

DD2424: Assignment 1

Sam Shahriari

April 8, 2024

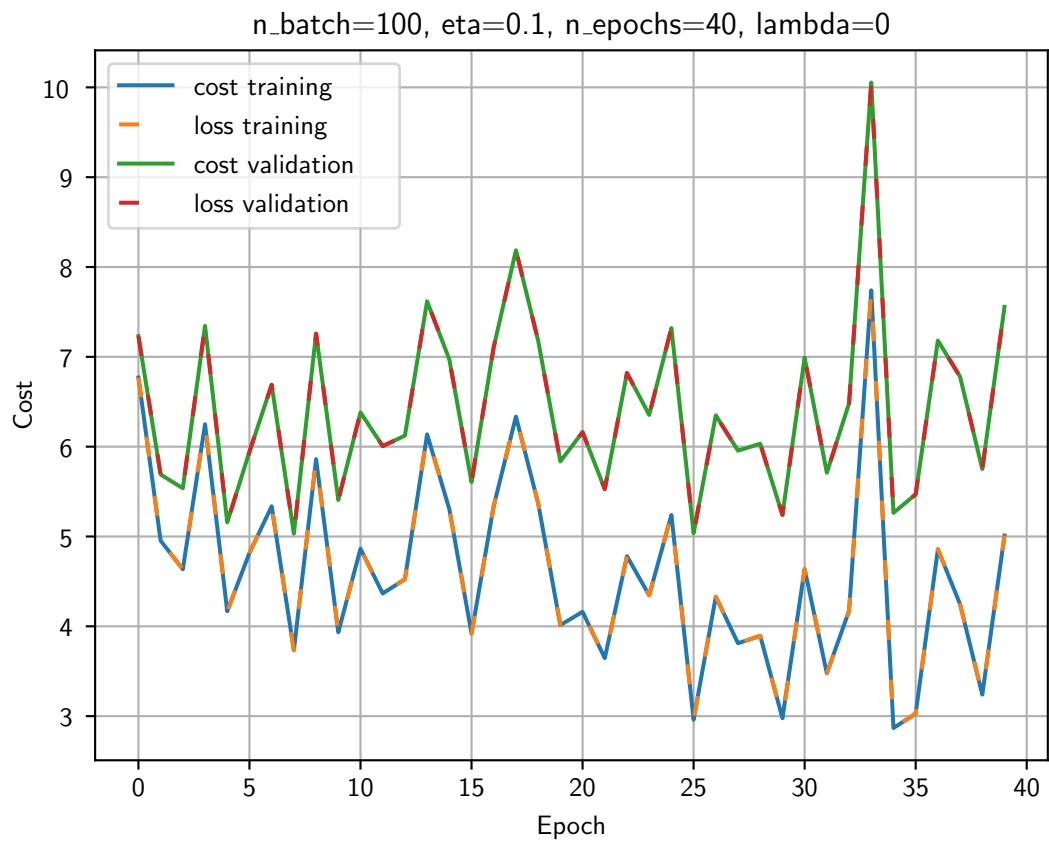
1 Gradients

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. The batches tested were all possible combinations of $\text{featureSize} \in \{20, 500, 3072\}$, $\text{batchSize} \in \{1, 10, 100\}$ and $\lambda \in \{0, .1, 1\}$. None of the absolute 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 Mini-Batch Descent

