

DD2424: Assignment 3

Sam Shahriari

May 1, 2024

1 Gradient

To check that my analytical calculation of the gradient was correct, I compared it to a numerical estimation for the same points and features. The numerical estimation was calculated by the given function `ComputeGradsNumSlow()` which uses centered difference formula. It was tested on different batch sizes and lambdas and also with and without batch normalization. None of the absolute or relative errors were above 10^{-6} and therefore I draw the conclusion that my implementation is correct. More detailed results of the testing can be found in [Appendix A](#).

2 3-layer network

Two 3-layer networks, with 50 nodes per layer, were trained with cyclical learning rate on training set consisting of 45000 images and validation set of 5000 images. The following hyperparameters were used: $\eta_{min} = 10^{-5}$, $\eta_{max} = 10^{-1}$, $\lambda = 0.005$, $n_s = 2250$ and $n_{batch} = 100$. The weights were He-initialized. After two training cycles the accuracies on the test data for batch normalization was 53.53% and without batch normalization the accuracy was 52.97%. As the difference in accuracy is small for the 3-layer network it does not seem that batch normalization improves the results in a significant way.

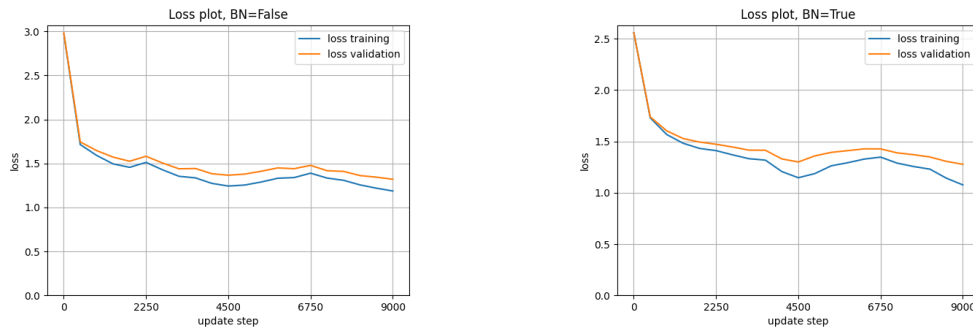


Figure 1: Loss plots for 3-layer networks

3 9-layer network

For the 9-layer networks the following number of hidden nodes were used per layer; 50, 30, 20, 20, 10, 10, 10, 10. The other hyperparameters was the same as for the 3-layer networks. Here it was clear that batch normalization greatly improved the results, 52.27% accuracy compared to 41.75%.

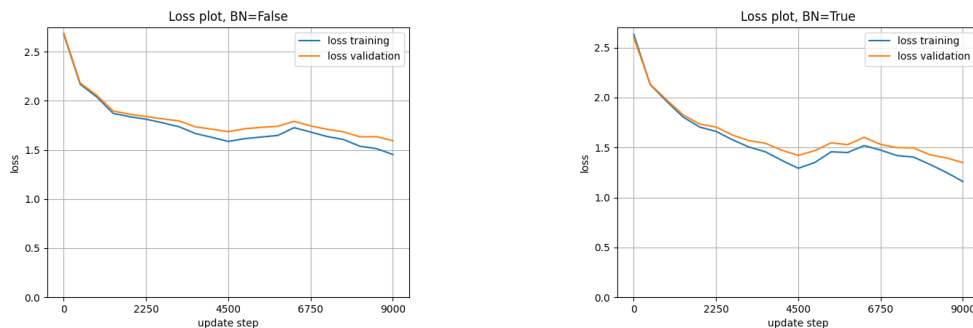


Figure 2: Loss plots for 9-layer networks

4 Search for lambda

The search was a coarse search with a wide possible range of lambda. 8 random lambda values were chosen in the \log_{10} range from -1 to -5 . Other parameter settings were cycles= 2, batch size = 100, $n_s = 2250$, $\eta_{min} = 10^{-5}$, $\eta_{max} = 10^{-1}$. The three best performing lambdas were:

- Accuracy 0.5174 lambda 0.003933267138576786 log lambda -2.4052465562930467

- Accuracy 0.5068 lambda 0.010751078961286456 log lambda -1.968547948453605
- Accuracy 0.5052 lambda 7.889600922457556e-05 log lambda -4.102944964034926

The best lambda was then used for a finer search. Now 20 values were generated in the log range from -2.405 ± 1 . The other hyperparameters were the same as in the coarse search. The three best performing lambdas were:

- Accuracy 0.5236 lambda 0.013089805517608464 log lambda -1.8830668059516829
- Accuracy 0.5234 lambda 0.0087248837545424 log lambda -2.0592403506263293
- Accuracy 0.5186 lambda 0.010736512208547537 log lambda -1.9691367777228215

Lastly, the best lambda = 0.0131 was used to train the model. When testing the model on the test data an accuracy of 53.26% was achieved. A plot of the loss function can be seen in [Figure 3](#) and all the lambdas and their respective accuracy can be seen in [Appendix B](#).

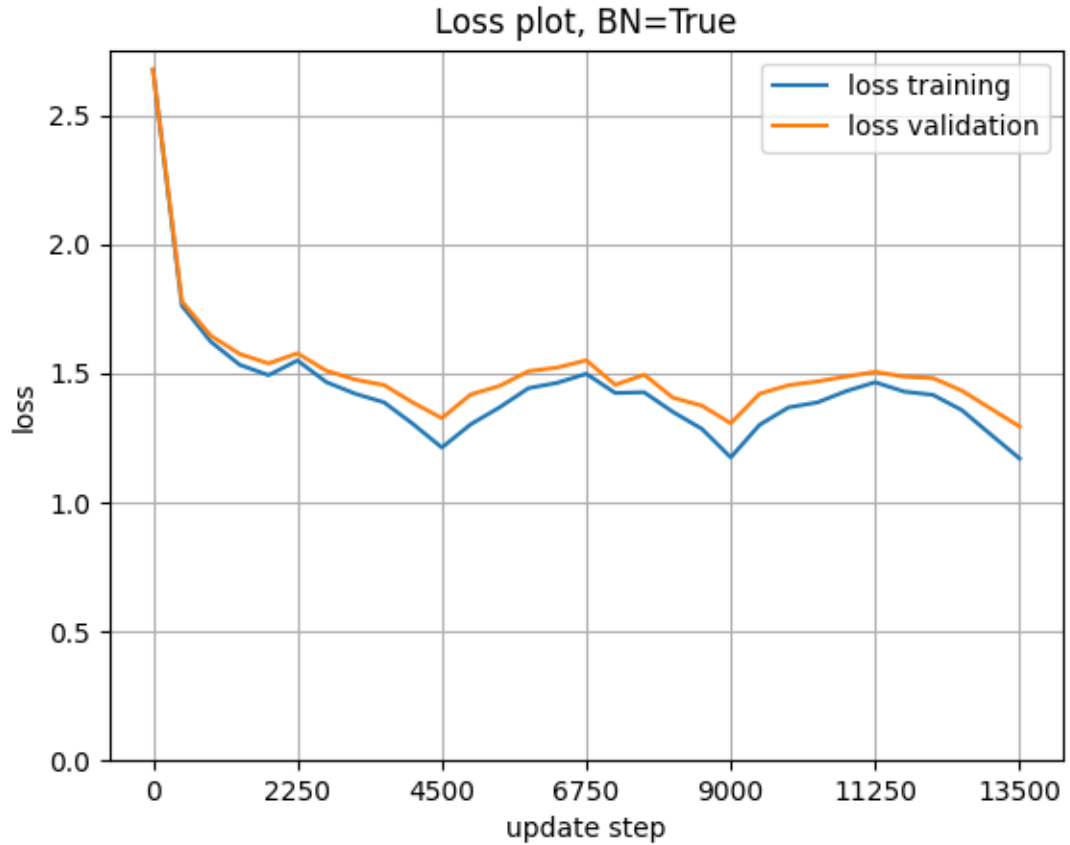


Figure 3: Plot of the loss function for the training and validation set with the best lambda.

