DD2424 Assignment 01

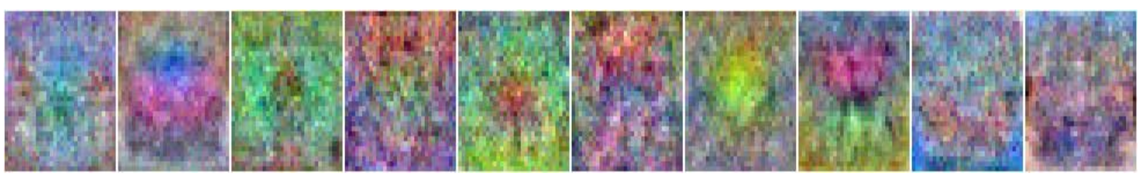
Harsha Holenarasipura Narasanna *harshahn@kth.se*

1. Gradient for the network parameter *W* and *b* were implemented based on the analytical solution. Correctness of the obtained analytical gradients are validated in comparison with numerically computed gradients(fast) measured in terms of relative error. Relative error is averaged and obtained from the *rerr(ga, gn)* function. Results were obtained as follows:
   a. lambda= 0, n = 50 images, *W*: 10x3072, *b*: 10x1; Relative error, b:1.42e-6, W: 2.09e-6;
   b. lambda= 0.1, n = 50 images, *W*: 10x3072, *b*: 10x1; Relative error, b:9.67e-7, W: 1.68e-5;
   c. lambda= 1, n = 50 images, *W*: 10x3072, *b*: 10x1; Relative error, b:1.43e-5, W: 1.28e-4;

Thus, the correctness of the implementation of analytical gradient is valid.
Comment: Even with significant value of hyperparameter *lamda* the relative error remain low.

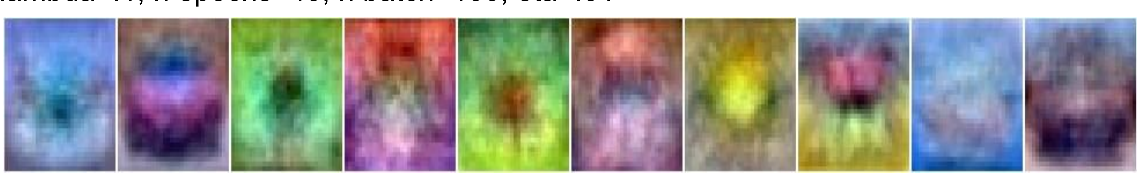
2. Mini-batch gradient descent algorithm. Includes random shuffle amongst 10,000 images.
   1. Class template/ image of weight matrix:
      a. lambda=0, n epochs=40, n batch=100, eta=.1

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

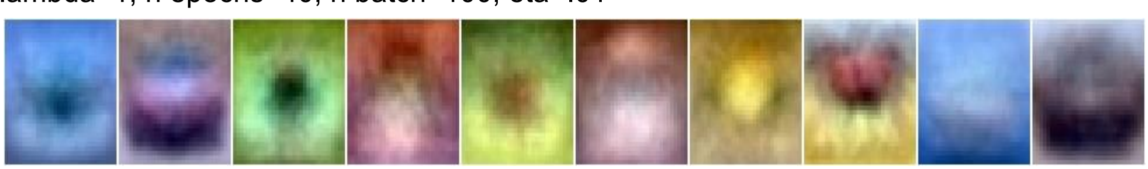


Comment: Visual complexion of class objects is better in the case of low learning rate (eta).

