**Checklist:**

[ ] Early stopping YES

[ ] Cost functions

[ ] Quadratic YES

[ ] Cross-entropy YES

[ ] log-likelihood YES

[ ] allow choice of cost function with a parameter YES

[ ] Momentum YES

[ ] L2 Regularization YES

[ ] Better initial weights YES

[ ] Transfer functions

[ ] tanh YES

[ ] softmax YES

[ ] ReLU YES

[ ] Minibatch shuffling YES

[ ] Learning rate schedule description YES

[ ] Returning learned network YES

[ ] Returning accuracy and costs for plotting YES

[ ] Did NOT include the MNIST data with my submission YES

Incomplete, Details: Running my network with the combination of a softmax transfer function and a log likelihood cost function is not working as expected.

Also, I have not yet plotted the performance for the data partitions in the early stopping scenario.

Not sure, Details \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Learning Rate Schedule:**

*Note – it wasn’t clear where the discussion on learning rate schedule should go so I’ve decided to just put it here. Also note that I’ve added an additional section toward the end to demonstrate the early stopping criteria as requested.*

Setting a variable learning rate schedule can be done in much the same was as early stopping. Specifically, we keep the learning rate constant until the validation accuracy starts to get worse. Then, we start decreasing the rate by a suggested factor or 2 or 10 and continue doing this with each epoch until we terminate the training.

**Instructions:**

1. Ensure that all three datasets along with “net2.m” and “call\_function.m” are saved in the present working directory
2. Open the “call\_function.m” program
3. Choose the dataset you would like to evaluate by uncommenting the corresponding block of code
4. Set the desired network input parameters
5. Execute “call\_function.m” to pass the parameters to the “net2.m” function

**Description:**

The first part of the function – rows 5 through 49 – are responsible for defining the inputs to the function as passed in the parameters. For example, the “input” (non-target) values are set for the train, test, and validation datasets based on the values passed in the “split” variable. Rows 29 through 49 set the initial weights and biases – allowing for the possibility that these may be passed into the function as parameters.

Row 61 begins the backpropagation algorithm where the first step is completing the tasks that need to occur with each epoch. We shuffle the training input and training targets in rows 65 through 68. In order to calculate momentum, we initialize and empty “velocity cell” with zeros in rows 81 through 85. We also initialize empty cells for errors and weighted inputs.

Row 99 begins the processing of the minibatches. This section should look similar to project one except we determine the activation for each layer in rows 116 through 130 based on the activation function passed as a parameter. Rows 144 through 180 are responsible for calculating the error at the output layer. This is – in principle – the same as what was done for project one. However, we now calculate the error at the output layer differently depending on the cost function and activation passed as parameters.

Once the error at the output layer has been calculated, rows 186 through 190 calculate the number of correct “guesses” on the training data for that minibatch and adds it to the total correct guesses for the training input.

In order to complete L2 regularization, we need to know the sum of the squared weights in the network – this is calculated for each minibatch in rows 195 through 198. We next calculate the cost on the training data for each minibatch in rows 200 through 206 – again allowing different cost calculations based on the input parameters.

After backpropogating the error in the output layer backward through the rest of the network, we net update the weights and biases using stochastic gradient descent in rows 214 through 228. One difference from project one is that we first apply momentum to the weights based on their previous velocities in rows 217 through 219.

Next, we want to know the performance of the model on the test and validation inputs. Rows 239 through 260 pass the test input through the network in its current state – again allowing for different activations based on the input parameters. Similarly, rows 271 through 301 pass the validation data through the current network.

We next need to calculate the cost. Having collected the incremental contribution to the cost for the training data in each minibatch, we get the cost for the entire training input in rows 305 through 309. Calculating cost for the training and validation sets occurs only at the very end of the epoch and not incrementally like the training input. Because we’re using L2 regularization, we need to recalculate the sum of the squared weights of the full network – this takes place in rows 317 through 320. We can then calculate the cost on the training data in rows 322 through 328 based on the cost function passed as a parameter. Calculating the cost on the validation input in done in a similar way in rows 337 through 343. One difference here is that we first check to confirm whether the cost on the validation data has gone up for a predetermined number of epochs for the stopping criteria. We define the previous cost in rows 331 through 334 and then compare it to the new cost in rows 345 through 349.

