High Level Computer Vision

Assignment – 2

Submitted By:

Student 1: Hitesh Kotte - 7010571 – hiko00001@stud.uni-saarland.de

Student 2: Rahul Mudambi Venkatesh – 7015710 – ramu00001@stud.uni-saarland.de

Student 3: Anush Onkarappa – 7010620 – anon00001@stud.uni-saarland.de

Question 1: Implementing the feedforward model:

Output:

Your scores:
[[0.3644621 0.22911264 0.40642526]
 [0.47590629 0.17217039 0.35192332]
 [0.43035767 0.26164229 0.30800004]
 [0.41583127 0.2983228 0.28584593]
 [0.36328815 0.32279939 0.31391246]]

correct scores:
[[0.3644621 0.22911264 0.40642526]
 [0.47590629 0.17217039 0.35192332]
 [0.43035767 0.26164229 0.30800004]
 [0.41583127 0.2983228 0.28584593]
 [0.36328815 0.32279939 0.31391246]]

Difference between your scores and correct scores:
2.9173411436600046e-08
Difference between your loss and correct loss:
1.794120407794253e-13

Question 2: Backpropagation:

Output:

W2 max relative error: 3.440708e-09
b2 max relative error: 3.865070e-11
W1 max relative error: 3.561318e-09
b1 max relative error: 1.125423e-09
Final training loss: 0.017149607938732

a) To verify the loss function defined in Eq 12 i.e,

$$J(\theta, \{x_i, y_i\}_{i=1}^{N}) = \frac{1}{N} \sum_{i=1}^{N} -\log\left[\frac{\exp^{z_{y_i}^{(3)}}}{\Sigma_j \exp^{z_j^{(3)}}}\right]$$

Let us derive the above equation wrt z.

wrt, $\frac{\partial J}{\partial z^{(3)}} = \frac{\partial J}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}}$

Let,

$$\frac{\partial J}{\partial a^{(3)}} \cdot \left[\frac{1}{N} \cdot \sum_{i=1}^{N} -\log(a^{(3)})\right]$$

$$= \frac{\partial}{\partial a^{(3)}} \cdot \left[\frac{1}{N} \cdot \sum_{i=1}^{N} \underbrace{-\log(a^{(3)})}_{\text{(Using log to transform)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}}\right.$$

$$= -\frac{1}{N} \cdot \sum_{i=1}^{N} \cdot \frac{1}{(a^{(3)})} \cdot (1) \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}}$$

Let us try to simplify the above equation.

$$= \left(-\frac{1}{N} \cdot \sum_{i=1}^{N} \frac{1}{(a^{(3)})}\right) \cdot \left(\frac{\partial}{\partial z^{(3)}} \underbrace{\psi(z^{(3)})}_{\text{softmax}}\right)$$

From the softmax function we get,

- For $i=j$

$$\psi(z^{(3)}) = S_i(1-S_j)$$

- For $i \neq j$

$$\psi(z^{(3)}) = -S_j S_i$$

$$\therefore \psi(z^{(3)}) = \begin{cases} S_i(1-S_j) & , i=j \\ -S_j S_i & , i \neq j \end{cases}$$

Substituting the values of $\psi(z^{(3)})$ in ①

- For $i=j$

$$= -\frac{1}{N} \cdot \sum_{i=1}^{N} \frac{1}{\cancel{\psi(z_{y_i}^{(3)})}} \cdot \cancel{\psi(z_{y_i}^{(3)})} \left(1 - \psi(z_{y_i}^{(3)})\right)$$

Now, by multiplying by '-'

$$= \frac{1}{N} \left(\psi(z_{y_i}^{(3)}) - 1 \right)$$

- For $i \neq j$

$$= -\frac{1}{N} \cdot \sum_{i=1}^{N} -\frac{1}{\cancel{\psi(z_{y_i}^{(3)})}} \cdot -\psi(z_{y_j}^{(3)}) \cdot \cancel{\psi(z_{y_i}^{(3)})}$$

$$= \frac{1}{N} \cdot \psi(z_{y_i}^{(3)})$$

$\therefore$ The General form is as follows:

$$\frac{\partial J}{\partial z} = \frac{1}{N} \left(\psi(z^{(3)}) - \Delta \right)$$