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



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



DD2424 Assignment 01

Harsha Holenarasipura Narasanna *harshahn@kth.se*

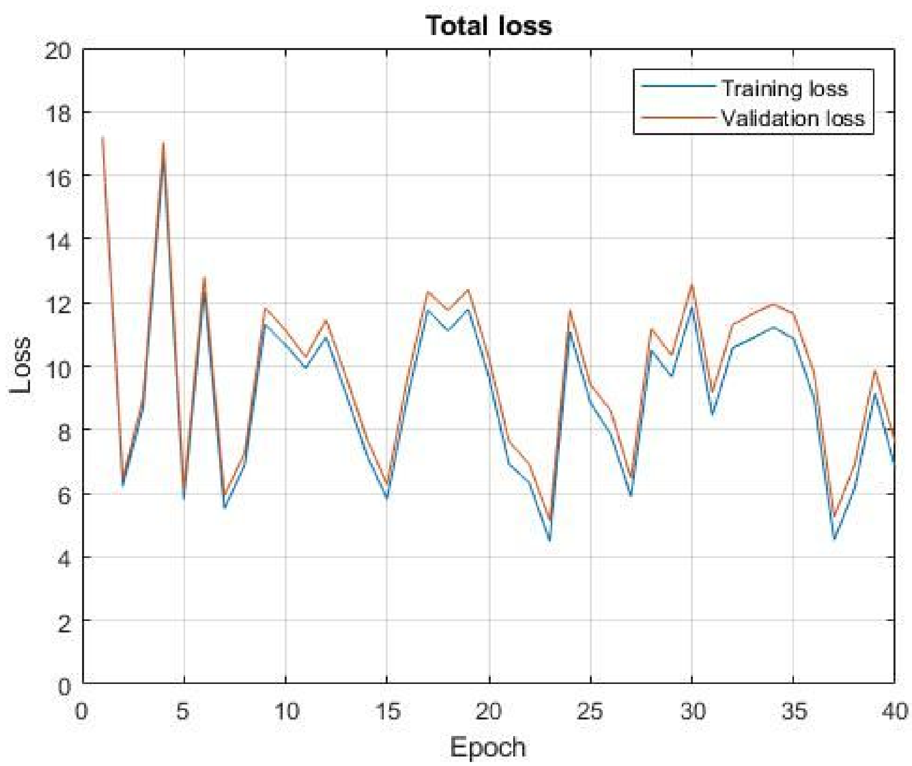
Comment: Complexion of class objects is blurred in the the case of higher lamda which can be attributed to being more generalized.
Overall, the objects resemble close to objects in the case of 3 or c. Possibly due to moderately lower yet optimal values for regularization & learning rate.

Harsha Holenarasipura Narasanna *harshahn@kth.se*

2. Accuracy on the test data:
    a. lambda=0, n epochs=40, n batch=100, eta=.1 Accuracy = 25.01%
    b. lambda=0, n epochs=40, n batch=100, eta=.01 Accuracy = 36.72%
    c. lambda=.1, n epochs=40, n batch=100, eta=.01 Accuracy = 31.60%
    d. lambda=1, n epochs=40, n batch=100, eta=.01 Accuracy = 18.37%

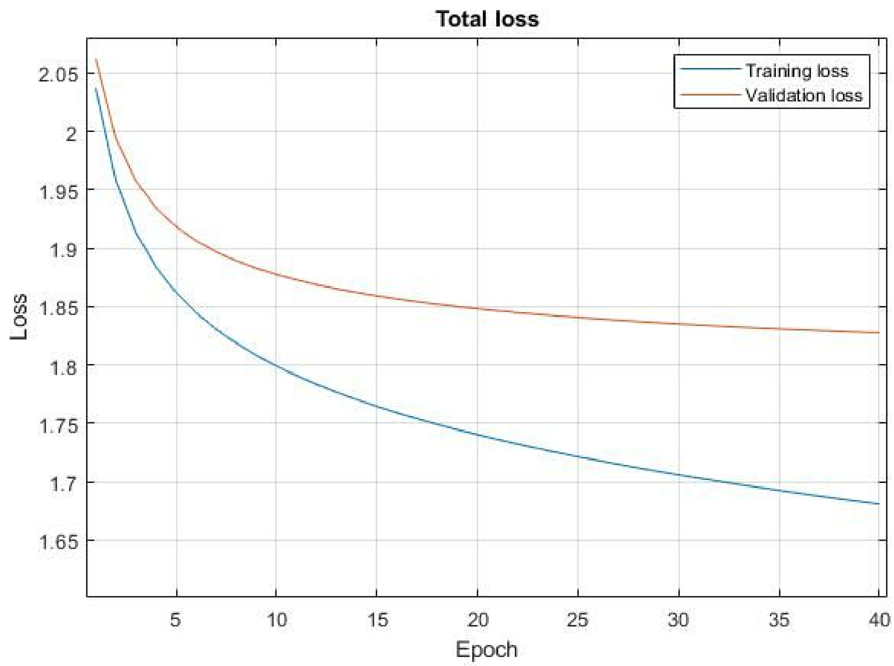
Comment: (b) Moderately lower regularization & learning rate had a moderate boost in accuracy. However, increase in regularization led to exponential decay in accuracy.