Finally, we’re ready to calculate the correct guesses for the test and validation data similar to what we had done for the training data – this occurs in rows 360 through 371. Rows 374 through 376 add the results for the train, test, and validate inputs to respective lists so that they can be returned by the function for graphing purposes. Rows 379 through 382 then print the results for the epoch.

**Code:**

Note – double-click into the object below to scroll through the code

****

**Analysis:**

One thing that is clear is that my network does not perform well with the hyperparameters required for this assignment. Specially, the network needs a much lower lambda value (~.001) in order for learning to occur. However, having read the discussion forum I am not necessarily concerned that this is an issue. Another thing that came up relates to updating the weights when both momentum and L2 regularization are in use. The reason for this is that each of these two methods have their own calculations for updating weights – should these be combined or used in sequence? My network currently first updates weights based on momentum and then updates using the rules for L2 regularization, but I think that these two more likely should be performed in parallel. However, I was unsure how to combine these two formulas without impacting the overall results.

I also noticed that my network is not calculating costs when the ReLU activation function is used with either the log-likelihood or cross entropy cost functions. The reason for this is that, in the event that the weighted input is less than zero, the activation will be zero when ReLU activation function is used. Both the log-likelihood and cross entropy cost functions require that you take the natural log of the activations, which when the activation is zero does not exist. I am not sure what I need to do to update my network to account for this.

Another obvious issue is that my network is not working when log-likelihood cost function is used together with the softmax activation function. At this point I am not sure why this is, but I am out of time and cannot look into it any further prior to the project deadline.

**Ideas for enhancements:**

My current network can certainly be enhanced by fine tuning the hyperparameters. Although the framework has been set to experiment with different values and network types, this is far from a trivial exercise.

**Early stopping:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| MNIST | 40 | [3 2] | 1 | .1 | Sigmoid | quad | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | 0.461609 | 0/40000 | 0.000000 ||567.296277 | 0/5000 | 0.000000 || 547.560898 | 0/5000 | 0.000000

2 | 0.456530 | 0/40000 | 0.000000 ||569.002584 | 0/5000 | 0.000000 || 546.891584 | 0/5000 | 0.000000

3 | 0.456494 | 0/40000 | 0.000000 ||567.468377 | 0/5000 | 0.000000 || 548.415464 | 0/5000 | 0.000000

4 | 0.456396 | 0/40000 | 0.000000 ||568.030518 | 0/5000 | 0.000000 || 549.618533 | 0/5000 | 0.000000

5 | 0.456353 | 0/40000 | 0.000000 ||568.045226 | 0/5000 | 0.000000 || 549.638483 | 0/5000 | 0.000000

Early stopping due to consecutive increases in the cost on the validation set.

**Outputs:**

Note – double-click into the objects to scroll through the results

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| iris.csv | 40 | 20 | 10 | .1 | Sigmoid | Cross | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | 957.647460 | 37/120 | 0.308333 ||3904.576741 | 6/15 | 0.400000 || 3905.139823 | 7/15 | 0.466667

2 | 358.433843 | 37/120 | 0.308333 ||1811.876448 | 6/15 | 0.400000 || 1812.362604 | 7/15 | 0.466667

3 | 188.971616 | 37/120 | 0.308333 ||1120.654711 | 6/15 | 0.400000 || 1121.258780 | 7/15 | 0.466667

4 | 134.694180 | 37/120 | 0.308333 ||899.281473 | 6/15 | 0.400000 || 899.825279 | 7/15 | 0.466667

5 | 116.407969 | 27/120 | 0.225000 ||774.299195 | 6/15 | 0.400000 || 774.883854 | 7/15 | 0.466667

6 | 104.189206 | 37/120 | 0.308333 ||840.734839 | 3/15 | 0.200000 || 841.077163 | 3/15 | 0.200000

7 | 101.531597 | 35/120 | 0.291667 ||816.363850 | 3/15 | 0.200000 || 816.758977 | 3/15 | 0.200000

8 | 102.887159 | 30/120 | 0.250000 ||700.869499 | 6/15 | 0.400000 || 701.344171 | 7/15 | 0.466667

9 | 96.011205 | 37/120 | 0.308333 ||740.611498 | 5/15 | 0.333333 || 740.985127 | 7/15 | 0.466667

10 | 95.219404 | 32/120 | 0.266667 ||732.510937 | 6/15 | 0.400000 || 733.067826 | 7/15 | 0.466667

