ML Model: Handwritten Digit Classification

Shu Huang

January 24, 2021

Recognition of hand-wrriten digits of 0 and 1 is implemented based on Machine Learing using pytorch.

1 Logistic Regression Model

1.1 Optimizer: SGD, with no momentum added.

NoMomentum Learning rate: 0.0001
Epoch: [1/ 10], Batch ID: [198/ 198], Averaged Loss: 0.1669
Epoch: [2/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0478
Epoch: [3/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0324
Epoch: [4/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0256
Epoch: [5/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0216
Epoch: [6/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0190
Epoch: [7/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0172
Epoch: [8/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0157
Epoch: [9/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0146
Epoch: [10/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0137
Accuracy of the model on the test images: 99.716309 %

1.2 Optimizer: SGD, with 0.9 momentum added.

Momentum: 0.9 Learning rate: 0.0001
Epoch: [1/ 10], Batch ID: [198/ 198], Averaged Loss: 0.1991
Epoch: [2/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0474
Epoch: [3/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0326
Epoch: [4/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0260
Epoch: [5/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0221
Epoch: [6/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0195
Epoch: [7/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0176
Epoch: [8/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0161
Epoch: [9/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0150
Epoch: [10/ 10], Batch ID: [198/ 198], Averaged Loss: 0.0141
Accuracy of the model on the test images: 99.810875 %

1.3 Disscusion

I. For optimizer of SGD without moment (see 1.1) and without momentum (see 1.2). momentum, it can be seen that adding momentum smoothing the variance of the gradients taking by SGD, it also accelerates the convergence.

II. After trying different step sizes (Learning rate), with 0.9 momentum added:

learing rate =0.0001, loss starts from 0.2 to 0.01. Accuracy of the model on the test images: 99.810875 %

learing rate = 0.001 , loss starts from 0.034 to 0.004. Accuracy of the model on the test images: 99.905434 %

learing rate = 0.01, loss starts from 0.015 to 0.0014. Accuracy of the model on the test images: 99.905434 %

learing rate = 0.1, loss starts from 0.054 to 0.000. Accuracy of the model on the test images: 99.905434 %

learing rate = 0.3, , loss is nan. Accuracy of the model on the test images: 46.335697 %

It shows that, for very small learning rate (e.g. 0.0001), 10 epochs is not enough to give a very good convergence, as the learning rate gets lager to 0.001, the convergence improved a little bit. However, too large learning rate leading to loss with value "nan", which shows the learning rate is no longer suitable to do the Logistic Regression.

2 SVM Model

2.1 Optimizer: SGD, with no momentum added.

```
noMomentum Learning rate: 0.0001
Epoch: [ 1/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.5200
Epoch: [ 2/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0670
Epoch: [ 3/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0399
Epoch: [ 4/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0293
Epoch: [ 5/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0239
Epoch: [ 6/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0206
Epoch: [ 7/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0184
Epoch: [ 8/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0169
Epoch: [ 9/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0157
Epoch: [ 10/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0147
Accuracy of the model on the test images: 99.763596%
```

2.2 Optimizer: SGD, with 0.9 momentum added.

```
0.9 Momentum Learning rate: 0.0001
Epoch: [ 1/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0718
Epoch: [ 2/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0129
Epoch: [ 3/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0102
Epoch: [ 4/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0087
Epoch: [ 5/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0077
Epoch: [ 6/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0070
Epoch: [ 7/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0064
Epoch: [ 8/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0060
Epoch: [ 9/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0056
Epoch: [ 10/ 10], Batch ID: [ 198/ 198], Averaged Loss: 0.0053
Accuracy of the model on the test images: 99.858162 %
```

2.3 Disscusion

I. For SVM model, compared with the case without momentum (see 1.2), the tests using optimizer of SGD without moment (see 1.1) reduces variance of gradients taking by SGD, it also accelerates the convergence.

II. According to the outputs, with 0.9 momentum added, at learning rate of

learing rate = 0.0001, loss starts from 0.07 to 0.005. Accuracy of the model on the test images: 99.858162 %

learing rate = 0.001, loss starts from 0.0147 to 0.002. Accuracy of the model on the test images: 99.952721 %

learing rate = 0.01 , loss starts from 0.0145 to 0.0006. Accuracy of the model on the test images: 99.952721 %

learing rate = 0.1 , loss starts from 0.0715 to 0.0019. Accuracy of the model on the test images: 99.905434 %

learing rate = 0.3, loss starts from 0.0246 to 0.0004. Accuracy of the model on the test images: 99.905434 %

learing rate = 10, loss starts from 8.4074 to 0.0341. Accuracy of the model on the test images: 99.905434 %

It can be seen that, for very small learning rate (e.g. 0.0001), 10 epochs is not enough to give a very good convergence, as the learning rate gets lager to 0.001, the convergence improved a little bit. When keeping enlarging the learning rate, the convergence reached a plateau (99.952721 %), and after learning rated reached 0.1 the convergence keeps the same even when learning rate is made to be 10.