Yue Zhou
DD2424 assignment 1
29 March 2020

Report for assignment 1

Yue Zhou

1. CHECK GRADIENTS

After using back propagation to get the gradients for W and b , I check whether it is right through this kind of way:

Got a mini-batch of data for example d×2 X and K×2 Y . Then use the following codes to get the analytical gradient and the numerical gradient:

gnw, gnb=ComputeGradsNum(X=xmini, Y= Ymini, P=P, W=W, b=b, h=0.0001,lamda=0)

gaw, gab=compute_gradient(X=xmini, Y= Ymini, W=W,b=b,lamda=0)

err=np.abs(gaw-gnw)/(np.abs(gaw)+np.abs(gnw))

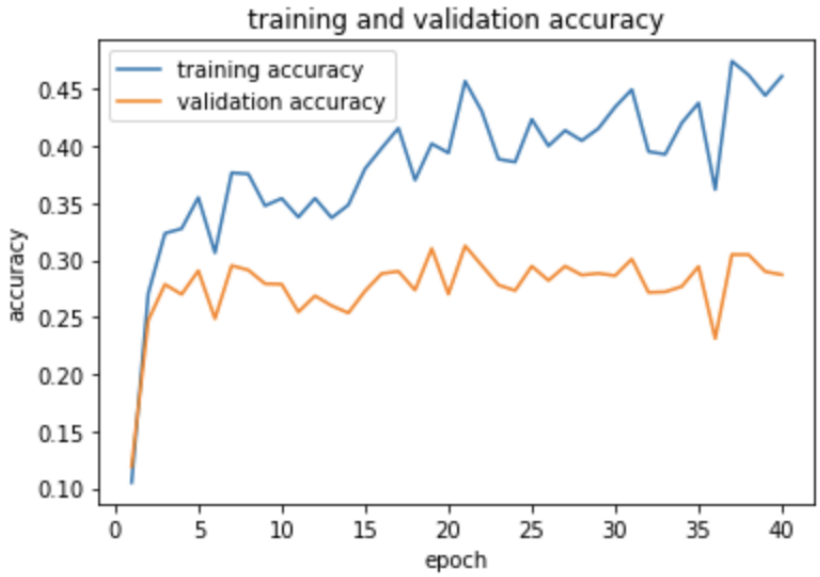
The absolute error between this two is very small. This is the error between grad_w.

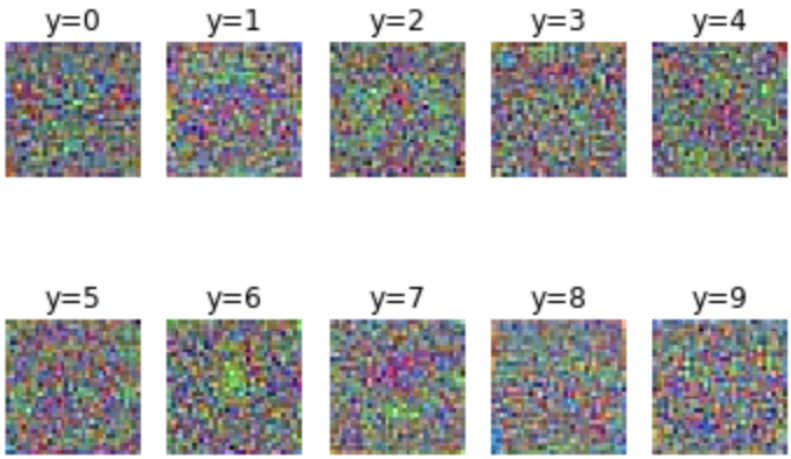
```
[[4.67850713e-05 2.66712263e-05 2.10687190e-05 ... 9.06341705e-06
  7.60801193e-05 6.81903120e-04]
 [2.88651447e-05 2.12013820e-05 1.73986199e-05 ... 7.13827549e-06
  2.31841381e-04 5.50258835e-05]
 [5.97120495e-05 2.52699102e-05 1.92084983e-05 ... 8.45687810e-06
  3.71154499e-05 1.07968922e-04]
 ...
 [3.05238419e-05 2.62694307e-05 2.22844363e-05 ... 8.87111345e-06
  5.01205656e-05 3.79257456e-05]
 [4.60323343e-05 2.56126610e-05 2.01100179e-05 ... 8.60855116e-06
  6.57145867e-05 2.45720249e-03]
 [5.61924607e-06 8.07036856e-05 1.07220832e-05 ... 1.55931767e-06
  1.51003827e-06 2.32427999e-06]]
```

2. RESULTS

2.1 lamda=0, n_epochs=40, n_batch=100, eta=0.1

For training data, the final loss is 3 . 1 7 , the accuracy is 46.13%. For validation data, final the loss is 5.69, the accuracy is 28.72%.