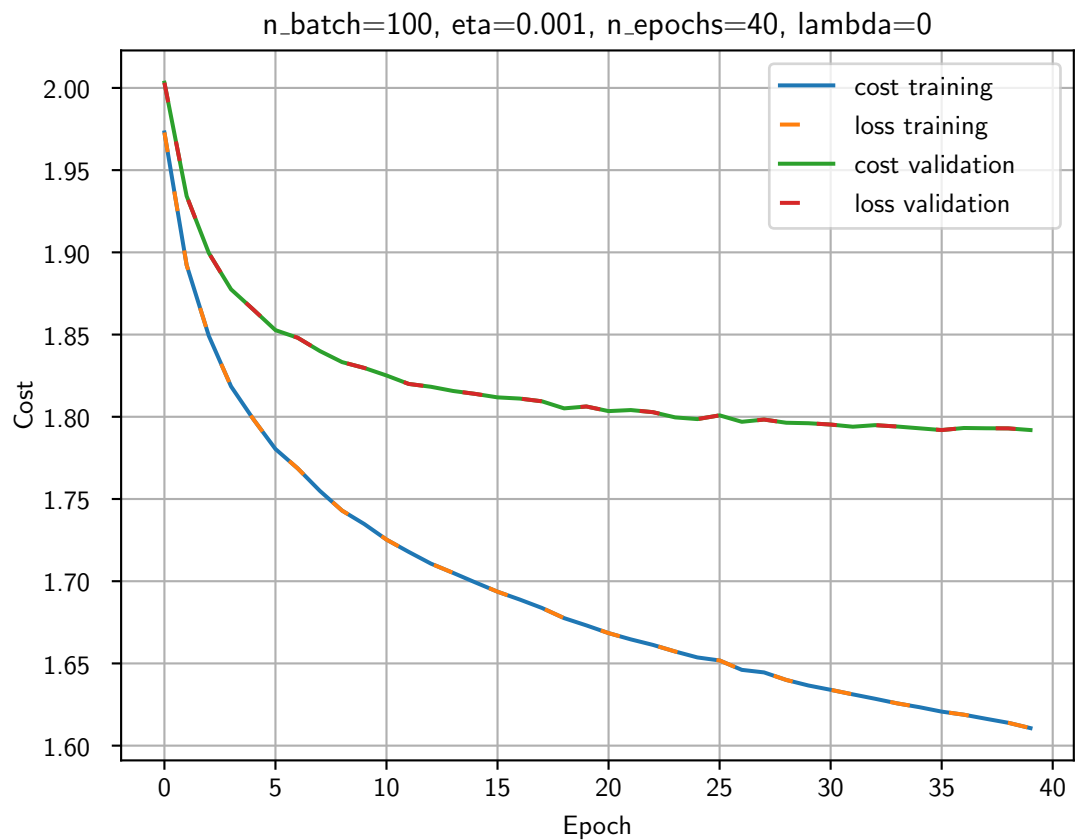
2.1 $\lambda=0$, $n_{\text{epochs}}=40$, $n_{\text{batch}}=100$, $\eta=.1$

Accuracy on training set is 0.4226 and on test 0.2887.



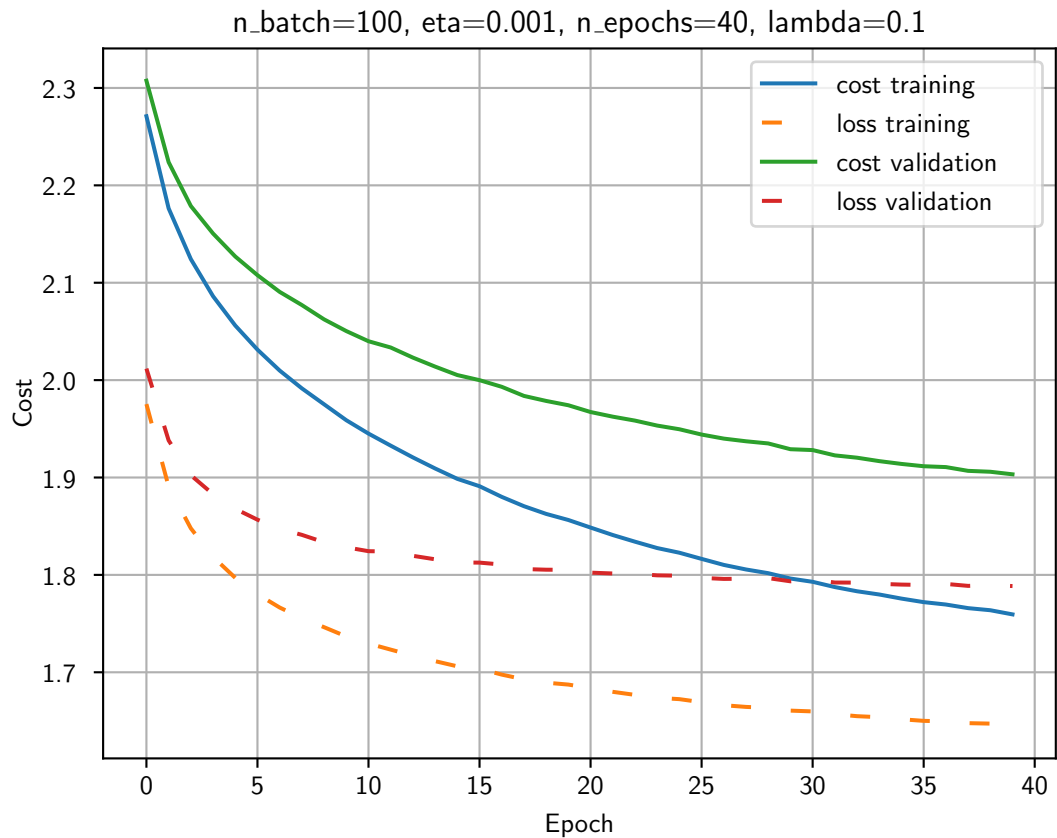
2.2 $\lambda=0$, $n_{\text{epochs}}=40$, $n_{\text{batch}}=100$, $\eta=0.001$

Accuracy on training set is 0.4557 and on test 0.3908.