11 | 95.922202 | 28/120 | 0.233333 ||692.121025 | 6/15 | 0.400000 || 692.660323 | 7/15 | 0.466667

12 | 91.906102 | 37/120 | 0.308333 ||777.781127 | 0/15 | 0.000000 || 778.132388 | 1/15 | 0.066667

13 | 94.404284 | 24/120 | 0.200000 ||721.697659 | 4/15 | 0.266667 || 722.116195 | 7/15 | 0.466667

14 | 96.676852 | 23/120 | 0.191667 ||575.196999 | 6/15 | 0.400000 || 575.901558 | 7/15 | 0.466667

15 | 89.166924 | 36/120 | 0.300000 ||711.126553 | 4/15 | 0.266667 || 711.546295 | 7/15 | 0.466667

16 | 95.199749 | 25/120 | 0.208333 ||602.288008 | 6/15 | 0.400000 || 602.762081 | 7/15 | 0.466667

17 | 88.370705 | 36/120 | 0.300000 ||683.976194 | 6/15 | 0.400000 || 684.477390 | 7/15 | 0.466667

18 | 90.229506 | 28/120 | 0.233333 ||671.380216 | 6/15 | 0.400000 || 671.932925 | 7/15 | 0.466667

19 | 89.247783 | 33/120 | 0.275000 ||677.887677 | 6/15 | 0.400000 || 678.439368 | 7/15 | 0.466667

20 | 87.194850 | 28/120 | 0.233333 ||749.647607 | 0/15 | 0.000000 || 750.050604 | 1/15 | 0.066667

21 | 93.435943 | 18/120 | 0.150000 ||563.279602 | 6/15 | 0.400000 || 563.931091 | 7/15 | 0.466667

22 | 87.953132 | 30/120 | 0.250000 ||593.409060 | 6/15 | 0.400000 || 593.950473 | 7/15 | 0.466667

23 | 85.919064 | 37/120 | 0.308333 ||653.479425 | 6/15 | 0.400000 || 653.939462 | 7/15 | 0.466667

24 | 87.044792 | 28/120 | 0.233333 ||656.296599 | 6/15 | 0.400000 || 656.804704 | 7/15 | 0.466667

25 | 87.990294 | 26/120 | 0.216667 ||620.392998 | 6/15 | 0.400000 || 620.992373 | 7/15 | 0.466667

26 | 83.684625 | 30/120 | 0.250000 ||745.148999 | 0/15 | 0.000000 || 745.447673 | 0/15 | 0.000000

27 | 86.147188 | 20/120 | 0.166667 ||715.047930 | 0/15 | 0.000000 || 715.434210 | 1/15 | 0.066667

28 | 91.990829 | 19/120 | 0.158333 ||509.892018 | 6/15 | 0.400000 || 510.430232 | 7/15 | 0.466667

29 | 78.569714 | 31/120 | 0.258333 ||703.674075 | 0/15 | 0.000000 || 704.037947 | 0/15 | 0.000000

30 | 84.545920 | 25/120 | 0.208333 ||661.153049 | 4/15 | 0.266667 || 661.696984 | 7/15 | 0.466667

31 | 81.782806 | 29/120 | 0.241667 ||714.756832 | 0/15 | 0.000000 || 715.111652 | 0/15 | 0.000000

32 | 85.268764 | 19/120 | 0.158333 ||600.432913 | 6/15 | 0.400000 || 600.978724 | 7/15 | 0.466667

33 | 83.063379 | 26/120 | 0.216667 ||620.223965 | 4/15 | 0.266667 || 620.672736 | 7/15 | 0.466667

34 | 82.742529 | 21/120 | 0.175000 ||594.421847 | 6/15 | 0.400000 || 594.959470 | 7/15 | 0.466667

35 | 78.248083 | 33/120 | 0.275000 ||701.440386 | 0/15 | 0.000000 || 701.952309 | 0/15 | 0.000000

36 | 81.014819 | 25/120 | 0.208333 ||684.956592 | 0/15 | 0.000000 || 685.463846 | 0/15 | 0.000000

37 | 82.406137 | 22/120 | 0.183333 ||616.193329 | 6/15 | 0.400000 || 616.758388 | 7/15 | 0.466667