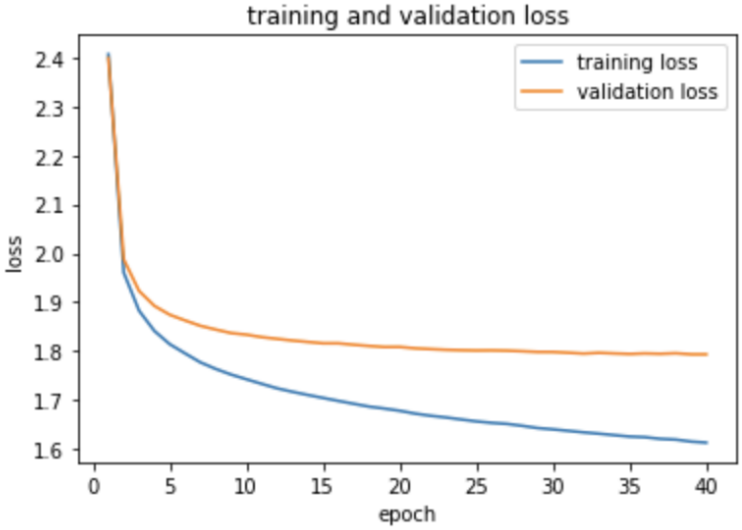


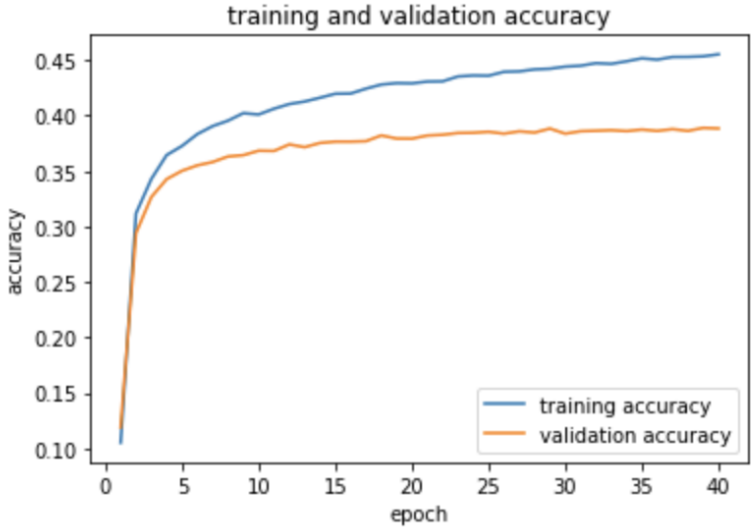


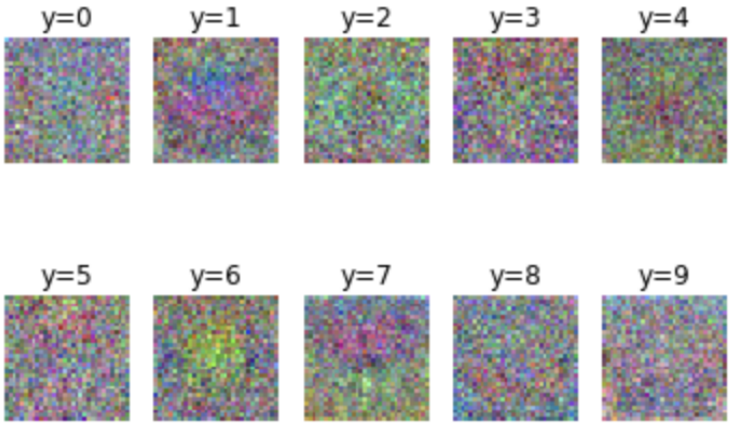
2.2 lamda=0, n_epochs=40, n_batch=100, eta=0.001

For training data, the final loss is 1.61, the accuracy is 45.54%

. For validation data, final the loss is 1.79, the accuracy is 38.85%.



