DD2424 DEEP LEARNING IN DATA SCIENCE
Assignment 1

Rithika Harish Kumar(rihk@kth.se)

AIM

To train and test a one-layer network with multiple outputs to classify images from the CIFAR-10 dataset. Training is done using mini-batch gradient descent applied to a cost function that computes the cross-entropy loss of the classifier applied to the labelled training data with L2 regularization.

RESULTS

Gradient Computation:

The finite difference method is used to compute gradients numerically.

Without regularization:

```
Gradient of weight
Difference: 3.632e-04
Mean of the values  Analytical: 1.304e-02  Numerical: 1.304e-02
Min of values  Analytical: 2.595e-07  Numerical: 2.465e-07
Max of values  Analytical: 6.044e-02  Numerical: 6.044e-02
Gradient of bias
Difference: 4.476e-07
Mean of the values  Analytical: 2.639e-02  Numerical: 2.639e-02
Min of values  Analytical: 5.874e-05  Numerical: 5.879e-05
Max of values  Analytical: 6.257e-02  Numerical: 6.257e-02
```

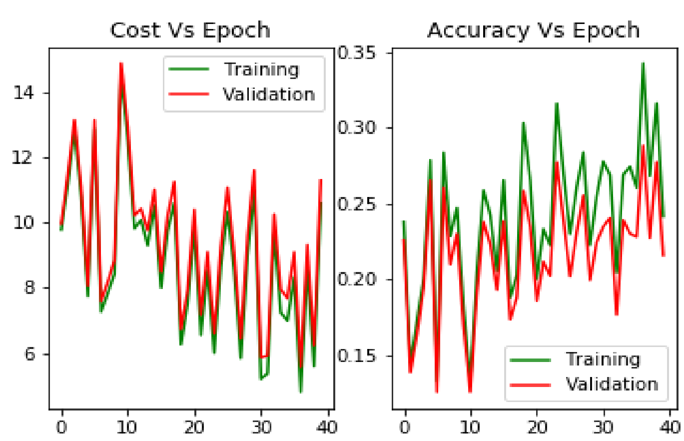
With regularization(rate=1.0):

```
Gradient of weight
Difference: 6.182e-02
Mean of the values  Analytical: 3.613e-02  Numerical: 3.613e-02
Min of values  Analytical: 3.325e-06  Numerical: 1.315e-06
Max of values  Analytical: 1.737e-01  Numerical: 1.737e-01
Gradient of bias
Difference: 4.476e-07
Mean of the values  Analytical: 3.773e-02  Numerical: 3.773e-02
Min of values  Analytical: 3.946e-04  Numerical: 3.946e-04
Max of values  Analytical: 7.878e-02  Numerical: 7.878e-02
```

The learning works although the difference is not very low. So, we can conclude that the analytical gradient computation is correct.

Cases:

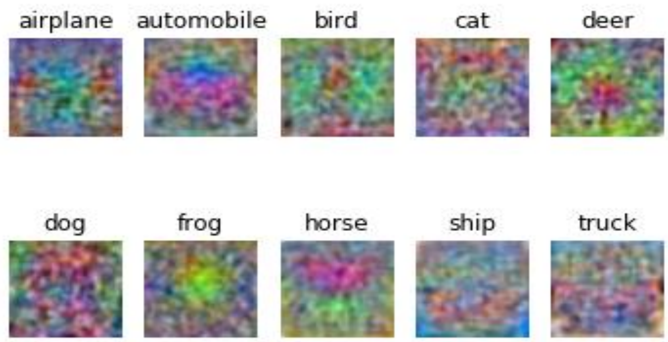
lambda=0, n epochs=40, n batch=100, eta=.1



DD2424 DEEP LEARNING IN DATA SCIENCE
Assignment 1

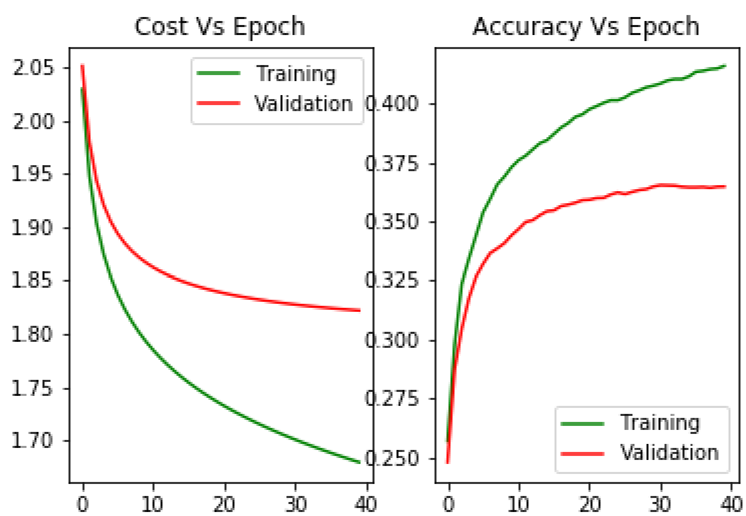
Rithika Harish Kumar(rihk@kth.se)