38 | 80.782370 | 24/120 | 0.200000 ||603.541664 | 4/15 | 0.266667 || 604.006445 | 7/15 | 0.466667

39 | 80.557711 | 32/120 | 0.266667 ||587.732127 | 4/15 | 0.266667 || 588.094537 | 7/15 | 0.466667

40 | 79.313703 | 26/120 | 0.216667 ||608.830622 | 2/15 | 0.133333 || 609.132120 | 2/15 | 0.133333

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| iris.csv | 40 | 20 | 10 | .1 | Relu | Cross | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

2 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

3 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

4 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

5 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

6 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

7 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

8 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

9 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

10 | NaN | 8/120 | 0.066667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

11 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

12 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

13 | NaN | 6/120 | 0.050000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

14 | NaN | 3/120 | 0.025000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

15 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

16 | NaN | 0/120 | 0.000000 ||NaN | 7/15 | 0.466667 || NaN | 2/15 | 0.133333

17 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

18 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

19 | NaN | 7/120 | 0.058333 ||NaN | 7/15 | 0.466667 || NaN | 2/15 | 0.133333

20 | NaN | 3/120 | 0.025000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

21 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

22 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

23 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

24 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

25 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

26 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

27 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

28 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

29 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

30 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

31 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

32 | NaN | 9/120 | 0.075000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

33 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

34 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

35 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

36 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

37 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

38 | NaN | 3/120 | 0.025000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

39 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

40 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| iris.csv | 40 | 20 | 10 | .1 | Relu | Cross | 0 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | NaN | 3/120 | 0.025000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

2 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

3 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

4 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

5 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

6 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

7 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

8 | NaN | 6/120 | 0.050000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

9 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

10 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

11 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

12 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

13 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

14 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

15 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

16 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

17 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

18 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

19 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

20 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

21 | NaN | 1/120 | 0.008333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

22 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

23 | NaN | 5/120 | 0.041667 ||NaN | 6/15 | 0.400000 || NaN | 4/15 | 0.266667

24 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

25 | NaN | 6/120 | 0.050000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

26 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

27 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

28 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

29 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

30 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

31 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

32 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

33 | NaN | 3/120 | 0.025000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

34 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

35 | NaN | 2/120 | 0.016667 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

36 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

37 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

38 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

39 | NaN | 4/120 | 0.033333 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

40 | NaN | 0/120 | 0.000000 ||NaN | 0/15 | 0.000000 || NaN | 0/15 | 0.000000

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| MNIST | 30 | 30 | 10 | 3 | Sigmoid | quad | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | 0.309118 | 1/40000 | 0.000025 ||558.502458 | 0/5000 | 0.000000 || 538.758274 | 0/5000 | 0.000000

2 | 0.159860 | 0/40000 | 0.000000 ||558.554915 | 0/5000 | 0.000000 || 538.044667 | 0/5000 | 0.000000

3 | 0.153193 | 0/40000 | 0.000000 ||559.154397 | 0/5000 | 0.000000 || 537.426891 | 0/5000 | 0.000000

4 | 0.150203 | 0/40000 | 0.000000 ||558.574095 | 0/5000 | 0.000000 || 537.219938 | 0/5000 | 0.000000

5 | 0.148066 | 0/40000 | 0.000000 ||558.413646 | 0/5000 | 0.000000 || 537.191691 | 0/5000 | 0.000000

6 | 0.145118 | 0/40000 | 0.000000 ||558.414573 | 0/5000 | 0.000000 || 537.495665 | 0/5000 | 0.000000

7 | 0.144186 | 0/40000 | 0.000000 ||558.869165 | 0/5000 | 0.000000 || 538.866308 | 0/5000 | 0.000000

8 | 0.144022 | 0/40000 | 0.000000 ||558.836404 | 0/5000 | 0.000000 || 538.668769 | 0/5000 | 0.000000

9 | 0.141347 | 0/40000 | 0.000000 ||558.441689 | 0/5000 | 0.000000 || 537.216631 | 0/5000 | 0.000000

10 | 0.140761 | 0/40000 | 0.000000 ||559.403229 | 0/5000 | 0.000000 || 538.542029 | 0/5000 | 0.000000

11 | 0.140835 | 0/40000 | 0.000000 ||559.316545 | 0/5000 | 0.000000 || 537.725219 | 0/5000 | 0.000000