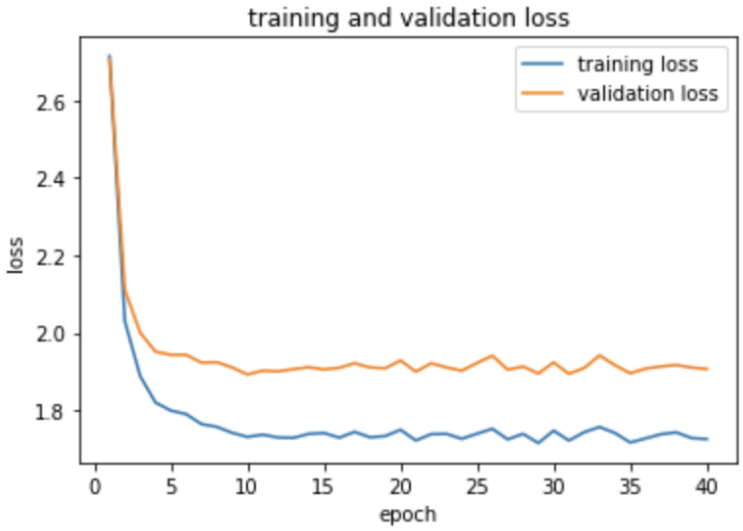


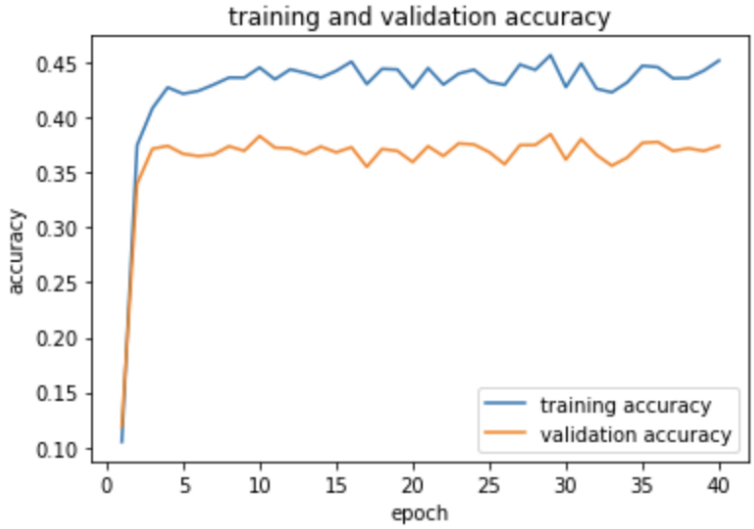


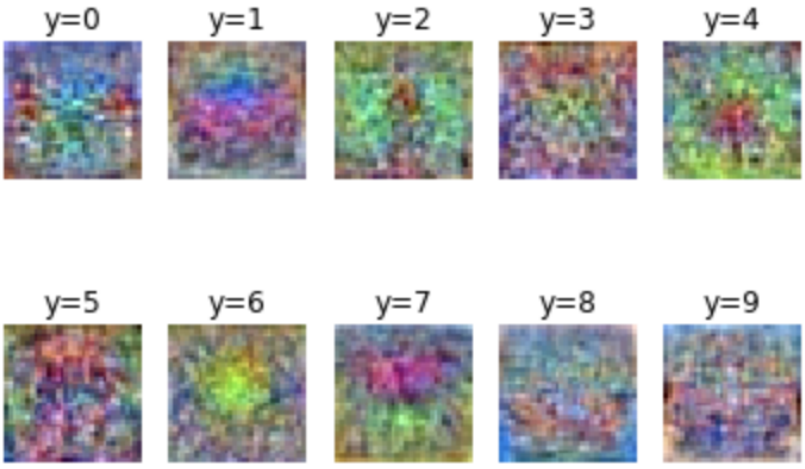
2.3 lamda=0.1, n_epochs=40, n_batch=100, eta=0.001

For training data, the final loss is 1.73, the accuracy is 45.17%

. For validation data, final the loss is 1.91, the accuracy is 37.4%.



