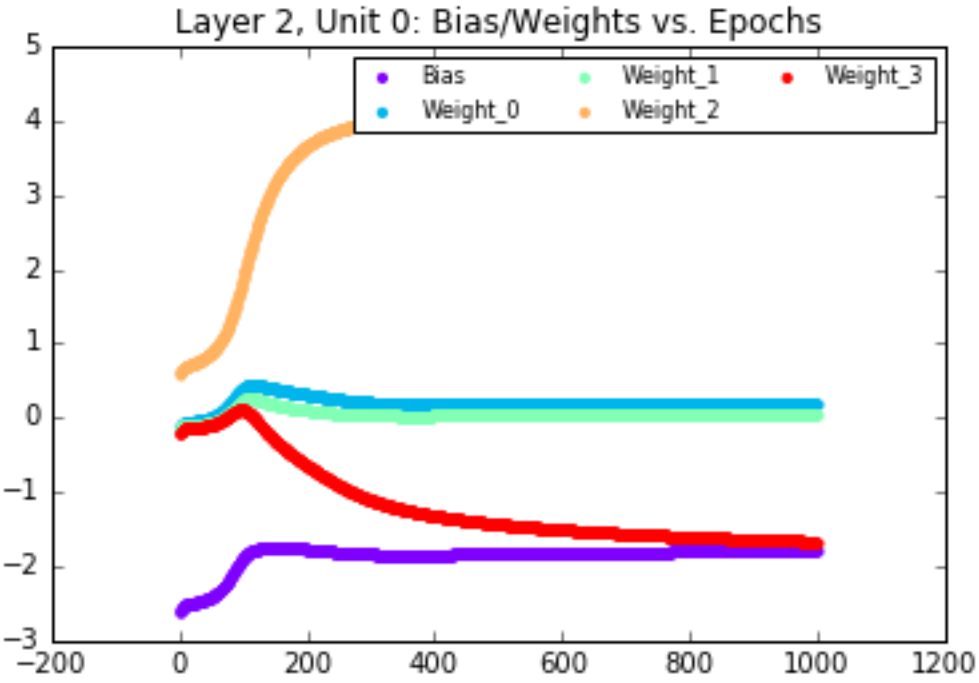
2 inputs, 2 hidden layers, first of size 4, second of size 4, 4 output units