12 | 0.140127 | 0/40000 | 0.000000 ||558.503691 | 0/5000 | 0.000000 || 539.516868 | 0/5000 | 0.000000

13 | 0.140310 | 0/40000 | 0.000000 ||558.406055 | 0/5000 | 0.000000 || 537.806274 | 0/5000 | 0.000000

14 | 0.140446 | 0/40000 | 0.000000 ||558.896948 | 0/5000 | 0.000000 || 539.370424 | 0/5000 | 0.000000

15 | 0.138703 | 0/40000 | 0.000000 ||558.684010 | 0/5000 | 0.000000 || 539.136595 | 0/5000 | 0.000000

16 | 0.138959 | 0/40000 | 0.000000 ||560.779046 | 0/5000 | 0.000000 || 537.915869 | 0/5000 | 0.000000

17 | 0.138601 | 0/40000 | 0.000000 ||558.515653 | 0/5000 | 0.000000 || 537.649856 | 0/5000 | 0.000000

18 | 0.139085 | 0/40000 | 0.000000 ||559.048579 | 0/5000 | 0.000000 || 537.415986 | 0/5000 | 0.000000

19 | 0.138474 | 0/40000 | 0.000000 ||558.391402 | 0/5000 | 0.000000 || 537.779413 | 0/5000 | 0.000000

20 | 0.138611 | 0/40000 | 0.000000 ||558.421023 | 0/5000 | 0.000000 || 538.403466 | 0/5000 | 0.000000

21 | 0.137376 | 0/40000 | 0.000000 ||558.472330 | 0/5000 | 0.000000 || 537.216140 | 0/5000 | 0.000000

22 | 0.138230 | 0/40000 | 0.000000 ||566.701490 | 0/5000 | 0.000000 || 547.050611 | 0/5000 | 0.000000

23 | 0.137933 | 0/40000 | 0.000000 ||558.663970 | 0/5000 | 0.000000 || 537.247102 | 0/5000 | 0.000000

24 | 0.138045 | 0/40000 | 0.000000 ||558.442581 | 0/5000 | 0.000000 || 538.911868 | 0/5000 | 0.000000

25 | 0.138275 | 0/40000 | 0.000000 ||558.387594 | 0/5000 | 0.000000 || 537.279910 | 0/5000 | 0.000000

26 | 0.137767 | 0/40000 | 0.000000 ||558.494194 | 0/5000 | 0.000000 || 538.535464 | 0/5000 | 0.000000

27 | 0.137107 | 0/40000 | 0.000000 ||558.940117 | 0/5000 | 0.000000 || 537.347723 | 0/5000 | 0.000000

28 | 0.137566 | 0/40000 | 0.000000 ||559.417379 | 0/5000 | 0.000000 || 539.296668 | 0/5000 | 0.000000

29 | 0.136969 | 0/40000 | 0.000000 ||558.650226 | 0/5000 | 0.000000 || 540.081237 | 0/5000 | 0.000000

30 | 0.138429 | 0/40000 | 0.000000 ||558.699915 | 0/5000 | 0.000000 || 537.502130 | 0/5000 | 0.000000

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| MNIST | 30 | 30 | 10 | 3 | softmax | log | .3 | 0 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/2 | 40000.000000

5.000000e-05 | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

.. and so on, this combination of cost and transfer function is not working

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| MNIST | 30 | 30 | 10 | 1 | softmax | log | .3 | 0 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/2 | 40000.000000

5.000000e-05 | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

NaN | NaN | NaN/NaN | NaN ||NaN | NaN/NaN | NaN || NaN | NaN/NaN | NaN

.. and so on, this combination of cost and transfer function is not working

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| MNIST | 30 | 30 | 10 | 3 | Sigmoid | quad | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | 0.263531 | 4/40000 | 0.000100 ||587.482754 | 0/5000 | 0.000000 || 567.060137 | 0/5000 | 0.000000

2 | 0.180258 | 0/40000 | 0.000000 ||584.965618 | 0/5000 | 0.000000 || 561.195131 | 0/5000 | 0.000000

3 | 0.168703 | 0/40000 | 0.000000 ||582.810518 | 0/5000 | 0.000000 || 560.102160 | 0/5000 | 0.000000

4 | 0.162665 | 0/40000 | 0.000000 ||584.561680 | 0/5000 | 0.000000 || 563.711869 | 0/5000 | 0.000000