5 Sensitivity to initialization

Normally, the weights at each layer are initialized proportional to the number of nodes in that layer. To see how stable batch normalization actually is, it was tested to initialize every layer with the same constant variance. When training the model with batch normalization, the initialization does not really seem to be that important; all accuracies are in the same magnitude. However, the same is not true for the case without batch normalization. Here the accuracy is drastically worsen when the variance gets smaller. When $\sigma = 1e - 4$ the accuracy is just 10% which probably means that the model is trivial and classifies all pictures with the same label.

σ	Without batch normalization	With batch normalization
1e-1	53.02%	53.53%
1e-3	39.32%	53.20%
1e-4	10.00%	53.38%

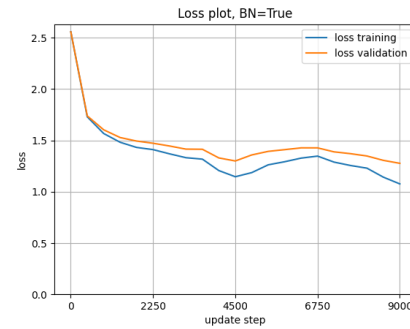
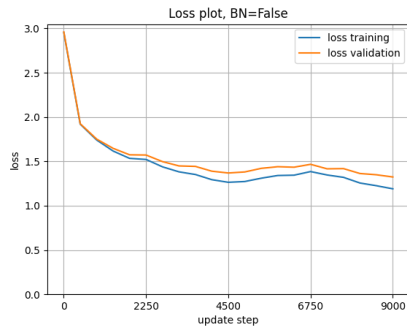


Figure 4: Loss plots for 3-layer networks with $\sigma = 1e - 1$

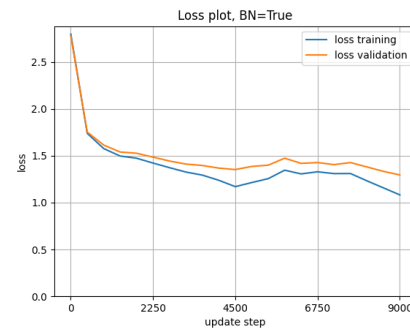
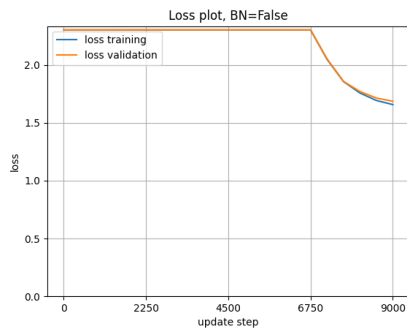


Figure 5: Loss plots for 3-layer networks with $\sigma = 1e - 3$

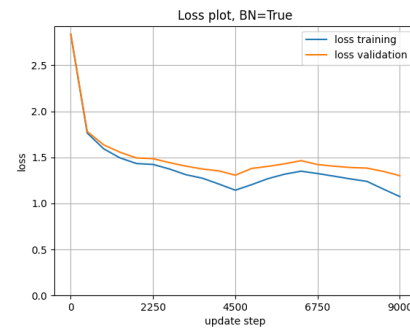
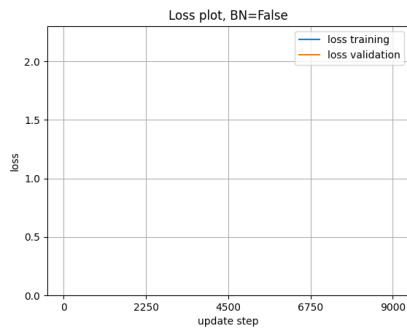


Figure 6: Loss plots for 3-layer networks with $\sigma = 1e - 4$

A Gradients

A.1 Without batch normalization

Hidden layer [10]

layer 0

num features: 10 , batch size: 1 , lambda: 0

SLOW W abs wrong 0 of 100 biggest error 3.816653243449153e-10

SLOW W rel wrong 0 of 100 biggest error 1.2174792905786544e-09

SLOW b abs wrong 0 of 10 biggest error 2.33975727681468e-10

SLOW b rel wrong 0 of 10 biggest error 3.7628044110350863e-10

layer 1

num features: 10 , batch size: 1 , lambda: 0

SLOW W abs wrong 0 of 100 biggest error 2.014396544534891e-10

SLOW W rel wrong 0 of 100 biggest error 1.9417829230840295e-08

SLOW b abs wrong 0 of 10 biggest error 1.543821737115536e-10

SLOW b rel wrong 0 of 10 biggest error 7.49267270517082e-10

layer 0

num features: 10 , batch size: 1 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 5.429762819919581e-10

SLOW W rel wrong 0 of 100 biggest error 1.8324007518343367e-09

SLOW b abs wrong 0 of 10 biggest error 4.560203326064993e-10

SLOW b rel wrong 0 of 10 biggest error 7.333732161266963e-10

layer 1

num features: 10 , batch size: 1 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 4.3690111424521394e-10

SLOW W rel wrong 0 of 100 biggest error 1.4898622561071484e-09

SLOW b abs wrong 0 of 10 biggest error 3.764267786365849e-10

SLOW b rel wrong 0 of 10 biggest error 1.8269224885092043e-09

layer 0

num features: 10 , batch size: 1 , lambda: 1

SLOW W abs wrong 0 of 100 biggest error 1.4289345262596953e-09

```
SLOW W rel wrong 0 of 100 biggest error 5.0106393544156714e-09
SLOW b abs wrong 0 of 10 biggest error 4.560203326064993e-10
SLOW b rel wrong 0 of 10 biggest error 7.333732161266963e-10
```

layer 1

```
num features: 10 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 of 100 biggest error 1.2536531587142852e-09
SLOW W rel wrong 0 of 100 biggest error 4.275041794399312e-10
SLOW b abs wrong 0 of 10 biggest error 4.0393534350702254e-10
SLOW b rel wrong 0 of 10 biggest error 1.9604305666816544e-09
```

layer 0

```
num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 100 biggest error 3.1322301002889574e-10
SLOW W rel wrong 0 of 100 biggest error 1.002230345908774e-09
SLOW b abs wrong 0 of 10 biggest error 3.183766167369484e-10
SLOW b rel wrong 0 of 10 biggest error 4.281986952057651e-09
```

layer 1

```
num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 100 biggest error 3.0777190952041567e-10
SLOW W rel wrong 0 of 100 biggest error 4.825364977997872e-08
SLOW b abs wrong 0 of 10 biggest error 2.9009543725533327e-10
SLOW b rel wrong 0 of 10 biggest error 1.4986104300467567e-09
```

layer 0

```
num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 100 biggest error 4.817825652314589e-10
SLOW W rel wrong 0 of 100 biggest error 1.5231167548308392e-09
SLOW b abs wrong 0 of 10 biggest error 3.8788997014371773e-10
SLOW b rel wrong 0 of 10 biggest error 5.216902572847383e-09
```

layer 1

```
num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 100 biggest error 4.915475526612312e-10
SLOW W rel wrong 0 of 100 biggest error 1.6405285918965569e-09
```