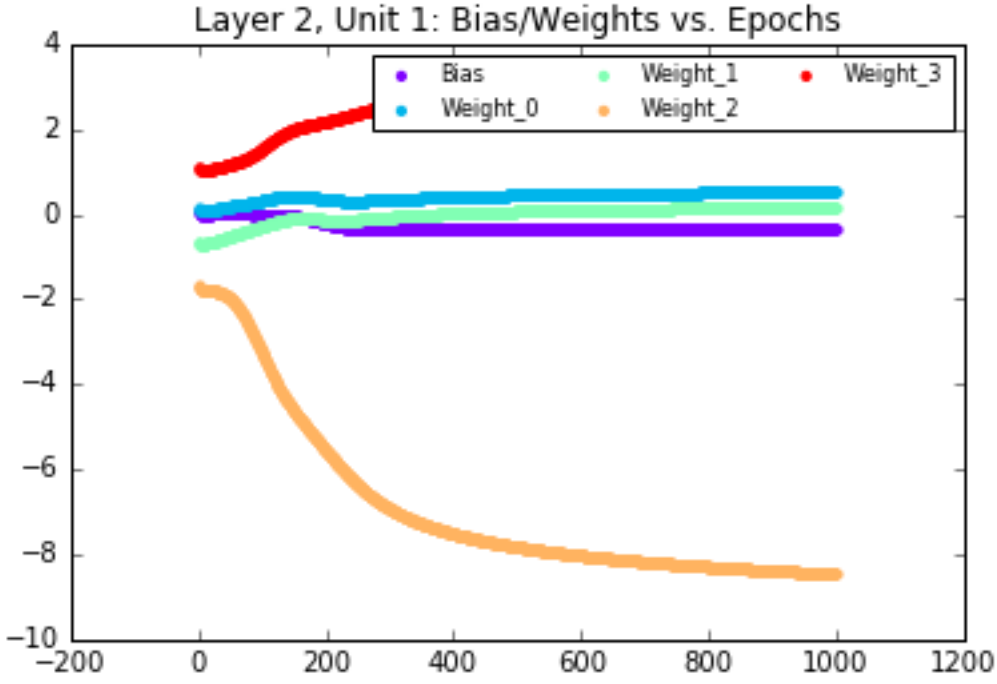


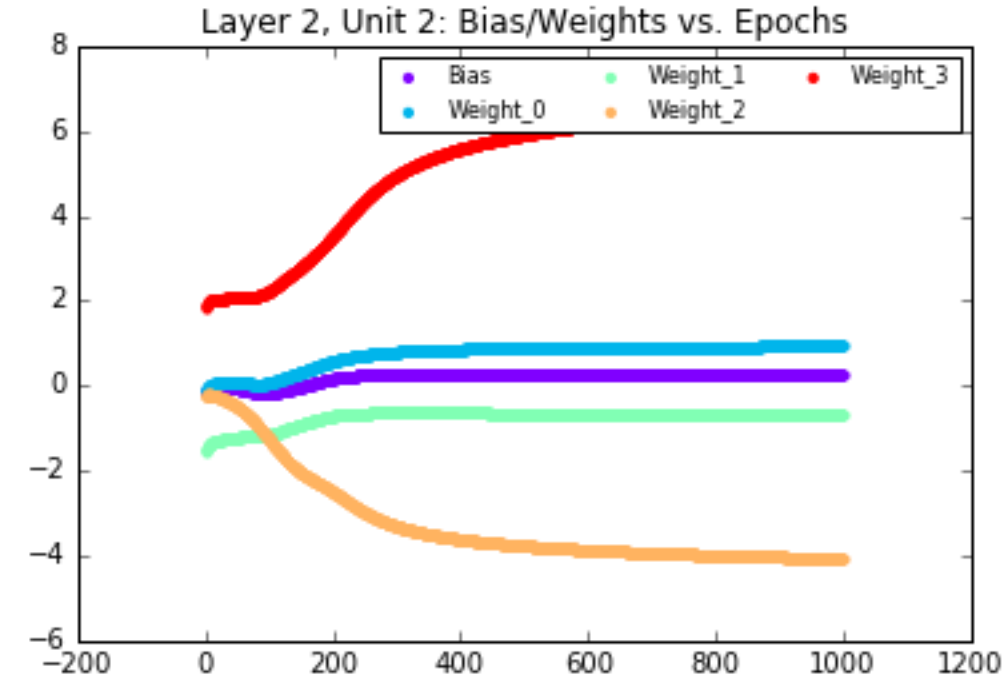


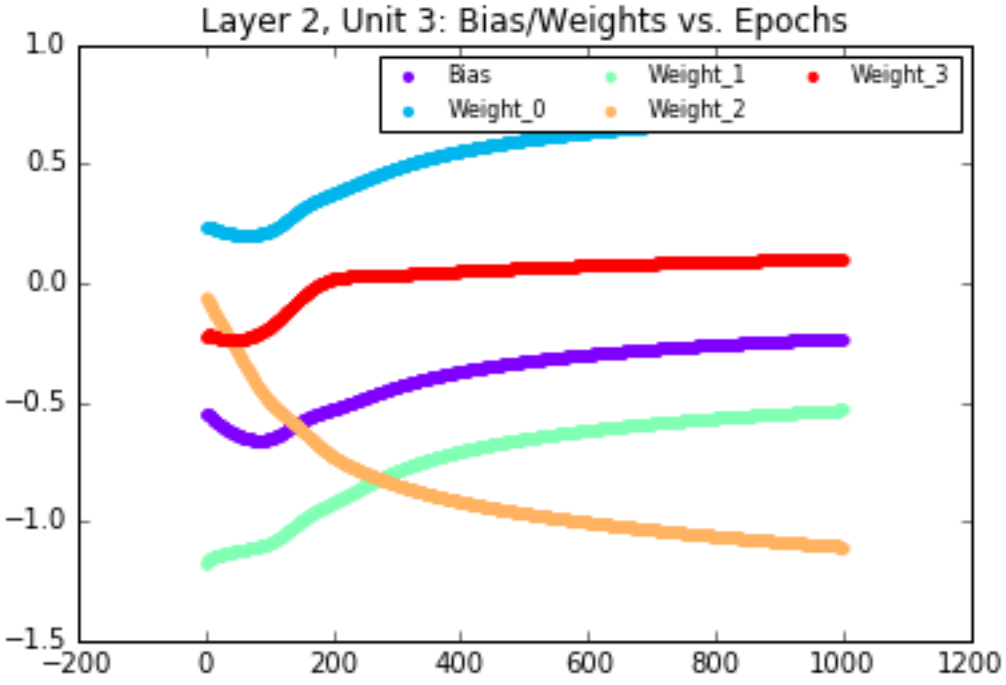

Layer 1, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