5 | 0.157559 | 0/40000 | 0.000000 ||583.168506 | 0/5000 | 0.000000 || 560.183466 | 0/5000 | 0.000000

6 | 0.154052 | 0/40000 | 0.000000 ||582.835098 | 0/5000 | 0.000000 || 560.764669 | 0/5000 | 0.000000

7 | 0.150848 | 0/40000 | 0.000000 ||583.543059 | 0/5000 | 0.000000 || 560.253556 | 0/5000 | 0.000000

8 | 0.147884 | 0/40000 | 0.000000 ||586.107138 | 0/5000 | 0.000000 || 562.272017 | 0/5000 | 0.000000

9 | 0.145281 | 0/40000 | 0.000000 ||583.333195 | 0/5000 | 0.000000 || 560.217706 | 0/5000 | 0.000000

10 | 0.143297 | 0/40000 | 0.000000 ||585.112370 | 0/5000 | 0.000000 || 564.640001 | 0/5000 | 0.000000

11 | 0.143860 | 0/40000 | 0.000000 ||582.952720 | 0/5000 | 0.000000 || 560.077580 | 0/5000 | 0.000000

12 | 0.141956 | 0/40000 | 0.000000 ||583.799323 | 0/5000 | 0.000000 || 560.307867 | 0/5000 | 0.000000

13 | 0.140685 | 0/40000 | 0.000000 ||588.423922 | 0/5000 | 0.000000 || 563.890536 | 0/5000 | 0.000000

14 | 0.140271 | 0/40000 | 0.000000 ||585.212937 | 0/5000 | 0.000000 || 564.386250 | 0/5000 | 0.000000

15 | 0.140585 | 0/40000 | 0.000000 ||583.587286 | 0/5000 | 0.000000 || 561.847990 | 0/5000 | 0.000000

16 | 0.139674 | 0/40000 | 0.000000 ||582.746157 | 0/5000 | 0.000000 || 560.123265 | 0/5000 | 0.000000

17 | 0.138259 | 0/40000 | 0.000000 ||582.777965 | 0/5000 | 0.000000 || 560.404663 | 0/5000 | 0.000000

18 | 0.138788 | 0/40000 | 0.000000 ||582.907242 | 0/5000 | 0.000000 || 561.036772 | 0/5000 | 0.000000

19 | 0.138222 | 0/40000 | 0.000000 ||583.356424 | 0/5000 | 0.000000 || 560.146686 | 0/5000 | 0.000000

20 | 0.138269 | 0/40000 | 0.000000 ||582.750573 | 0/5000 | 0.000000 || 560.114905 | 0/5000 | 0.000000

21 | 0.138096 | 0/40000 | 0.000000 ||583.061824 | 0/5000 | 0.000000 || 561.489621 | 0/5000 | 0.000000

22 | 0.138521 | 0/40000 | 0.000000 ||585.534063 | 0/5000 | 0.000000 || 561.730620 | 0/5000 | 0.000000

23 | 0.138384 | 0/40000 | 0.000000 ||585.108931 | 0/5000 | 0.000000 || 561.360080 | 0/5000 | 0.000000

24 | 0.137136 | 0/40000 | 0.000000 ||586.565983 | 0/5000 | 0.000000 || 562.510963 | 0/5000 | 0.000000

25 | 0.137067 | 0/40000 | 0.000000 ||582.710566 | 0/5000 | 0.000000 || 560.270695 | 0/5000 | 0.000000

26 | 0.137336 | 0/40000 | 0.000000 ||582.817686 | 0/5000 | 0.000000 || 560.063728 | 0/5000 | 0.000000

27 | 0.136091 | 0/40000 | 0.000000 ||584.416506 | 0/5000 | 0.000000 || 561.147668 | 0/5000 | 0.000000

28 | 0.137955 | 0/40000 | 0.000000 ||583.985598 | 0/5000 | 0.000000 || 560.489137 | 0/5000 | 0.000000

29 | 0.137132 | 0/40000 | 0.000000 ||583.552990 | 0/5000 | 0.000000 || 560.157922 | 0/5000 | 0.000000

30 | 0.136356 | 0/40000 | 0.000000 ||583.538679 | 0/5000 | 0.000000 || 560.284268 | 0/5000 | 0.000000