SLOW b abs wrong 0 of 10 biggest error 3.177282603683551e-10
SLOW b rel wrong 0 of 10 biggest error 1.6413594416155358e-09

layer 0

num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 100 biggest error 1.269836347184139e-09
SLOW W rel wrong 0 of 100 biggest error 3.2795591013262285e-09
SLOW b abs wrong 0 of 10 biggest error 9.82848309682982e-10
SLOW b rel wrong 0 of 10 biggest error 1.3218758591783335e-08

layer 1

num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 100 biggest error 1.427862800218449e-09
SLOW W rel wrong 0 of 100 biggest error 4.858541029455862e-10
SLOW b abs wrong 0 of 10 biggest error 7.618174702184177e-10
SLOW b rel wrong 0 of 10 biggest error 3.935489704570203e-09

layer 0

num features: 10 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 of 100 biggest error 4.375390293132053e-10
SLOW W rel wrong 0 of 100 biggest error 3.3185970317690952e-09
SLOW b abs wrong 0 of 10 biggest error 1.6197404528739412e-10
SLOW b rel wrong 0 of 10 biggest error 3.5635811937903967e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 of 100 biggest error 4.391686376538362e-10
SLOW W rel wrong 0 of 100 biggest error 6.637140870508526e-08
SLOW b abs wrong 0 of 10 biggest error 2.25316550139798e-10
SLOW b rel wrong 0 of 10 biggest error 1.8171392041218887e-09

layer 0

num features: 10 , batch size: 100 , lambda: 0.1
SLOW W abs wrong 0 of 100 biggest error 6.291454093809712e-10
SLOW W rel wrong 0 of 100 biggest error 4.552376734390091e-09
SLOW b abs wrong 0 of 10 biggest error 3.013975041765704e-10

SLOW b rel wrong 0 of 10 biggest error 6.631028297362521e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 7.356595124741183e-10

SLOW W rel wrong 0 of 100 biggest error 2.4532928003810457e-09

SLOW b abs wrong 0 of 10 biggest error 2.8837873061737795e-10

SLOW b rel wrong 0 of 10 biggest error 2.32572483785417e-09

layer 0

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 100 biggest error 1.4534506653784973e-09

SLOW W rel wrong 0 of 100 biggest error 6.309098294643479e-09

SLOW b abs wrong 0 of 10 biggest error 8.666166013970855e-10

SLOW b rel wrong 0 of 10 biggest error 1.9066379539299874e-08

layer 1

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 100 biggest error 1.5136956167671745e-09

SLOW W rel wrong 0 of 100 biggest error 5.15018331125187e-10

SLOW b abs wrong 0 of 10 biggest error 5.722568252597426e-10

SLOW b rel wrong 0 of 10 biggest error 4.615152821044962e-09

Hidden layer [10, 11]

layer 0

num features: 10 , batch size: 1 , lambda: 0

SLOW W abs wrong 0 of 100 biggest error 3.2780280182984067e-10

SLOW W rel wrong 0 of 100 biggest error 7.759598034878195e-10

SLOW b abs wrong 0 of 10 biggest error 8.993437938809024e-11

SLOW b rel wrong 5 of 10 biggest error 0.08993437938809024

layer 1

num features: 10 , batch size: 1 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 3.1623973475699077e-10

SLOW W rel wrong 0 of 110 biggest error 2.308690118505315e-08

SLOW b abs wrong 0 of 11 biggest error 1.1298817784166637e-10

SLOW b rel wrong 0 of 11 biggest error 4.153039989176215e-10

layer 2

num features: 10 , batch size: 1 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 2.2619771312092085e-10

SLOW W rel wrong 0 of 110 biggest error 4.000063594242261e-08

SLOW b abs wrong 0 of 10 biggest error 2.1889029477861754e-10

SLOW b rel wrong 0 of 10 biggest error 1.078791301664163e-09

layer 0

num features: 10 , batch size: 1 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 8.603201762102941e-10

SLOW W rel wrong 0 of 100 biggest error 2.0359059260943626e-09

SLOW b abs wrong 0 of 10 biggest error 2.372824367546933e-10

SLOW b rel wrong 5 of 10 biggest error 0.23728243675469327

layer 1

num features: 10 , batch size: 1 , lambda: 0.1

SLOW W abs wrong 0 of 110 biggest error 5.999792893410238e-10

SLOW W rel wrong 0 of 110 biggest error 2.045969828159974e-09

SLOW b abs wrong 0 of 11 biggest error 4.799983466252655e-10

SLOW b rel wrong 0 of 11 biggest error 1.764301687444395e-09

layer 2

num features: 10 , batch size: 1 , lambda: 0.1

SLOW W abs wrong 0 of 110 biggest error 6.747589188782488e-10

SLOW W rel wrong 0 of 110 biggest error 1.8769885349659557e-09

SLOW b abs wrong 0 of 10 biggest error 3.137955312260132e-10

SLOW b rel wrong 0 of 10 biggest error 1.5465276335347916e-09

layer 0

num features: 10 , batch size: 1 , lambda: 1

SLOW W abs wrong 0 of 100 biggest error 2.3967262480217855e-09

SLOW W rel wrong 0 of 100 biggest error 5.656620833809046e-09

SLOW b abs wrong 0 of 10 biggest error 7.98244040312035e-10

SLOW b rel wrong 5 of 10 biggest error 0.798244040312035

layer 1

num features: 10 , batch size: 1 , lambda: 1

SLOW W abs wrong 0 of 110 biggest error 1.9051911385320475e-09

SLOW W rel wrong 0 of 110 biggest error 6.496830233249509e-10

SLOW b abs wrong 0 of 11 biggest error 1.3681767663253908e-09

SLOW b rel wrong 0 of 11 biggest error 5.0289268588556515e-09

layer 2

num features: 10 , batch size: 1 , lambda: 1

SLOW W abs wrong 0 of 110 biggest error 1.9167385456597685e-09

SLOW W rel wrong 0 of 110 biggest error 5.342009945935117e-10

SLOW b abs wrong 0 of 10 biggest error 1.1814977746160693e-09

SLOW b rel wrong 0 of 10 biggest error 5.8229604171371905e-09

layer 0

num features: 10 , batch size: 10 , lambda: 0

SLOW W abs wrong 0 of 100 biggest error 3.310144918583502e-10

SLOW W rel wrong 0 of 100 biggest error 4.22288548852521e-09

SLOW b abs wrong 0 of 10 biggest error 1.8541097997204492e-10

SLOW b rel wrong 0 of 10 biggest error 3.376329969373532e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 4.157322964614546e-10

SLOW W rel wrong 0 of 110 biggest error 3.055737692458823e-08

SLOW b abs wrong 0 of 11 biggest error 3.19271765275575e-10

SLOW b rel wrong 0 of 11 biggest error 2.0682343907533025e-09

layer 2

num features: 10 , batch size: 10 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 4.011891999167183e-10