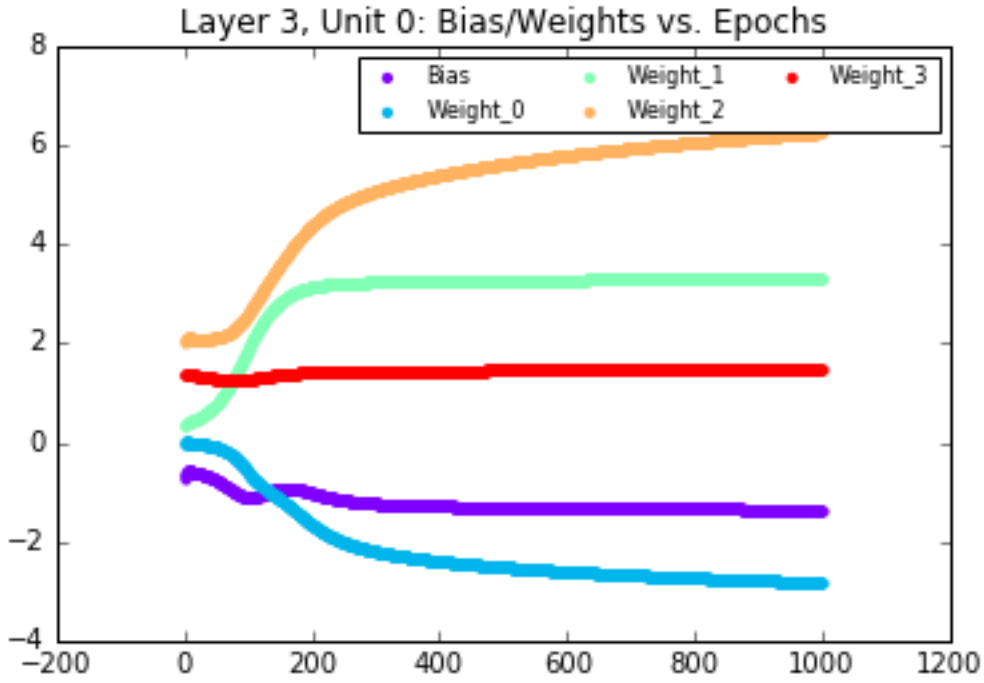

Layer 1, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
4
3
2
1
0
−1
−2
−3
−200
0
200
400
600
800
1000
1200