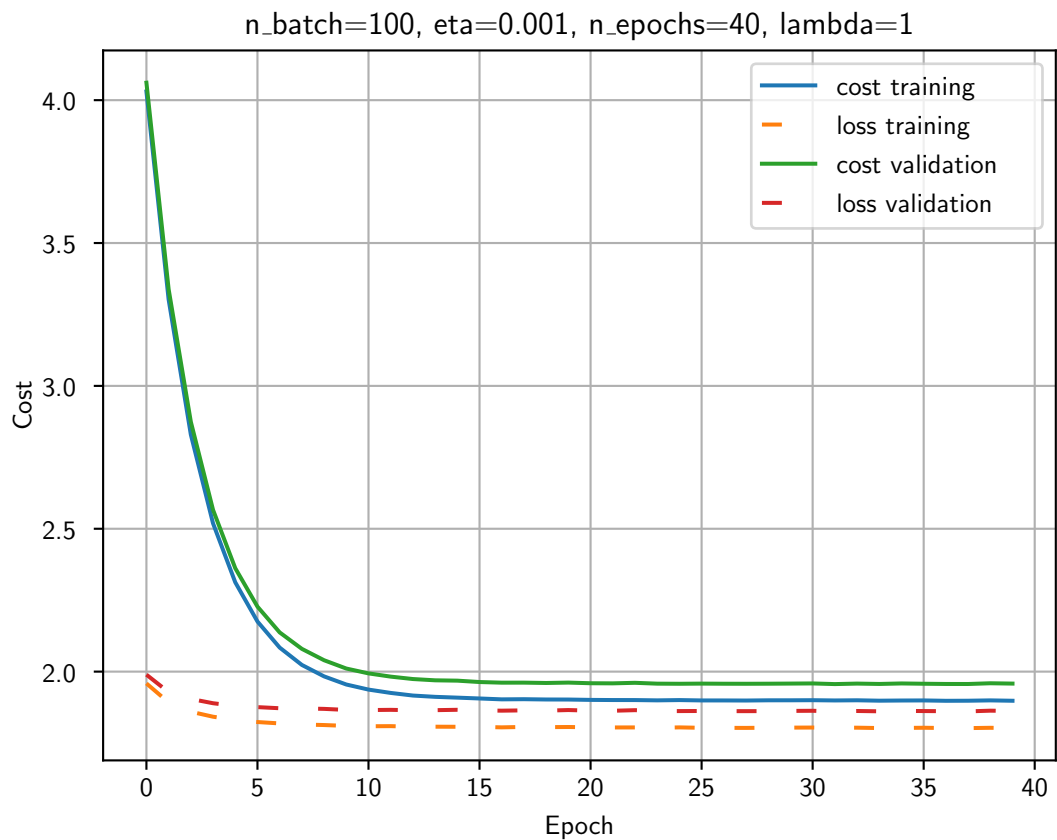
2.3 $\lambda=0.1$, $n_{\text{epochs}}=40$, $n_{\text{batch}}=100$, $\eta=0.001$

Accuracy on training set is 0.4508 and on test 0.3935.



2.4 $\lambda=1$, $n_{\text{epochs}}=40$, $n_{\text{batch}}=100$, $\eta=.001$

Accuracy on training set is 0.4009 and on test 0.3738.



2.5 Commentary

From the results we see that the learning rate choice is very important when it comes to have a decrease in the cost function. When taking a too big of a step, we sometimes overshoot and therefore increase the cost function, leading to a worse predicting model. This almost never happened when taking smaller steps. The test result on the model with .1 learning rate had 10 percentage points lower accuracy compared to the models with .001 of a learning rate.

The amount of regularization is a good example of the bias-variance tradeoff. When we increase λ , we initially see a slight increase in accuracy that is then later decreased to be worse than without regularization when λ is increased more. So without regularization we might overfit the model but with too much regularization we instead underfit the model.

