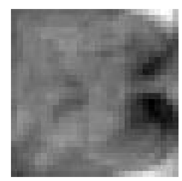
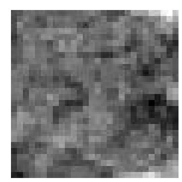
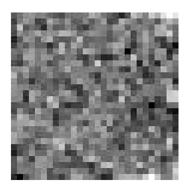
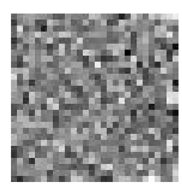
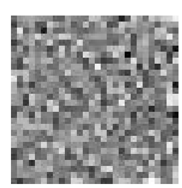
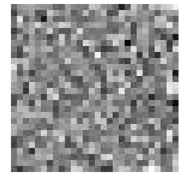
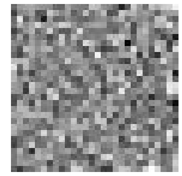
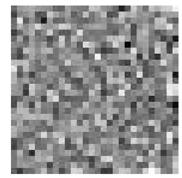
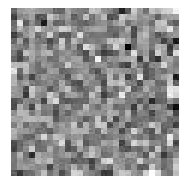
REPORT

Below are the accuracy calculated at different values of *lambda* for both L1 and L2 regularizations.

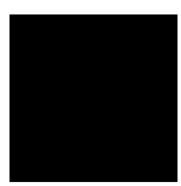
The third column shows an image which represents the weight vectors as an image. If the image is uniform that is either black or white, that means that the weights are similar to each other while sharp changes represent sharp difference in weights. We would hence prefer a more uniform image.

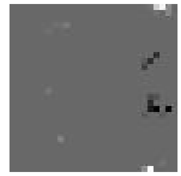
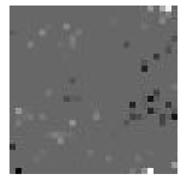
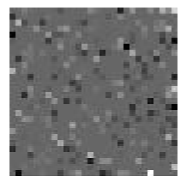
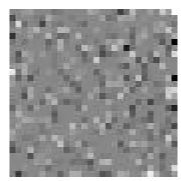
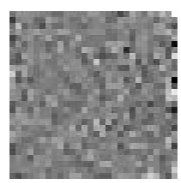
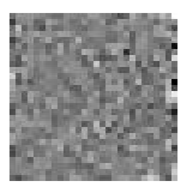
Regularization : L2

Lambda	Accuracy	Weight Image
10000000.0	0.955516014235	
1000000.0	**0.960854092527**	
100000.0	0.952846975089	
10000.0	0.940391459075	
1000.0	0.936832740214	

100.0	0.936832740214	
10.0	0.935943060498	
1.0	0.936832740214	
0.1	0.934163701068	

Regularization : L1

Lambda	Accuracy	Weights Image
10000000.0	0.492882562278	
1000000.0	0.492882562278	

100000.0	0.891459074733	
10000.0	0.944839857651	
1000.0	**0.955516014235**	
100.0	0.953736654804	
10.0	0.940391459075	
1.0	0.939501779359	
0.1	0.939501779359	

Observations

1. We see that we get max accuracy on the test data with L1 loss with gamma = 1000 (C=0.001) and L2 loss with gamma = 10^6 (C=0.000001) .
2. We observe that generally with increasing lambda (and decreasing C), the accuracy improves although it begins to decrease after a certain threshold especially in the case of L1 regularization.
3. With higher lambda, the weight vector is more uniform that is with higher regularization, the weights tend to be close by.

Inferences and Conclusions

1. We observe that with increasing lambda, the images are much more smoother(all weights have almost the same value) and less grainy(some weights have high value while some have low value). This is because lambda is directly proportional to the regularization involved and by the definition of regularization, it tries to even out the weights (specially L2).
2. Increasing the lambda means that we're giving more weightage to the regularization term in the loss function, which will tend to make the weights more evenly distributed, which theoretically should make the classifier more general at cost of wrongly classifying a few of the data points. Hence, for the testing data, as lambda increases, the accuracy also increases till a certain threshold after which the objective of the classifier in a way changes to just making the weights uniform.
3. Decreasing lambda on the other hand means we're giving more weightage to correctly classifying the training data w.r.t. to the regularization term, which may sometimes lead to overfitting. Thus a lower accuracy on the testing/validation data.
4. We observed that our weights were in the range [-1, 1], so, between L1 and L2 we see:
    a. Since L2 squares the weights, the regularization loss would be smaller. Hence, weights would react slower to change in lambda.
    b. Whereas L1 directly takes sum of the absolute values of the weights, so, the regularization loss would be higher, hence, the change in weights would be more prominent with changing lambda.