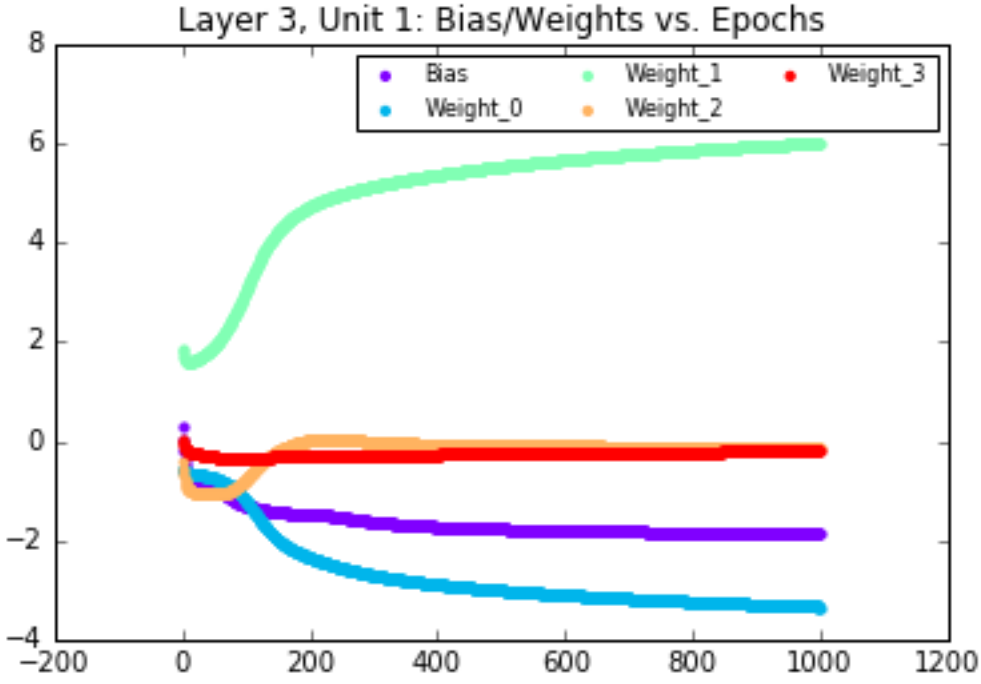

Layer 2, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
5
4
3
2
1
0
−1
−2
−3
−200
0
200
400
600
800
1000
1200