Δ is either 1 or 0.

$$\Delta_{ij} = \begin{cases} 1 & , y_i = j \\ 0 & , y_i \neq j \end{cases}$$

b) To verify the partial derivative of the loss wrt $w^{(2)}$

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial W^{(2)}}$$

$$= \frac{1}{N}\left(\psi(z^{(3)}) - \Delta\right) a^{(2)'}$$

During Backpropagation we can solve using ①

$$\frac{\partial J}{\partial w^{(2)}} = \frac{\partial J}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial w^{(2)}}$$

$$= \frac{1}{N}\left(\psi(z^{(3)}) - \Delta\right) \cdot \frac{\partial}{\partial w^{(2)}}\left(a^{(2)} \cdot w^{(2)} + b^{(2)}\right)$$

$$= \frac{1}{N}\left(\psi(z^{(3)}) - \Delta\right) \cdot \left(a^{(2)}(1)\right)' \quad \text{—— ②}$$

Keeping in mind the L2 regularization,

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N} -\log\left[\frac{\exp z_{y_i}^{(3)}}{\Sigma_j \exp z_j^{(3)}}\right] + \lambda\left(\|w^{(1)}\|_2^2 + \|w^{(2)}\|_2^2\right)$$

Deriving wrt $w^{(2)}$,

$$\frac{\partial J}{\partial w^{(3)}} = \frac{\partial}{\partial w^{(2)}} \frac{1}{N}\sum_{i=1}^{N} -\log\left[\frac{\exp^{z_{y_i}^{(3)}}}{\Sigma_j \exp^{z_j^{(3)}}}\right] + \frac{\partial}{\partial w^{(2)}} \cdot \lambda\left(\|w^{(1)}\|_2^2 + \|w^{(2)}\|_2^2\right)$$

Using ②

$$= \frac{1}{N}\left(\psi(z^{(3)}) - \Delta\right) a^{(2)'} + 2\lambda w^{(2)}.$$

c) Finally, let us derive the expressions for the derivatives of the regularized loss in Eq. 13 wrt $\omega^{(1)}, b^{(1)}, b^{(2)}$ now.

$$\cdot\ \frac{\partial J}{\partial b^{(2)}} = \frac{\partial J}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial b^{(2)}}$$

$$= \frac{1}{N} \cdot \left(\psi(z^{(3)}) - \Delta \right) \cdot \frac{\partial}{\partial b^{(2)}} \left(a^{(2)} \omega^{(2)} + b^{(2)} \right)$$

$$= \frac{1}{N} \cdot \left(\psi(z^{(3)}) - \Delta \right)$$

$$\cdot\ \frac{\partial J}{\partial \omega^{(1)}} = \frac{\partial J}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial \omega^{(1)}}$$

$$= \frac{1}{N} \left(\psi(z^{(3)}) - \Delta \right) \cdot \frac{\partial}{\partial a^{(2)}} \left(a^{(2)} \omega^{(2)} + b^{(2)} \right)$$

$$= \frac{\partial}{\partial z^{(2)}} \phi(z^{(2)}) \cdot \frac{\partial}{\partial \omega^{(1)}} \cdot \left(a^{(1)} \cdot \omega^{(1)} + b^{(1)} \right) + \frac{\partial}{\partial \omega^{(1)}} \cdot$$

$$\lambda \left(\|\omega^{(1)}\|_2^2 + \|\omega^{(2)}\|_2^2 \right).$$

$$= \frac{1}{N} \left(\psi(z^{(3)}) - \Delta \right) . (\omega^{(2)})(1) . \phi(z^{(2)}) . a^{(1)}(1) + 2\lambda\omega^{(1)}$$

$$\cdot\ \frac{\partial J}{\partial b^{(1)}} = \frac{\partial J}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial a^{(2)}} \cdot \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial b^{(1)}}$$

$$= \frac{1}{N} \left(\psi(z^{(3)}) - \Delta \right) \cdot \frac{\partial}{\partial a^{(2)}} \left(a^{(2)} \cdot \omega^{(2)} + b^{(2)} \right) . \frac{\partial}{\partial z^{(2)}} \phi(z^{(2)}) . 1.$$

$$= \frac{1}{N} \left(\psi(z^{(3)}) - \Delta \right) . (\omega^{(2)}) . \phi(z^{(2)})$$