A Gradients

```
-----
num features: 20 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 biggest error 3.2132971283793665e-10
SLOW W rel wrong 0 biggest error 1.472636264539525e-10
SLOW b abs wrong 0 biggest error 1.3739215320995868e-10
SLOW b rel wrong 0 biggest error 6.68000411647355e-10
-----
num features: 20 , batch size: 1 , lambda: 0.1
SLOW W abs wrong 0 biggest error 4.429095978863984e-10
SLOW W rel wrong 0 biggest error 2.0297654510181836e-10
SLOW b abs wrong 0 biggest error 1.3739215320995868e-10
SLOW b rel wrong 0 biggest error 6.68000411647355e-10
-----
num features: 20 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 biggest error 4.3186516168747247e-10
SLOW W rel wrong 0 biggest error 1.978583529399071e-10
SLOW b abs wrong 0 biggest error 1.3739215320995868e-10
SLOW b rel wrong 0 biggest error 6.68000411647355e-10
-----
num features: 20 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 biggest error 4.0456757388618314e-10
SLOW W rel wrong 0 biggest error 1.2490967794502652e-09
SLOW b abs wrong 0 biggest error 3.172324937461557e-10
SLOW b rel wrong 0 biggest error 1.5812284573983417e-09
-----
num features: 20 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 biggest error 5.780703034308221e-10
SLOW W rel wrong 0 biggest error 1.7683872787302242e-09
SLOW b abs wrong 0 biggest error 3.172324937461557e-10
SLOW b rel wrong 0 biggest error 1.5812284573983417e-09
-----
num features: 20 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 biggest error 5.310642064992876e-10
SLOW W rel wrong 0 biggest error 1.1148174877824885e-09
SLOW b abs wrong 0 biggest error 3.172324937461557e-10
SLOW b rel wrong 0 biggest error 1.5812284573983417e-09
-----
num features: 20 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 biggest error 3.853053189006861e-10
```

```
SLOW W rel wrong 0 biggest error 2.6938586167678283e-09
SLOW b abs wrong 0 biggest error 1.9724555322397919e-10
SLOW b rel wrong 0 biggest error 1.6560127689865365e-09
-----
num features: 20 , batch size: 100 , lambda: 0.1
SLOW W abs wrong 0 biggest error 5.556532109429746e-10
SLOW W rel wrong 0 biggest error 3.6952214748672334e-09
SLOW b abs wrong 0 biggest error 1.9724555322397919e-10
SLOW b rel wrong 0 biggest error 1.6560127689865365e-09
-----
num features: 20 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 biggest error 4.680341600393767e-10
SLOW W rel wrong 0 biggest error 2.0753058284123862e-09
SLOW b abs wrong 0 biggest error 1.9724555322397919e-10
SLOW b rel wrong 0 biggest error 1.6560127689865365e-09
-----
num features: 500 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 biggest error 3.207355353529451e-10
SLOW W rel wrong 0 biggest error 9.786973290771547e-11
SLOW b abs wrong 0 biggest error 1.7522611139852984e-10
SLOW b rel wrong 0 biggest error 6.679553910074137e-10
-----
num features: 500 , batch size: 1 , lambda: 0.1
SLOW W abs wrong 0 biggest error 4.2350634587781655e-10
SLOW W rel wrong 0 biggest error 1.290928113474338e-10
SLOW b abs wrong 0 biggest error 1.7522611139852984e-10
SLOW b rel wrong 0 biggest error 6.679553910074137e-10
-----
num features: 500 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 biggest error 4.202084791238292e-10
SLOW W rel wrong 0 biggest error 1.268808174139857e-10
SLOW b abs wrong 0 biggest error 3.9727071632356115e-10
SLOW b rel wrong 0 biggest error 1.5143811304136833e-09
-----
num features: 500 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 biggest error 4.716383117386869e-10
SLOW W rel wrong 0 biggest error 9.544808036654983e-10
SLOW b abs wrong 0 biggest error 2.6485158510780593e-10
SLOW b rel wrong 0 biggest error 1.3149849681780395e-09
-----
num features: 500 , batch size: 10 , lambda: 0.1
```



```
SLOW W abs wrong 0 biggest error 5.561131138920317e-10