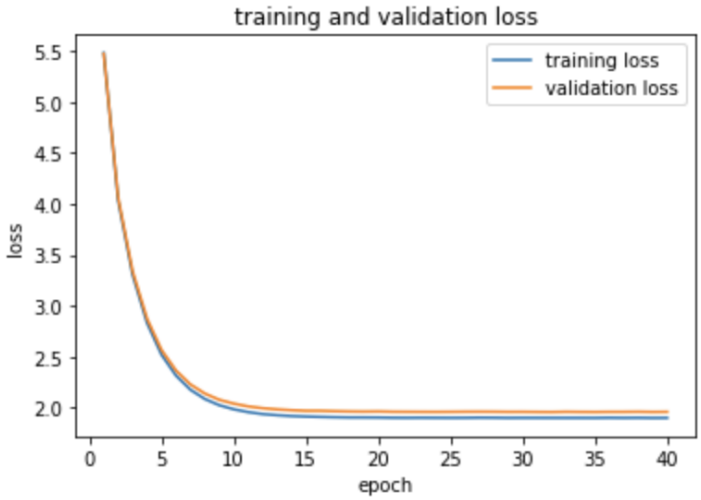


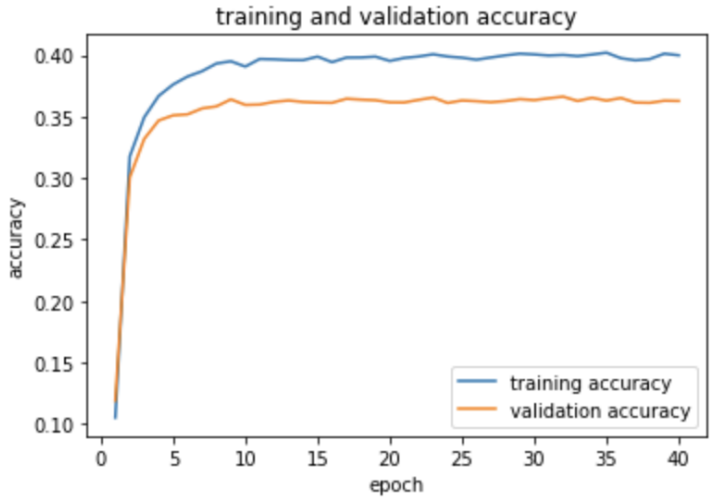


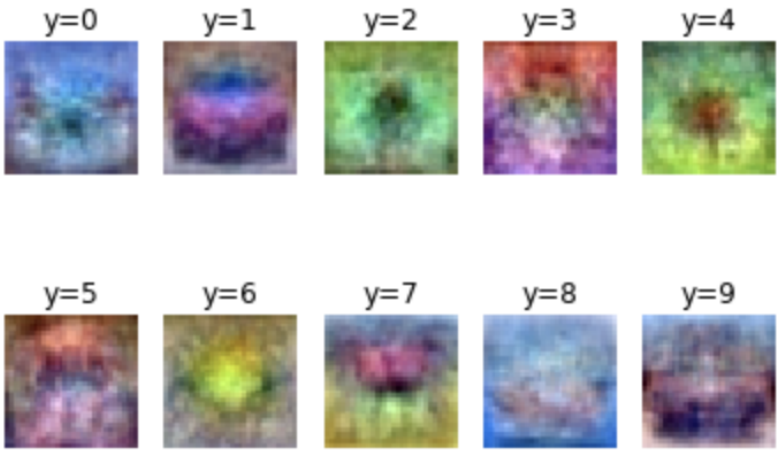
2.4 lamda=1, n_epochs=40, n_batch=100, eta=0.001

For training data, the final loss is 1.90, the accuracy is 40%

. For validation data, final the loss is 1.96, the accuracy is 36.28%







3. Effect of parameters

By increasing regularisation terms, the loss is larger but accuracy increases and validation set is more similar to training set. However, increasing too much may lead to higher loss, like in 2.4. In some cases, regularisation can get rid of overfitting and make the model more general.

We can see from the results that the learning rate is very important in building the model and learning the best w and b. Too large learning rate will lost local minima and the loss will fluctuate a lot like in 2.1. Too small learning rate will slowly converge to local minima but need more time and epochs. So we have to learn the optimum learning rate.