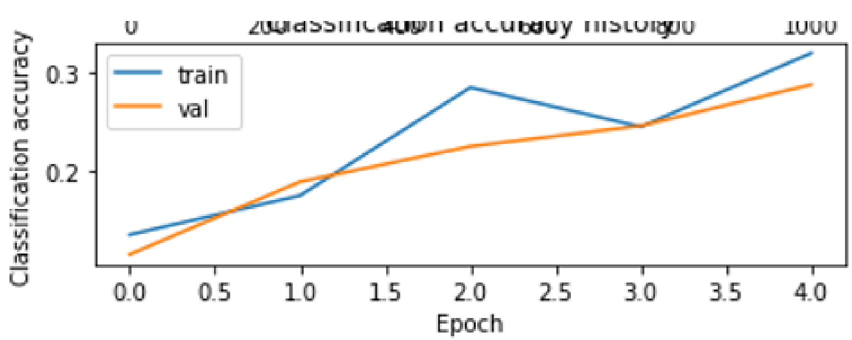
Question 3: Stochastic Gradient Descent Training:

Output:

3a)
Train data shape: (49000, 3072)
Train labels shape: (49000,)
Validation data shape: (1000, 3072)
Validation labels shape: (1000,)
Test data shape: (1000, 3072)
Test labels shape: (1000,)

iteration 0 / 1000: loss 2.302954
iteration 100 / 1000: loss 2.302550
iteration 200 / 1000: loss 2.297648
iteration 300 / 1000: loss 2.259602
iteration 400 / 1000: loss 2.204170
iteration 500 / 1000: loss 2.118565
iteration 600 / 1000: loss 2.051535
iteration 700 / 1000: loss 1.988466
iteration 800 / 1000: loss 2.006591
iteration 900 / 1000: loss 1.951473

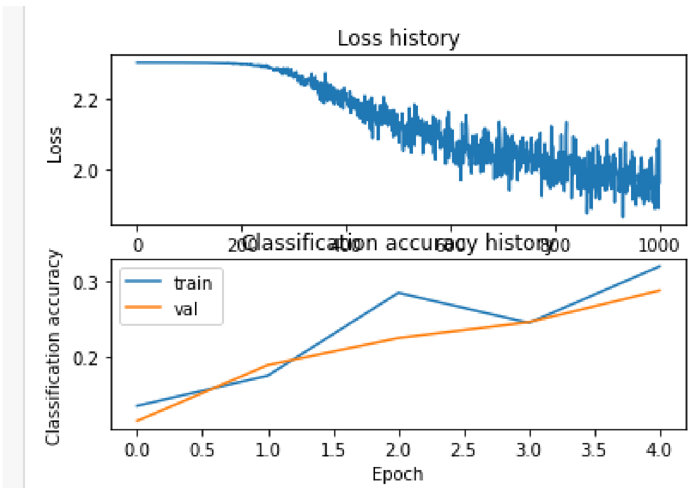
Validation accuracy: 0.287



3b)
Train data shape: (49000, 3072)
Train labels shape: (49000,)
Validation data shape: (1000, 3072)
Validation labels shape: (1000,)
Test data shape: (1000, 3072)
Test labels shape: (1000,)

Best Model
Hidden layer size: 70
Number of iterators: 2000
Batch size: 600
Learning Rate: 0.001
Regularization strength: 0.25
Learning Rate Decay: 0.95
New Validation accuracy: 0.511

Test accuracy: 0.518



Analysis:

Model : 1
Hidden layer size: 50
Number of iterators: 1000
Batch size: 400
Learning Rate: 0.001
Regularization strength: 0.25
Learning Rate Decay: 0.95
New Validation accuracy: 0.472
Test accuracy: 0.476
iteration 0 / 1000: loss 2.303057
iteration 100 / 1000: loss 1.941223
iteration 200 / 1000: loss 1.757708
iteration 300 / 1000: loss 1.718844
iteration 400 / 1000: loss 1.646715
iteration 500 / 1000: loss 1.489401
iteration 600 / 1000: loss 1.529393
iteration 700 / 1000: loss 1.494463