SLOW W rel wrong 0 of 110 biggest error 3.784408644874036e-08

SLOW b abs wrong 0 of 10 biggest error 3.0174389376025346e-10

SLOW b rel wrong 0 of 10 biggest error 1.4411710872890965e-09

layer 0

num features: 10 , batch size: 10 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 6.988984252442876e-10

SLOW W rel wrong 0 of 100 biggest error 7.904839170053428e-09

SLOW b abs wrong 0 of 10 biggest error 3.4571384643911074e-10

SLOW b rel wrong 0 of 10 biggest error 6.295441730234953e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0.1

SLOW W abs wrong 0 of 110 biggest error 6.177657968509109e-10

SLOW W rel wrong 0 of 110 biggest error 2.1614708349375864e-09

SLOW b abs wrong 0 of 11 biggest error 7.062865500784277e-10

SLOW b rel wrong 0 of 11 biggest error 4.575306342350281e-09

layer 2

num features: 10 , batch size: 10 , lambda: 0.1

SLOW W abs wrong 0 of 110 biggest error 7.899906423469716e-10

SLOW W rel wrong 0 of 110 biggest error 2.1818672974413835e-09

SLOW b abs wrong 0 of 10 biggest error 3.0174389376025346e-10

SLOW b rel wrong 0 of 10 biggest error 1.4411710872890965e-09

layer 0

num features: 10 , batch size: 10 , lambda: 1

SLOW W abs wrong 0 of 100 biggest error 3.0005775867453233e-09

SLOW W rel wrong 0 of 100 biggest error 1.1483680696040504e-08

SLOW b abs wrong 0 of 10 biggest error 1.2101389404439722e-09

SLOW b rel wrong 0 of 10 biggest error 2.2036604155497965e-08

layer 1

num features: 10 , batch size: 10 , lambda: 1

SLOW W abs wrong 0 of 110 biggest error 1.7359864656896207e-09

SLOW W rel wrong 0 of 110 biggest error 5.934890812047437e-10

SLOW b abs wrong 0 of 11 biggest error 1.211124048272616e-09

SLOW b rel wrong 0 of 11 biggest error 7.845630840371141e-09

layer 2

num features: 10 , batch size: 10 , lambda: 1

SLOW W abs wrong 0 of 110 biggest error 1.7682271202801303e-09

SLOW W rel wrong 0 of 110 biggest error 4.92456195832206e-10

SLOW b abs wrong 0 of 10 biggest error 1.5149927978796995e-09

SLOW b rel wrong 0 of 10 biggest error 7.2358177345262185e-09

layer 0

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 100 biggest error 3.9827103767708927e-10

SLOW W rel wrong 0 of 100 biggest error 6.123094748571906e-09

SLOW b abs wrong 0 of 10 biggest error 2.336965135296687e-10

SLOW b rel wrong 0 of 10 biggest error 2.9465819822814825e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 3.031853972794516e-10

SLOW W rel wrong 0 of 110 biggest error 5.369569725553318e-08

SLOW b abs wrong 0 of 11 biggest error 3.3958598433248e-10

SLOW b rel wrong 0 of 11 biggest error 3.680960628825551e-09

layer 2

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 3.3310030038077865e-10

SLOW W rel wrong 0 of 110 biggest error 3.766201064474824e-08

SLOW b abs wrong 0 of 10 biggest error 3.51507122020589e-10

SLOW b rel wrong 0 of 10 biggest error 2.892732279466673e-09

layer 0

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 7.853921020484211e-10

SLOW W rel wrong 0 of 100 biggest error 1.0570141511146505e-08

SLOW b abs wrong 0 of 10 biggest error 5.021436200874918e-10

SLOW b rel wrong 0 of 10 biggest error 6.331319715129423e-09

layer 1

```
num features: 10 , batch size: 100 , lambda: 0.1
SLOW W abs wrong 0 of 110 biggest error 6.845263939014856e-10
SLOW W rel wrong 0 of 110 biggest error 2.3043505605761464e-09
SLOW b abs wrong 0 of 11 biggest error 5.616305892575113e-10
SLOW b rel wrong 0 of 11 biggest error 6.087825123480107e-09
```

layer 2

```
num features: 10 , batch size: 100 , lambda: 0.1
SLOW W abs wrong 0 of 110 biggest error 7.43009542780726e-10
SLOW W rel wrong 0 of 110 biggest error 2.0541998375113104e-09
SLOW b abs wrong 0 of 10 biggest error 6.786267520153544e-10
SLOW b rel wrong 0 of 10 biggest error 5.5847673867315154e-09
```

layer 0

```
num features: 10 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 of 100 biggest error 2.9537939733037533e-09
SLOW W rel wrong 0 of 100 biggest error 1.2468487147717476e-08
SLOW b abs wrong 0 of 10 biggest error 1.2975254620739562e-09
SLOW b rel wrong 0 of 10 biggest error 1.6359958008587047e-08
```

layer 1

```
num features: 10 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 of 110 biggest error 2.347205207975378e-09
SLOW W rel wrong 0 of 110 biggest error 7.993745149074607e-10
SLOW b abs wrong 0 of 11 biggest error 1.4780323570953713e-09
SLOW b rel wrong 0 of 11 biggest error 1.602121160946253e-08
```

layer 2

```
num features: 10 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 of 110 biggest error 2.1793760685007157e-09
SLOW W rel wrong 0 of 110 biggest error 6.070245803077635e-10
SLOW b abs wrong 0 of 10 biggest error 9.627925713628915e-10
SLOW b rel wrong 0 of 10 biggest error 7.92331356930237e-09
```

Hidden layer [10, 11, 12]

layer 0

```
num features: 10 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 of 100 biggest error 2.682418011670795e-10
SLOW W rel wrong 0 of 100 biggest error 8.753739537226628e-10
SLOW b abs wrong 0 of 10 biggest error 2.142868660293118e-10
SLOW b rel wrong 0 of 10 biggest error 3.8911679307662153e-10
```

layer 1

```
num features: 10 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 of 110 biggest error 2.0783318908099493e-10
SLOW W rel wrong 0 of 110 biggest error 1.5300561355244167e-08
SLOW b abs wrong 0 of 11 biggest error 2.442720747897198e-10
SLOW b rel wrong 0 of 11 biggest error 9.054202123287524e-10
```

layer 2

```
num features: 10 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 of 132 biggest error 3.9809963442091756e-10
SLOW W rel wrong 0 of 132 biggest error 1.8767724327407137e-08
SLOW b abs wrong 0 of 12 biggest error 1.0997905364185101e-10
SLOW b rel wrong 0 of 12 biggest error 1.4449807418518444e-10
```

layer 3

```
num features: 10 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 of 120 biggest error 2.495510335919285e-10
SLOW W rel wrong 0 of 120 biggest error 6.087307860109151e-08
SLOW b abs wrong 0 of 10 biggest error 1.9330549660967478e-10
SLOW b rel wrong 0 of 10 biggest error 9.601942119111676e-10
```

layer 0

```
num features: 10 , batch size: 1 , lambda: 0.1
SLOW W abs wrong 0 of 100 biggest error 9.496751406101733e-10
SLOW W rel wrong 0 of 100 biggest error 3.2631297924259116e-09
SLOW b abs wrong 0 of 10 biggest error 4.6147617671632446e-10
SLOW b rel wrong 0 of 10 biggest error 8.379801025255265e-10
```