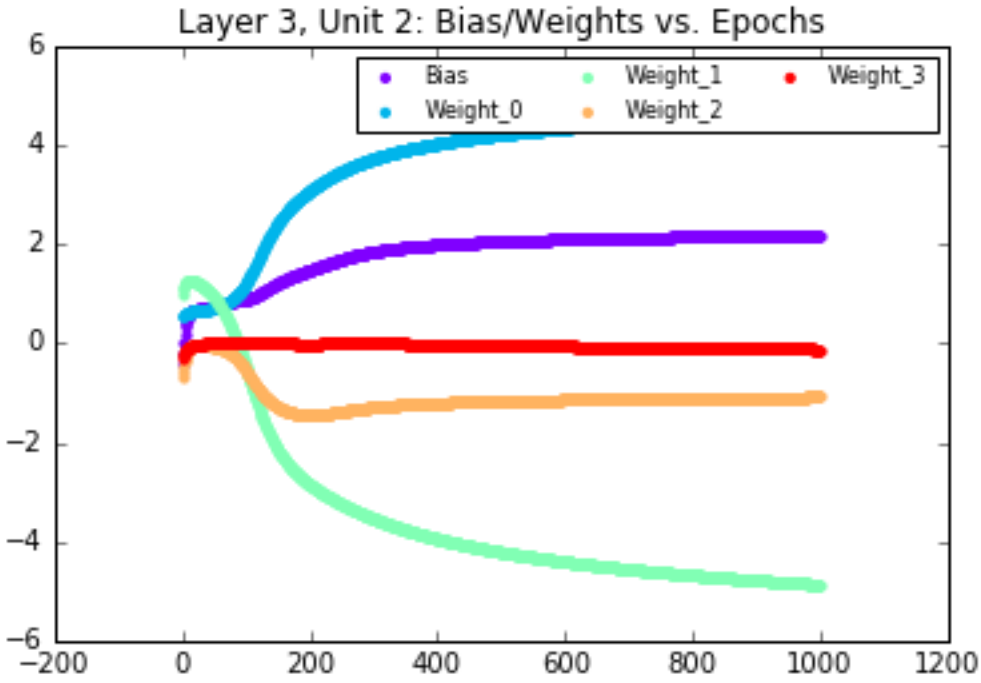

Layer 2, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200

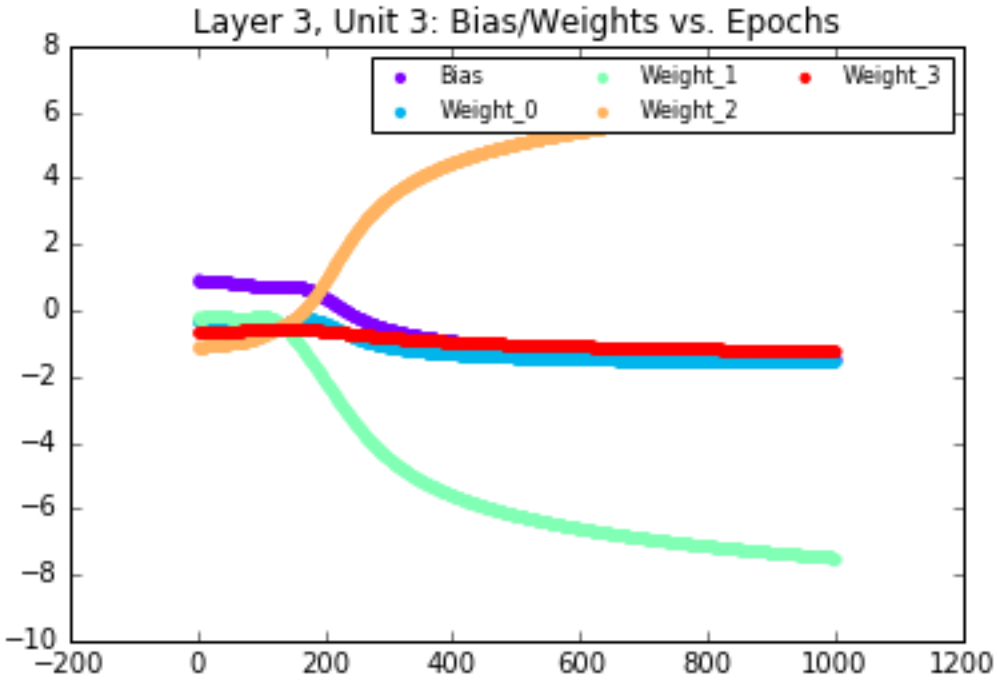

Layer 2, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200

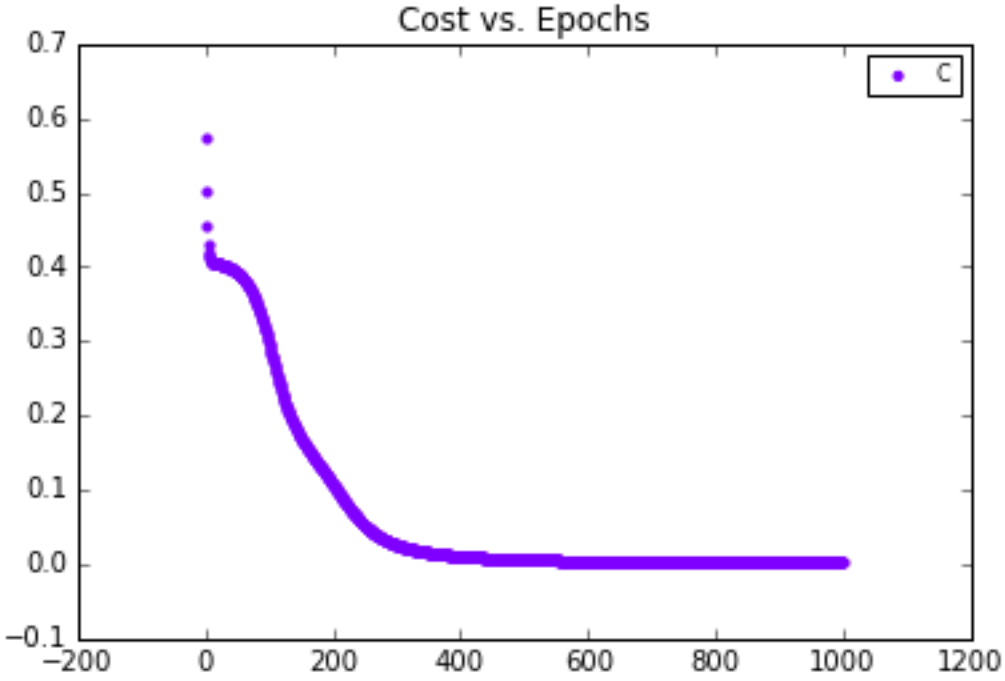

Layer 2, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
1.0
0.5
0.0
−0.5
−1.0
−1.5
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 0: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 1: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 2: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
6
4
2
0
−2
−4
−6
−200
0
200
400
600
800
1000
1200


Layer 3, Unit 3: Bias/Weights vs. Epochs
Bias
Weight_0
Weight_1
Weight_2
Weight_3
8
6
4
2
0
−2
−4
−6
−8
−10
−200
0
200
400
600
800
1000
1200


Cost vs. Epochs
C
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−200
0
200
400
600
800
1000
1200