SLOW W rel wrong 0 biggest error 1.1225342746745826e-09
SLOW b abs wrong 0 biggest error 2.6485158510780593e-10
SLOW b rel wrong 0 biggest error 1.3149849681780395e-09
-----
num features: 500 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 biggest error 5.350384579827505e-10
SLOW W rel wrong 0 biggest error 1.0196851572952885e-09
SLOW b abs wrong 0 biggest error 4.2174509717884945e-10
SLOW b rel wrong 0 biggest error 2.0939593884900943e-09
-----
num features: 500 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 biggest error 5.018856823191253e-10
SLOW W rel wrong 0 biggest error 2.1716818414356464e-09
SLOW b abs wrong 0 biggest error 2.630494919109161e-10
SLOW b rel wrong 0 biggest error 2.232682680432535e-09
-----
num features: 500 , batch size: 100 , lambda: 0.1
SLOW W abs wrong 0 biggest error 5.857871687720007e-10
SLOW W rel wrong 0 biggest error 2.4814091023576336e-09
SLOW b abs wrong 0 biggest error 2.630494919109161e-10
SLOW b rel wrong 0 biggest error 2.232682680432535e-09
-----
num features: 500 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 biggest error 5.936309430132347e-10
SLOW W rel wrong 0 biggest error 1.9167948282773386e-09
SLOW b abs wrong 0 biggest error 4.850940968359474e-10
SLOW b rel wrong 0 biggest error 4.117328570064186e-09
-----
num features: 3072 , batch size: 1 , lambda: 0
SLOW W abs wrong 0 biggest error 3.66935370976762e-10
SLOW W rel wrong 0 biggest error 1.0323877944563545e-10
SLOW b abs wrong 0 biggest error 1.7566850751826735e-10
SLOW b rel wrong 0 biggest error 4.725696125618173e-10
-----
num features: 3072 , batch size: 1 , lambda: 0.1
SLOW W abs wrong 0 biggest error 5.128785174335349e-10
SLOW W rel wrong 0 biggest error 1.4418298187011428e-10
SLOW b abs wrong 0 biggest error 1.7566850751826735e-10
SLOW b rel wrong 0 biggest error 4.725696125618173e-10
-----
```

```
num features: 3072 , batch size: 1 , lambda: 1
SLOW W abs wrong 0 biggest error 1.0698807100667507e-09
SLOW W rel wrong 0 biggest error 2.9858202116651607e-10
SLOW b abs wrong 0 biggest error 4.844569190254333e-10
SLOW b rel wrong 0 biggest error 1.3032479285049644e-09
-----
num features: 3072 , batch size: 10 , lambda: 0
SLOW W abs wrong 0 biggest error 6.085367348918069e-10
SLOW W rel wrong 0 biggest error 1.096311421126265e-09
SLOW b abs wrong 0 biggest error 2.857255543275983e-10
SLOW b rel wrong 0 biggest error 1.4779599994184556e-09
-----
num features: 3072 , batch size: 10 , lambda: 0.1
SLOW W abs wrong 0 biggest error 7.142557517658688e-10
SLOW W rel wrong 0 biggest error 1.2806952839337178e-09
SLOW b abs wrong 0 biggest error 2.857255543275983e-10
SLOW b rel wrong 0 biggest error 1.4779599994184556e-09
-----
num features: 3072 , batch size: 10 , lambda: 1
SLOW W abs wrong 0 biggest error 1.124553210968271e-09
SLOW W rel wrong 0 biggest error 1.8878599940806494e-09
SLOW b abs wrong 0 biggest error 4.3242488062533724e-10
SLOW b rel wrong 0 biggest error 2.2367851479772863e-09
-----
num features: 3072 , batch size: 100 , lambda: 0
SLOW W abs wrong 0 biggest error 5.52689449762056e-10
SLOW W rel wrong 0 biggest error 2.2846031625379247e-09
SLOW b abs wrong 0 biggest error 2.44458724096841e-10
SLOW b rel wrong 0 biggest error 2.3156471891467504e-09
-----
num features: 3072 , batch size: 100 , lambda: 0.1
SLOW W abs wrong 0 biggest error 6.909972663504682e-10
SLOW W rel wrong 0 biggest error 2.8233784610757455e-09
SLOW b abs wrong 0 biggest error 2.44458724096841e-10
SLOW b rel wrong 0 biggest error 2.3156471891467504e-09
-----
num features: 3072 , batch size: 100 , lambda: 1
SLOW W abs wrong 0 biggest error 1.1095955959605774e-09
SLOW W rel wrong 0 biggest error 3.7124017452773073e-09
SLOW b abs wrong 0 biggest error 4.665033290218723e-10
SLOW b rel wrong 0 biggest error 4.4189755410371955e-09
```