layer 1

```
num features: 10 , batch size: 1 , lambda: 0.1
```

```
SLOW W abs wrong 0 of 110 biggest error 8.489057096161545e-10
SLOW W rel wrong 0 of 110 biggest error 2.8948257042922355e-09
SLOW b abs wrong 0 of 11 biggest error 3.0274316387135514e-10
SLOW b rel wrong 0 of 11 biggest error 1.122149471852018e-09
```

layer 2

```
num features: 10 , batch size: 1 , lambda: 0.1
SLOW W abs wrong 0 of 132 biggest error 8.729134637341929e-10
SLOW W rel wrong 0 of 132 biggest error 2.3521641761056158e-09
SLOW b abs wrong 0 of 12 biggest error 3.5915415086407165e-10
SLOW b rel wrong 0 of 12 biggest error 4.718815212256458e-10
```

layer 3

```
num features: 10 , batch size: 1 , lambda: 0.1
SLOW W abs wrong 0 of 120 biggest error 8.792848948946386e-10
SLOW W rel wrong 0 of 120 biggest error 2.5496202416109578e-09
SLOW b abs wrong 0 of 10 biggest error 5.210006731592998e-10
SLOW b rel wrong 0 of 10 biggest error 2.5879338122470336e-09
```

layer 0

```
num features: 10 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 of 100 biggest error 7.916878458358312e-09
SLOW W rel wrong 0 of 100 biggest error 2.788134050339798e-08
SLOW b abs wrong 0 of 10 biggest error 2.0061591832210013e-09
SLOW b rel wrong 0 of 10 biggest error 3.642921482968489e-09
```

layer 1

```
num features: 10 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 of 110 biggest error 4.507371365036761e-09
SLOW W rel wrong 0 of 110 biggest error 1.5370440248539583e-09
SLOW b abs wrong 0 of 11 biggest error 2.4457894320928375e-09
SLOW b rel wrong 0 of 11 biggest error 9.065576524959831e-09
```

layer 2

```
num features: 10 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 of 132 biggest error 6.21763818209331e-09
```


SLOW W rel wrong 0 of 132 biggest error 1.6754130162391101e-09
SLOW b abs wrong 0 of 12 biggest error 1.4172026885361788e-09
SLOW b rel wrong 0 of 12 biggest error 1.862018743045596e-09

layer 3

num features: 10 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 of 120 biggest error 5.664210767086786e-09
SLOW W rel wrong 0 of 120 biggest error 1.6489976187438782e-09
SLOW b abs wrong 0 of 10 biggest error 2.6559473509379927e-09
SLOW b rel wrong 0 of 10 biggest error 1.3192719908326821e-08

layer 0

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 100 biggest error 3.458920649901387e-10
SLOW W rel wrong 0 of 100 biggest error 3.671014644034426e-09
SLOW b abs wrong 0 of 10 biggest error 3.030363182610074e-10
SLOW b rel wrong 0 of 10 biggest error 2.7185504620220743e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 110 biggest error 3.8540978855626865e-10
SLOW W rel wrong 0 of 110 biggest error 5.8844731867993286e-08
SLOW b abs wrong 0 of 11 biggest error 2.4006523158259796e-10
SLOW b rel wrong 0 of 11 biggest error 1.3532419673769166e-08

layer 2

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 132 biggest error 4.188384517683774e-10
SLOW W rel wrong 0 of 132 biggest error 1.635550249540048e-08
SLOW b abs wrong 0 of 12 biggest error 1.349440836850846e-10
SLOW b rel wrong 0 of 12 biggest error 8.89724517739925e-10

layer 3

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 120 biggest error 4.0651796100361903e-10
SLOW W rel wrong 0 of 120 biggest error 3.8665104648125546e-08

SLOW b abs wrong 0 of 10 biggest error 2.2294203289030534e-10
SLOW b rel wrong 0 of 10 biggest error 1.0975305874103587e-09

layer 0

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 100 biggest error 1.1479813488701751e-09
SLOW W rel wrong 0 of 100 biggest error 1.1019545255702454e-08
SLOW b abs wrong 0 of 10 biggest error 5.250809231860387e-10
SLOW b rel wrong 0 of 10 biggest error 4.710521149801265e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 110 biggest error 8.615314572857358e-10
SLOW W rel wrong 0 of 110 biggest error 3.118736918737021e-09
SLOW b abs wrong 0 of 11 biggest error 6.355045639033774e-10
SLOW b rel wrong 0 of 11 biggest error 3.582324023617402e-08

layer 2

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 132 biggest error 9.234999986951209e-10
SLOW W rel wrong 0 of 132 biggest error 2.4885575478727698e-09
SLOW b abs wrong 0 of 12 biggest error 5.790332935351472e-10
SLOW b rel wrong 0 of 12 biggest error 3.817730305599625e-09

layer 3

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 120 biggest error 8.011273120583917e-10
SLOW W rel wrong 0 of 120 biggest error 2.315480893623513e-09
SLOW b abs wrong 0 of 10 biggest error 6.22528500970887e-10
SLOW b rel wrong 0 of 10 biggest error 3.0646713968309676e-09

layer 0

num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 100 biggest error 7.409272834157932e-09
SLOW W rel wrong 0 of 100 biggest error 3.823822552497625e-08
SLOW b abs wrong 0 of 10 biggest error 2.9000842439314e-09

SLOW b rel wrong 0 of 10 biggest error 2.6016767252471953e-08

layer 1

num features: 10 , batch size: 10 , lambda: 1

SLOW W abs wrong 0 of 110 biggest error 4.7912367406866e-09

SLOW W rel wrong 0 of 110 biggest error 1.6433740109090479e-09

SLOW b abs wrong 0 of 11 biggest error 3.0135774015116468e-09

SLOW b rel wrong 0 of 11 biggest error 1.6987463718839716e-07

layer 2

num features: 10 , batch size: 10 , lambda: 1

SLOW W abs wrong 0 of 132 biggest error 4.2238524899929075e-09

SLOW W rel wrong 0 of 132 biggest error 1.138168644667248e-09

SLOW b abs wrong 0 of 12 biggest error 2.7732621338261154e-09

SLOW b rel wrong 0 of 12 biggest error 1.8284901769706578e-08

layer 3

num features: 10 , batch size: 10 , lambda: 1

SLOW W abs wrong 0 of 120 biggest error 5.598411290108629e-09

SLOW W rel wrong 0 of 120 biggest error 1.6293111644250124e-09

SLOW b abs wrong 0 of 10 biggest error 2.984008318218656e-09

SLOW b rel wrong 0 of 10 biggest error 1.4690098407523464e-08

layer 0

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 100 biggest error 3.350138118307644e-10

SLOW W rel wrong 0 of 100 biggest error 1.16299466221226e-08

SLOW b abs wrong 0 of 10 biggest error 1.530863665000215e-10

SLOW b rel wrong 0 of 10 biggest error 2.1193589420154887e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 110 biggest error 4.1018818333279916e-10

SLOW W rel wrong 0 of 110 biggest error 5.388234439947342e-08

SLOW b abs wrong 0 of 11 biggest error 3.34473814964098e-10

SLOW b rel wrong 0 of 11 biggest error 5.2308814867017325e-09

layer 2

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 132 biggest error 4.1852837303702276e-10