The learnt weight matrix

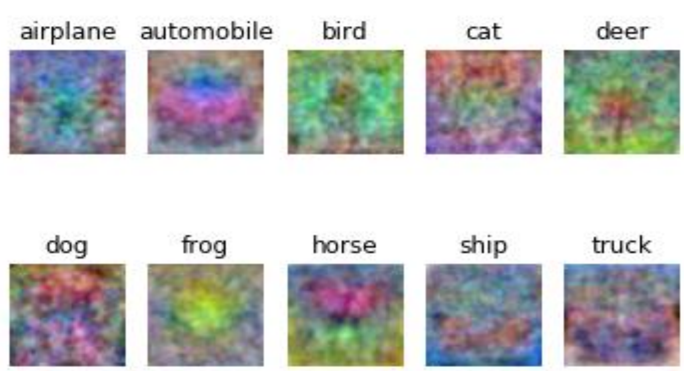


Accuracy for Training: 25.410%   Validation: 21.480%   Testing: 22.51

lambda=0, n epochs=40, n batch=100, eta=.01



Learnt weight matrix

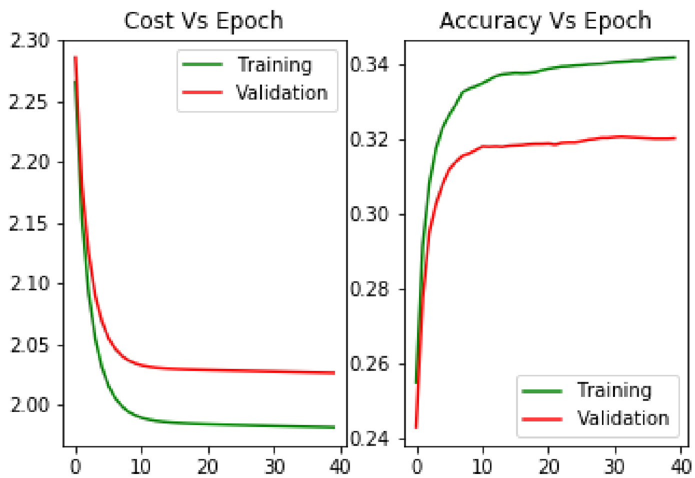


Accuracy for Training: 41.610%   Validation: 36.410% Testing: 36.85% (Max test accuracy)

DD2424 DEEP LEARNING IN DATA SCIENCE
Assignment 1

Rithika Harish Kumar(rihk@kth.se)

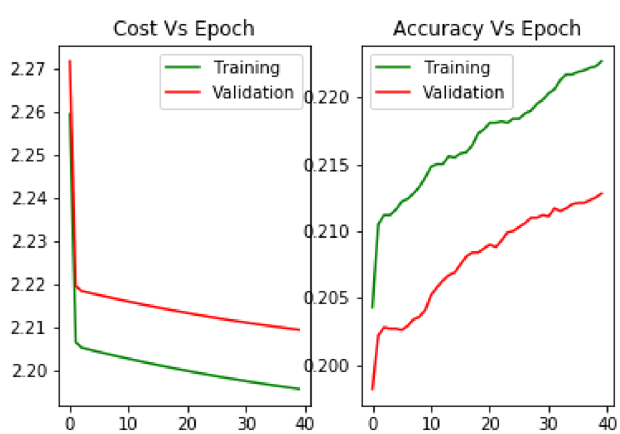
lambda=.1, n epochs=40, n batch=100, eta=.01



Learnt weight matrix

Accuracy for Training: 34.160% Validation: 32.010% Testing: 33.39%

lambda=1, n epochs=40, n batch=100, eta=.01



DD2424 DEEP LEARNING IN DATA SCIENCE
Assignment 1

Rithika Harish Kumar(rihk@kth.se)

Learnt weight matrix

Accuracy for Training: 22.270% and Validation: 21.250% Testing: 21.75%

Regularization:

If regularization is increased, generalization is enhanced and the test accuracy is raised. But if regularization is made high, the network may not be able to represent the data, and the test accuracy may decrease. Without regularization the gap between training and validation accuracy is wider i.e. overfitting. With regularization the gap is narrower, even if the overall accuracy is lower.

Learning rate:

If learning rate is increased, loss function convergences faster, but causes oscillations after some epochs. Learning rate of 0.1 is too high due to the oscillations we could see. We get a gradual and smooth convergence if the learning rate is maintained low.