Caleb Ralphs
4/25/2019

Homework 6

Hyperparameter Optimization

Constant: num epochs = 20

```
Sigma: 0.001 Batch Size: 50 Beta: 0.0
Loss: 0.08861917639444813
Accuracy: 0.9647
Sigma: 0.001 Batch Size: 50 Beta: 0.1
Loss: 0.05988529105874622
Accuracy: 0.9664
Sigma: 0.001 Batch Size: 50 Beta: 0.5
Loss: 0.07443424515181005
Accuracy: 0.9689
Sigma: 0.001 Batch Size: 100 Beta: 0.0
Loss: 0.040104688966884086
Accuracy: 0.9704
Sigma: 0.001 Batch Size: 100 Beta: 0.1
Loss: 0.03746548340538555
Accuracy: 0.9699
Sigma: 0.001 Batch Size: 100 Beta: 0.5
Loss: 0.09583407126147497
Accuracy: 0.9688
Sigma: 0.01 Batch Size: 50 Beta: 0.0
Loss: nan
Accuracy: 0.098
Sigma: 0.01 Batch Size: 50 Beta: 0.1
Loss: nan
Accuracy: 0.098
Sigma: 0.01 Batch Size: 50 Beta: 0.5
Loss: nan
Accuracy: 0.098
Sigma: 0.01 Batch Size: 100 Beta: 0.0
Loss: nan
Accuracy: 0.098
Sigma: 0.01 Batch Size: 100 Beta: 0.1
Loss: nan
Accuracy: 0.098
Sigma: 0.01 Batch Size: 100 Beta: 0.5
Loss: nan
Accuracy: 0.098
Sigma: 0.1 Batch Size: 50 Beta: 0.0
Loss: nan
Accuracy: 0.098
Sigma: 0.1 Batch Size: 50 Beta: 0.1
Loss: nan
Accuracy: 0.098
Sigma: 0.1 Batch Size: 50 Beta: 0.5
Loss: nan
Accuracy: 0.098
Sigma: 0.1 Batch Size: 100 Beta: 0.0
Loss: nan
Accuracy: 0.098
Sigma: 0.1 Batch Size: 100 Beta: 0.1
Loss: nan
Accuracy: 0.098
Sigma: 0.1 Batch Size: 100 Beta: 0.5
Loss: nan
Accuracy: 0.098
```

Optimal Hyperparameters by min loss (highlighted): sigma = .001, batch size = 100, beta = .1

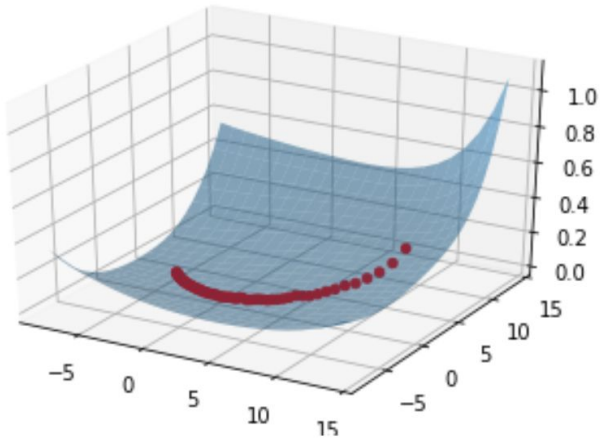
Metrics

Hyperparameters: sigma = .001, batch size = 100, beta = .1, num epochs = 100

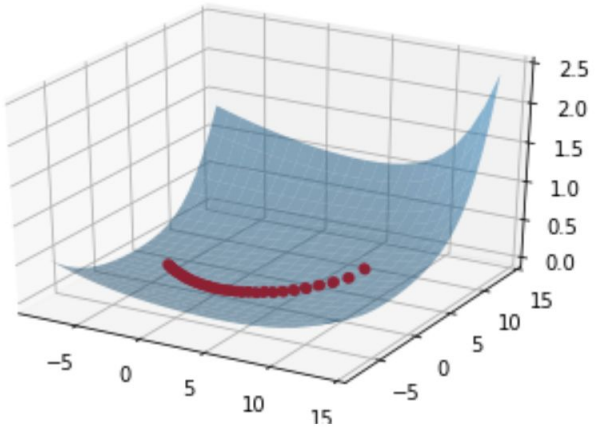
```
Grad Check: 7.11996115361948e-07
Epoch: 80 Cost: 0.030246989698216376
Epoch: 81 Cost: 0.02908002747652973
Epoch: 82 Cost: 0.029034305447621953
Epoch: 83 Cost: 0.02924248478735277
Epoch: 84 Cost: 0.027864703148763287
Epoch: 85 Cost: 0.029086049523045374
Epoch: 86 Cost: 0.02840434396567135
Epoch: 87 Cost: 0.027803311150724784
Epoch: 88 Cost: 0.02761494403378179
Epoch: 89 Cost: 0.028253720712918384
Epoch: 90 Cost: 0.02862816377034994
Epoch: 91 Cost: 0.028090218768880588
Epoch: 92 Cost: 0.02863584687123048
Epoch: 93 Cost: 0.0318853443630585
Epoch: 94 Cost: 0.026378924640707343
Epoch: 95 Cost: 0.02618016622554557
Epoch: 96 Cost: 0.026607692375151624
Epoch: 97 Cost: 0.026600770954973194
Epoch: 98 Cost: 0.02648656596714786
Epoch: 99 Cost: 0.027715661890226468
Loss: 0.027715661890226468
Accuracy: 0.9739
```

Mountains and Valleys

Single loss (not superimposed, epochs = 100)



Two losses (superimposed, epochs = 50)



Even with the superimposition, it still looks like a single gradient descent in respect to the loss. After multiple trials, the weights still were very similar and stacked ontop of one another. I re-initialized the weights before training a second time, but it did not affect the outcome. You can see the code that I have conditioned on the boolean ‘superimpose’ which shows how I changed the weights, and you can see the ‘plotSGDPath()’ function which has the code for superimposition of two different scatter sets.