SLOW W rel wrong 0 of 132 biggest error 4.7129708317498824e-08

SLOW b abs wrong 0 of 12 biggest error 2.976638284815647e-10

SLOW b rel wrong 0 of 12 biggest error 3.224857763229554e-09

layer 3

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 120 biggest error 3.263077221350285e-10

SLOW W rel wrong 0 of 120 biggest error 3.2134543193569454e-08

SLOW b abs wrong 0 of 10 biggest error 2.77394662795416e-10

SLOW b rel wrong 0 of 10 biggest error 2.315398022392262e-09

layer 0

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 100 biggest error 9.635492317222605e-10

SLOW W rel wrong 0 of 100 biggest error 2.472645603705556e-08

SLOW b abs wrong 0 of 10 biggest error 3.405430382130703e-10

SLOW b rel wrong 0 of 10 biggest error 4.7145474131943125e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 110 biggest error 7.487631070723921e-10

SLOW W rel wrong 0 of 110 biggest error 2.55449503094132e-09

SLOW b abs wrong 0 of 11 biggest error 3.217632167651363e-10

SLOW b rel wrong 0 of 11 biggest error 5.032098712597309e-09

layer 2

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 132 biggest error 7.476614813373139e-10

SLOW W rel wrong 0 of 132 biggest error 2.0139948361355007e-09

SLOW b abs wrong 0 of 12 biggest error 4.559690056082921e-10

SLOW b rel wrong 0 of 12 biggest error 4.9399189516203845e-09

layer 3

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 120 biggest error 1.0609056549926166e-09

SLOW W rel wrong 0 of 120 biggest error 3.0934433841117715e-09

SLOW b abs wrong 0 of 10 biggest error 4.705545258998356e-10

SLOW b rel wrong 0 of 10 biggest error 3.92769279594882e-09

layer 0

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 100 biggest error 6.916757076813607e-09

SLOW W rel wrong 0 of 100 biggest error 3.265009424644084e-08

SLOW b abs wrong 0 of 10 biggest error 2.428448631422442e-09

SLOW b rel wrong 0 of 10 biggest error 3.3619939122597864e-08

layer 1

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 110 biggest error 6.376669969831994e-09

SLOW W rel wrong 0 of 110 biggest error 2.1745864007876092e-09

SLOW b abs wrong 0 of 11 biggest error 2.5422092660154494e-09

SLOW b rel wrong 0 of 11 biggest error 3.9757956497578755e-08

layer 2

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 132 biggest error 4.188137614491438e-09

SLOW W rel wrong 0 of 132 biggest error 1.1285039177729964e-09

SLOW b abs wrong 0 of 12 biggest error 2.4220533963864588e-09

SLOW b rel wrong 0 of 12 biggest error 2.624026485897707e-08

layer 3

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 120 biggest error 5.657676771519959e-09

SLOW W rel wrong 0 of 120 biggest error 1.6480146505632893e-09

SLOW b abs wrong 0 of 10 biggest error 2.5034408920876494e-09

SLOW b rel wrong 0 of 10 biggest error 2.0896083696428822e-08

A.2 With batch normalization

Hidden layer [50]

layer 0

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 7.128798801270264e-10
SLOW W rel wrong 0 of 500 biggest error 7.832455233088892e-10
SLOW b abs wrong 0 of 50 biggest error 1.0408340855860843e-16
SLOW b rel wrong 0 of 50 biggest error 1.0408340855860843e-07
SLOW gamma abs wrong 0 of 50 biggest error 4.2855394103216704e-10
SLOW gamma rel wrong 0 of 50 biggest error 1.32734681958712e-09
SLOW beta abs wrong 0 of 50 biggest error 6.287028120333105e-10
SLOW beta rel wrong 0 of 50 biggest error 1.788374552680585e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 6.478785286034849e-10
SLOW W rel wrong 0 of 500 biggest error 1.3228975898799096e-09
SLOW b abs wrong 0 of 10 biggest error 4.6427356403810904e-10
SLOW b rel wrong 0 of 10 biggest error 8.959920131500254e-10

layer 0

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 500 biggest error 2.67553070670834e-09
SLOW W rel wrong 0 of 500 biggest error 2.5445243490694722e-09
SLOW b abs wrong 0 of 50 biggest error 1.0408340855860843e-16
SLOW b rel wrong 0 of 50 biggest error 1.0408340855860843e-07
SLOW gamma abs wrong 0 of 50 biggest error 1.0112075898849326e-09
SLOW gamma rel wrong 0 of 50 biggest error 3.1319818810752122e-09
SLOW beta abs wrong 0 of 50 biggest error 1.036731755998943e-09
SLOW beta rel wrong 0 of 50 biggest error 2.9490319669289687e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 500 biggest error 2.0893314855996437e-09
SLOW W rel wrong 0 of 500 biggest error 3.051377433224832e-09

SLOW b abs wrong 0 of 10 biggest error 8.029421016964022e-10
SLOW b rel wrong 0 of 10 biggest error 1.5495814663331095e-09

layer 0

num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 500 biggest error 1.8217281461119228e-08
SLOW W rel wrong 0 of 500 biggest error 3.273283790082922e-09
SLOW b abs wrong 0 of 50 biggest error 1.0408340855860843e-16
SLOW b rel wrong 0 of 50 biggest error 1.0408340855860843e-07
SLOW gamma abs wrong 0 of 50 biggest error 5.452099688385559e-09
SLOW gamma rel wrong 0 of 50 biggest error 1.6886619136019815e-08
SLOW beta abs wrong 0 of 50 biggest error 4.809678592898603e-09
SLOW beta rel wrong 0 of 50 biggest error 1.368135570174083e-08

layer 1

num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 500 biggest error 1.4334049780462266e-08
SLOW W rel wrong 0 of 500 biggest error 5.183162833689644e-09
SLOW b abs wrong 0 of 10 biggest error 5.243834200197028e-09
SLOW b rel wrong 0 of 10 biggest error 1.0119967893054161e-08

layer 0

num features: 10 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 5.967278633028927e-10
SLOW W rel wrong 0 of 500 biggest error 2.044386918206884e-09
SLOW b abs wrong 0 of 50 biggest error 4.163336342344337e-17
SLOW b rel wrong 0 of 50 biggest error 4.163336342344337e-08
SLOW gamma abs wrong 0 of 50 biggest error 4.517665668640758e-10
SLOW gamma rel wrong 0 of 50 biggest error 2.3201787479840212e-09
SLOW beta abs wrong 0 of 50 biggest error 3.598662878107062e-10
SLOW beta rel wrong 0 of 50 biggest error 1.905209899545019e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 5.355572270382414e-10
SLOW W rel wrong 0 of 500 biggest error 2.0829444941859307e-09
SLOW b abs wrong 0 of 10 biggest error 2.9758899511289627e-10

SLOW b rel wrong 0 of 10 biggest error 9.649852483257686e-10

layer 0

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 500 biggest error 2.618846445257006e-09