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| XOR | 20 | [3 2] | 1 | .1 | Sigmoid | cross | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | 8.115113 | 1/2 | 0.500000 ||0.130975 | 0/1 | 0.000000 || 0.233727 | 1/1 | 1.000000

2 | -0.155514 | 1/2 | 0.500000 ||-0.677061 | 0/1 | 0.000000 || -0.600749 | 1/1 | 1.000000

3 | -0.664403 | 1/2 | 0.500000 ||-0.725267 | 0/1 | 0.000000 || -0.655222 | 1/1 | 1.000000

4 | -0.695487 | 1/2 | 0.500000 ||-0.727657 | 0/1 | 0.000000 || -0.659263 | 1/1 | 1.000000

5 | -0.696304 | 1/2 | 0.500000 ||-0.724896 | 0/1 | 0.000000 || -0.662287 | 1/1 | 1.000000

6 | -0.697231 | 1/2 | 0.500000 ||-0.724037 | 0/1 | 0.000000 || -0.663023 | 1/1 | 1.000000

7 | -0.696216 | 1/2 | 0.500000 ||-0.726499 | 0/1 | 0.000000 || -0.660856 | 1/1 | 1.000000

8 | -0.697070 | 1/2 | 0.500000 ||-0.726975 | 0/1 | 0.000000 || -0.660361 | 1/1 | 1.000000

9 | -0.697250 | 1/2 | 0.500000 ||-0.727006 | 0/1 | 0.000000 || -0.660315 | 1/1 | 1.000000

10 | -0.697291 | 1/2 | 0.500000 ||-0.726930 | 0/1 | 0.000000 || -0.660381 | 1/1 | 1.000000

11 | -0.696485 | 1/2 | 0.500000 ||-0.724141 | 0/1 | 0.000000 || -0.663016 | 1/1 | 1.000000

12 | -0.696436 | 1/2 | 0.500000 ||-0.726076 | 0/1 | 0.000000 || -0.661242 | 1/1 | 1.000000

13 | -0.697103 | 1/2 | 0.500000 ||-0.726440 | 0/1 | 0.000000 || -0.660857 | 1/1 | 1.000000

14 | -0.696520 | 1/2 | 0.500000 ||-0.723754 | 0/1 | 0.000000 || -0.663376 | 1/1 | 1.000000

15 | -0.696415 | 1/2 | 0.500000 ||-0.725720 | 0/1 | 0.000000 || -0.661576 | 1/1 | 1.000000

16 | -0.696667 | 1/2 | 0.500000 ||-0.723388 | 0/1 | 0.000000 || -0.663707 | 1/1 | 1.000000

17 | -0.696365 | 1/2 | 0.500000 ||-0.725455 | 0/1 | 0.000000 || -0.661827 | 1/1 | 1.000000

18 | -0.696668 | 1/2 | 0.500000 ||-0.723145 | 0/1 | 0.000000 || -0.663935 | 1/1 | 1.000000

19 | -0.696357 | 1/2 | 0.500000 ||-0.725219 | 0/1 | 0.000000 || -0.662048 | 1/1 | 1.000000

20 | -0.696661 | 1/2 | 0.500000 ||-0.722912 | 0/1 | 0.000000 || -0.664155 | 1/1 | 1.000000

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| XOR | 20 | [3 2] | 1 | .1 | tanh | cross | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

2 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

3 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

4 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

5 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

6 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

7 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

8 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

9 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

10 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

11 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

12 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

13 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

14 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

15 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

16 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

17 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

18 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

19 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

20 | NaN | 0/2 | 0.000000 ||NaN | 0/1 | 0.000000 || NaN | 0/1 | 0.000000

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data set** | **Epochs** | **Hids** | **Batch size** | **eta** | **Trans** | **Cost** | **Mom** | **Reg** |
| XOR | 20 | [3 2] | 1 | .1 | relu | cross | .3 | 5 |

| TRAIN || TEST || VALIDATION

--------------------------------------------------------------------------

Ep | Cost | Corr | Acc || Cost | Corr | Acc || Cost | Corr | Acc

1 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

2 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

3 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

4 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

5 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

6 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

7 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

8 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

9 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

10 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

11 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

12 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

13 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

14 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

15 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

16 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

17 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

18 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

19 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000

20 | NaN | 1/2 | 0.500000 ||-Inf | 0/1 | 0.000000 || NaN | 1/1 | 1.000000