3. Graph of the total loss:
    a. Lambda = 0, n epochs = 40, n batch = 100, eta = 0.1

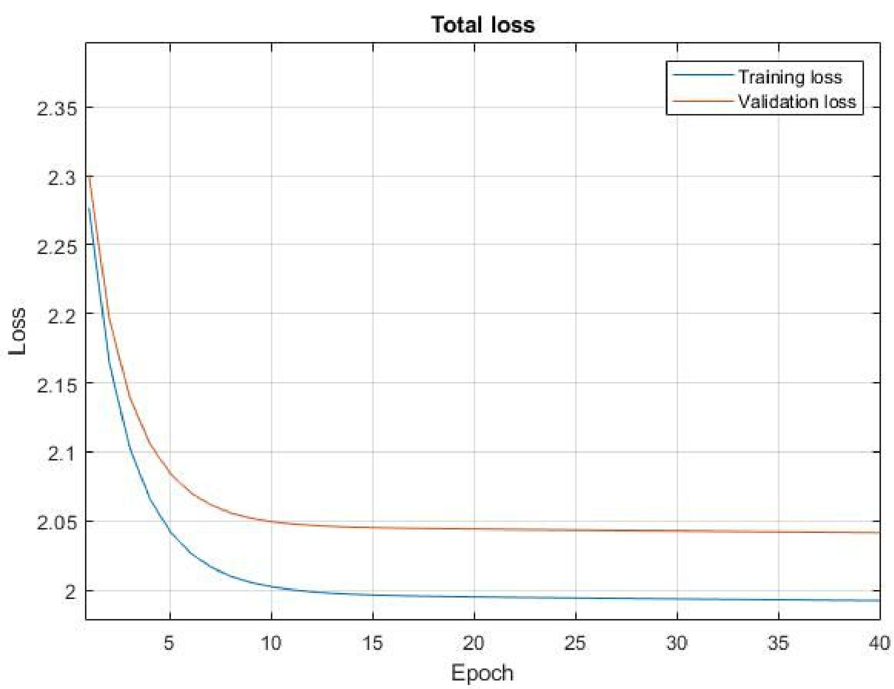


Comment: Consistent fluctuation in the higher values of the loss with no significant reduction over the course of increasing epoch indicates that the learning rate is high and thus, bouncing around the minima but not converging to the minima.

b. lambda=0, n epochs=40, n batch=100, eta=0.01



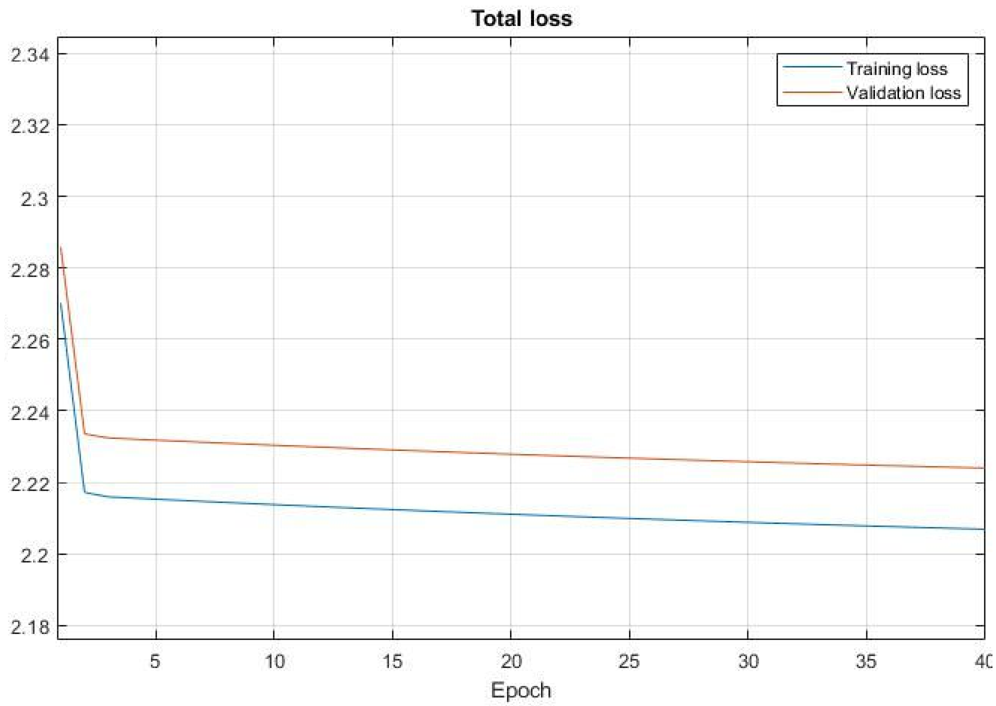
c. lambda=.1, n epochs=40, n batch=100, eta=.01



Harsha Holenarasipura Narasanna *harshahn@kth.se*

Comment: In *b*, Loss has monotonically reduced in an exponential fashion. However, having minimal loss for the training set has resulted in lesser accuracy possibly due to overfitting. With introduction of regularization in *c has* resulted in better performance for the test set. Thus, the overfitting issue is resolved through generalization.

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



Comment: Further increase in the regularization has stagnated the loss at lower values. Also, has performed poorly. Possible reason being that the loss is not optimal which has been ceased by higher regularization.