SLOW W rel wrong 0 of 500 biggest error 4.347678245231651e-09

SLOW b abs wrong 0 of 50 biggest error 4.163336342344337e-17

SLOW b rel wrong 0 of 50 biggest error 4.163336342344337e-08

SLOW gamma abs wrong 0 of 50 biggest error 8.962244158611243e-10

SLOW gamma rel wrong 0 of 50 biggest error 4.602821447234295e-09

SLOW beta abs wrong 0 of 50 biggest error 1.145609825753402e-09

SLOW beta rel wrong 0 of 50 biggest error 6.0651059990079225e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 500 biggest error 2.197769900097457e-09

SLOW W rel wrong 0 of 500 biggest error 4.500732792878929e-09

SLOW b abs wrong 0 of 10 biggest error 6.554732867591895e-10

SLOW b rel wrong 0 of 10 biggest error 2.1254887202877393e-09

layer 0

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 500 biggest error 1.771737417888275e-08

SLOW W rel wrong 0 of 500 biggest error 3.2647998334761153e-09

SLOW b abs wrong 0 of 50 biggest error 4.163336342344337e-17

SLOW b rel wrong 0 of 50 biggest error 4.163336342344337e-08

SLOW gamma abs wrong 0 of 50 biggest error 5.976705352850331e-09

SLOW gamma rel wrong 0 of 50 biggest error 3.0695110616315016e-08

SLOW beta abs wrong 0 of 50 biggest error 6.997275266151792e-09

SLOW beta rel wrong 0 of 50 biggest error 3.704508746294765e-08

layer 1

num features: 10 , batch size: 100 , lambda: 1

SLOW W abs wrong 0 of 500 biggest error 1.3269914111990744e-08

SLOW W rel wrong 0 of 500 biggest error 5.1652019270292135e-09

SLOW b abs wrong 0 of 10 biggest error 6.278208022902909e-09

SLOW b rel wrong 0 of 10 biggest error 2.0358206208947446e-08

Hidden layer [50, 50]

layer 0

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 4.753094575912775e-10
SLOW W rel wrong 0 of 500 biggest error 1.1261661857114025e-09
SLOW b abs wrong 0 of 50 biggest error 2.7755575615628914e-17
SLOW b rel wrong 0 of 50 biggest error 2.775557561562891e-08
SLOW gamma abs wrong 0 of 50 biggest error 4.197433373098747e-10
SLOW gamma rel wrong 0 of 50 biggest error 1.3883407749853116e-09
SLOW beta abs wrong 0 of 50 biggest error 3.34844291183245e-10
SLOW beta rel wrong 0 of 50 biggest error 1.4538433324843058e-09

layer 1

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 2500 biggest error 5.413139103627174e-10
SLOW W rel wrong 0 of 2500 biggest error 1.3621633227484055e-09
SLOW b abs wrong 0 of 50 biggest error 3.469446951953614e-17
SLOW b rel wrong 0 of 50 biggest error 3.469446951953614e-08
SLOW gamma abs wrong 0 of 50 biggest error 3.87784596100893e-10
SLOW gamma rel wrong 0 of 50 biggest error 1.602109045782743e-09
SLOW beta abs wrong 0 of 50 biggest error 3.4373592405594067e-10
SLOW beta rel wrong 0 of 50 biggest error 1.8943056604063247e-09

layer 2

num features: 10 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 4.092417897616496e-10
SLOW W rel wrong 0 of 500 biggest error 1.3735474662896633e-09
SLOW b abs wrong 0 of 10 biggest error 3.549494170806966e-10
SLOW b rel wrong 0 of 10 biggest error 1.2697469001416343e-09

layer 0

num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 500 biggest error 4.241303206212188e-09
SLOW W rel wrong 0 of 500 biggest error 6.221487198592001e-09
SLOW b abs wrong 0 of 50 biggest error 2.7755575615628914e-17
SLOW b rel wrong 0 of 50 biggest error 2.775557561562891e-08

```
SLOW gamma abs wrong 0 of 50 biggest error 1.3986553616673891e-09
SLOW gamma rel wrong 0 of 50 biggest error 4.626184851913746e-09
SLOW beta abs wrong 0 of 50 biggest error 1.415458361631039e-09
SLOW beta rel wrong 0 of 50 biggest error 6.145706394439544e-09
```

layer 1

```
num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 2500 biggest error 4.565241906728801e-09
SLOW W rel wrong 0 of 2500 biggest error 8.807999256896826e-09
SLOW b abs wrong 0 of 50 biggest error 3.469446951953614e-17
SLOW b rel wrong 0 of 50 biggest error 3.469446951953614e-08
SLOW gamma abs wrong 0 of 50 biggest error 1.5746430374885545e-09
SLOW gamma rel wrong 0 of 50 biggest error 6.505544262472107e-09
SLOW beta abs wrong 0 of 50 biggest error 1.644965424291156e-09
SLOW beta rel wrong 0 of 50 biggest error 9.065294303950348e-09
```

layer 2

```
num features: 10 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 of 500 biggest error 3.8990850791120746e-09
SLOW W rel wrong 0 of 500 biggest error 8.033916339962094e-09
SLOW b abs wrong 0 of 10 biggest error 1.2101135232756022e-09
SLOW b rel wrong 0 of 10 biggest error 4.3288925718938155e-09
```

layer 0

```
num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 500 biggest error 2.8414528729925337e-08
SLOW W rel wrong 0 of 500 biggest error 5.7930614465373185e-09
SLOW b abs wrong 0 of 50 biggest error 2.7755575615628914e-17
SLOW b rel wrong 0 of 50 biggest error 2.775557561562891e-08
SLOW gamma abs wrong 0 of 50 biggest error 1.0799786426263225e-08
SLOW gamma rel wrong 0 of 50 biggest error 3.572131472725439e-08
SLOW beta abs wrong 0 of 50 biggest error 1.217669705691482e-08
SLOW beta rel wrong 0 of 50 biggest error 5.286937927273436e-08
```

layer 1

```
num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 2500 biggest error 3.0222963708081974e-08
```

SLOW W rel wrong 0 of 2500 biggest error 1.0864304827685659e-08
SLOW b abs wrong 0 of 50 biggest error 3.469446951953614e-17
SLOW b rel wrong 0 of 50 biggest error 3.469446951953614e-08
SLOW gamma abs wrong 0 of 50 biggest error 1.2143384436891225e-08
SLOW gamma rel wrong 0 of 50 biggest error 5.016967215401986e-08
SLOW beta abs wrong 0 of 50 biggest error 1.2330538187721451e-08
SLOW beta rel wrong 0 of 50 biggest error 6.795277028145527e-08

layer 2

num features: 10 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 of 500 biggest error 2.2976485825054738e-08
SLOW W rel wrong 0 of 500 biggest error 8.76736910750253e-09
SLOW b abs wrong 0 of 10 biggest error 1.1830347451668999e-08
SLOW b rel wrong 0 of 10 biggest error 4.232024700280067e-08

layer 0

num features: 10 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 of 500 biggest error 5.608195200552599e-10
SLOW W rel wrong 0 of 500 biggest error 1.8290982739146668e-09
SLOW b abs wrong 0 of 50 biggest error 2.42861286636753e-17
SLOW b rel wrong 0 of 50 biggest error 2.42861286636753e-08
SLOW gamma abs wrong 0 of 50 biggest error 4.881679435686515e-10
SLOW gamma rel wrong 0 of 50 biggest error 5.66802676559808e-09
SLOW beta abs wrong 0 of 50 biggest error 4.877935919772591e-10
SLOW beta rel wrong 0 of 50 biggest error 6.762891381049424e-09