iteration 800 / 1000: loss 1.532311
iteration 900 / 1000: loss 1.517940
.
Model : 2
Hidden layer size: 50
Number of iterators: 1000
Batch size: 400
Learning Rate: 0.001
Regularization strength: 0.3
Learning Rate Decay: 0.95
New Validation accuracy: 0.461
Test accuracy: 0.471
iteration 0 / 1000: loss 2.303119
iteration 100 / 1000: loss 1.973539
iteration 200 / 1000: loss 1.803445
iteration 300 / 1000: loss 1.697027
iteration 400 / 1000: loss 1.689783
iteration 500 / 1000: loss 1.598333
iteration 600 / 1000: loss 1.685545
iteration 700 / 1000: loss 1.535282
iteration 800 / 1000: loss 1.450315
iteration 900 / 1000: loss 1.487978
.
Model : 3
Hidden layer size: 50
Number of iterators: 1000
Batch size: 400
Learning Rate: 0.001
Regularization strength: 0.35
Learning Rate Decay: 0.95
New Validation accuracy: 0.484
Test accuracy: 0.488
iteration 0 / 1000: loss 2.302965
iteration 100 / 1000: loss 1.962506
iteration 200 / 1000: loss 1.759919
iteration 300 / 1000: loss 1.716366
iteration 400 / 1000: loss 1.642749
iteration 500 / 1000: loss 1.588962
iteration 600 / 1000: loss 1.560887
iteration 700 / 1000: loss 1.531736
iteration 800 / 1000: loss 1.517979
iteration 900 / 1000: loss 1.560618
.
Model : 4
Hidden layer size: 50
Number of iterators: 1000
Batch size: 500
Learning Rate: 0.001
Regularization strength: 0.25
Learning Rate Decay: 0.95
New Validation accuracy: 0.478
Test accuracy: 0.467

iteration 0 / 1000: loss 2.303042
iteration 100 / 1000: loss 1.904162
iteration 200 / 1000: loss 1.794330
iteration 300 / 1000: loss 1.755557
iteration 400 / 1000: loss 1.672023
iteration 500 / 1000: loss 1.525590
iteration 600 / 1000: loss 1.496863
iteration 700 / 1000: loss 1.563132
iteration 800 / 1000: loss 1.555040
iteration 900 / 1000: loss 1.417689
.
Model : 5
Hidden layer size: 50
Number of iterators: 1000
Batch size: 500
Learning Rate: 0.001
Regularization strength: 0.3
Learning Rate Decay: 0.95
New Validation accuracy: 0.474
Test accuracy: 0.476
iteration 0 / 1000: loss 2.303102
iteration 100 / 1000: loss 1.913749
iteration 200 / 1000: loss 1.711483
iteration 300 / 1000: loss 1.681609
iteration 400 / 1000: loss 1.660277
iteration 500 / 1000: loss 1.618350
iteration 600 / 1000: loss 1.571224
iteration 700 / 1000: loss 1.562352
iteration 800 / 1000: loss 1.579448
iteration 900 / 1000: loss 1.512034

Question 4: Implement multi-layer perceptron using PyTorch library :

Output:

<u>Model 1</u>: (2-Layer MLP with Hidden Layer of size 50 Neurons)
Loss: 1.2983
Train Accuracy: 51.3%
Validation Accuracy: 50.7%

<u>Model 2</u>: (2-Layer MLP with Hidden Layer of size 60 Neurons)
Loss: 1.298
Train Accuracy: 49.95%
Validation Accuracy:52.8%

Model 3: (3-Layer MLP with Hidden Layer of size 50,60 Neurons)
Loss: 1.285
Train Accuracy: 49.90%
Validation Accuracy: 52.7%

Model 4: (4-Layer MLP with Hidden Layer of size 50,60,50 Neurons)
Loss: 2.406
Train Accuracy: 9.45%
Validation Accuracy: 7.8%

Model 5: (5-Layer MLP with Hidden Layer of size 50,60,60,50 Neurons)
Loss: 2.302
Train Accuracy: 9.15%
Validation Accuracy: 7.8%

Observation: Validation Accuracy is higher for Model 3 and Model 4 with almost similar Train Accuracy. Model 4 and 5 does not fit the model. The optimal model is Model 2 which is the best model and is simple.