layer 1

num features: 10 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 of 2500 biggest error 6.355727090456842e-10
SLOW W rel wrong 0 of 2500 biggest error 5.278553023287646e-09
SLOW b abs wrong 0 of 50 biggest error 1.734723475976807e-17
SLOW b rel wrong 0 of 50 biggest error 1.734723475976807e-08
SLOW gamma abs wrong 0 of 50 biggest error 4.608588545507164e-10
SLOW gamma rel wrong 0 of 50 biggest error 3.224326793973108e-09
SLOW beta abs wrong 0 of 50 biggest error 4.925912767961282e-10
SLOW beta rel wrong 0 of 50 biggest error 4.325055565439405e-09

layer 2

num features: 10 , batch size: 100 , lambda: 0

SLOW W abs wrong 0 of 500 biggest error 5.377113294555147e-10

SLOW W rel wrong 0 of 500 biggest error 2.5356421007309303e-09

SLOW b abs wrong 0 of 10 biggest error 3.224330247242335e-10

SLOW b rel wrong 0 of 10 biggest error 3.781982285058193e-09

layer 0

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 500 biggest error 4.334722997212381e-09

SLOW W rel wrong 0 of 500 biggest error 9.012744302359578e-09

SLOW b abs wrong 0 of 50 biggest error 2.42861286636753e-17

SLOW b rel wrong 0 of 50 biggest error 2.42861286636753e-08

SLOW gamma abs wrong 0 of 50 biggest error 1.7799334992019133e-09

SLOW gamma rel wrong 0 of 50 biggest error 2.0666475231269075e-08

SLOW beta abs wrong 0 of 50 biggest error 1.5327666531839146e-09

SLOW beta rel wrong 0 of 50 biggest error 2.1250657160048804e-08

layer 1

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 2500 biggest error 4.414065854149718e-09

SLOW W rel wrong 0 of 2500 biggest error 1.3186445365370889e-08

SLOW b abs wrong 0 of 50 biggest error 1.734723475976807e-17

SLOW b rel wrong 0 of 50 biggest error 1.734723475976807e-08

SLOW gamma abs wrong 0 of 50 biggest error 1.7885624856162075e-09

SLOW gamma rel wrong 0 of 50 biggest error 1.251339729750781e-08

SLOW beta abs wrong 0 of 50 biggest error 1.519548875617005e-09

SLOW beta rel wrong 0 of 50 biggest error 1.3341960426482677e-08

layer 2

num features: 10 , batch size: 100 , lambda: 0.1

SLOW W abs wrong 0 of 500 biggest error 4.503051272142233e-09

SLOW W rel wrong 0 of 500 biggest error 1.1273294207938397e-08

SLOW b abs wrong 0 of 10 biggest error 1.802895846503283e-09

SLOW b rel wrong 0 of 10 biggest error 2.114708987738483e-08

layer 0

```

num features: 10 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 of 500 biggest error 2.6120387602857065e-08
SLOW W rel wrong 0 of 500 biggest error 5.333447326294253e-09
SLOW b abs wrong 0 of 50 biggest error 2.42861286636753e-17
SLOW b rel wrong 0 of 50 biggest error 2.42861286636753e-08
SLOW gamma abs wrong 0 of 50 biggest error 1.1717508685560318e-08
SLOW gamma rel wrong 0 of 50 biggest error 1.36049803619569e-07
SLOW beta abs wrong 0 of 50 biggest error 1.2292009937686643e-08
SLOW beta rel wrong 0 of 50 biggest error 1.7041947543097358e-07

```

layer 1

```

num features: 10 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 of 2500 biggest error 2.7218096276815373e-08
SLOW W rel wrong 0 of 2500 biggest error 9.282741392783898e-09
SLOW b abs wrong 0 of 50 biggest error 1.734723475976807e-17
SLOW b rel wrong 0 of 50 biggest error 1.734723475976807e-08
SLOW gamma abs wrong 0 of 50 biggest error 1.384098639867215e-08
SLOW gamma rel wrong 0 of 50 biggest error 9.683629349763265e-08
SLOW beta abs wrong 0 of 50 biggest error 1.35380022608933e-08
SLOW beta rel wrong 0 of 50 biggest error 1.1886652237173365e-07

```

layer 2

```

num features: 10 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 of 500 biggest error 2.1088178092654175e-08
SLOW W rel wrong 0 of 500 biggest error 8.371959349765881e-09
SLOW b abs wrong 0 of 10 biggest error 1.0448192084080987e-08
SLOW b rel wrong 0 of 10 biggest error 1.225522026060306e-07

```

B Lambda Random Search

```

Accuracy 0.5174 lambda 0.003933267138576786 log lambda -2.4052465562930467
Accuracy 0.5068 lambda 0.010751078961286456 log lambda -1.968547948453605
Accuracy 0.5052 lambda 7.889600922457556e-05 log lambda -4.102944964034926
Accuracy 0.5034 lambda 7.746755362721288e-05 log lambda -4.110880158541829
Accuracy 0.4994 lambda 0.00041827305669248614 log lambda -3.3785401098555243
Accuracy 0.495 lambda 0.00010704442910478117 log lambda -3.970435929696102
Accuracy 0.4944 lambda 2.3982673574185582e-05 log lambda -4.620102403647806
Accuracy 0.4804 lambda 0.05408396000197605 log lambda -1.2669315170656206

```

Accuracy 0.5236 lambda 0.013089805517608464 log lambda -1.8830668059516829
Accuracy 0.5234 lambda 0.0087248837545424 log lambda -2.0592403506263293
Accuracy 0.5186 lambda 0.010736512208547537 log lambda -1.9691367777228215
Accuracy 0.5154 lambda 0.011063501929448962 log lambda -1.9561073840701921
Accuracy 0.5144 lambda 0.0010390316660296834 log lambda -2.983371216473245
Accuracy 0.513 lambda 0.0027148410276266032 log lambda -2.566255596220956
Accuracy 0.513 lambda 0.0025742886492584965 log lambda -2.589342758227356
Accuracy 0.5118 lambda 0.0012564482032855517 log lambda -2.900855410367016
Accuracy 0.5112 lambda 0.0010487565641182546 log lambda -2.979325307925899
Accuracy 0.5112 lambda 0.0007998196188185105 log lambda -3.097007947239104
Accuracy 0.5076 lambda 0.0009621417270758231 log lambda -3.0167609500480745
Accuracy 0.506 lambda 0.01392192959088928 log lambda -1.8563005669813204
Accuracy 0.5034 lambda 0.0009289449917670095 log lambda -3.032010002345819
Accuracy 0.5028 lambda 0.0007618795446937287 log lambda -3.1181136864163497
Accuracy 0.5016 lambda 0.0013880057954555737 log lambda -2.857608720531653
Accuracy 0.4954 lambda 0.0008220475617029411 log lambda -3.0851030544945255
Accuracy 0.4948 lambda 0.036514875640110266 log lambda -1.4375301741107374
Accuracy 0.4946 lambda 0.030438758357759336 log lambda -1.5165730670582203
Accuracy 0.492 lambda 0.035166855600196244 log lambda -1.4538664618375454
Accuracy 0.49 lambda 0.036957420946651554 log lambda -1.4322983433966967

Accuracy on test data: 0.5326