PREDICT-X MINDSPARK 2019

FINAL REPORT

CHINMAY PRAMOD BHARTI.

ADVAIT KUMAR.
IIT BOMBAY.

We didn't know much about adversarial attacks, so we read about it and were quite fascinated by the concept.

We had to make the model robust to 4 types of attacks so our approach was to include lots of parameters that are a long and broad neural network. We first tried it on using resnet by resizing and converting the image in RGB. But taking into account the complexity of resnet results were not worth that much.

A simple neural network consists of 4-5 layers. So we decided to extend it up to 20 layers, keeping the format of combinations of conv2d, max-pool, batch-norm along with dropout to avoid overfitting. Reason for selecting dropout over drop connect was the statistical comparison of performance.

After making a decent network we started to train it. We tried many sequences and finally arrived at this sequence
1)for 70 epochs, learning rate 0.01, SGD, training for all 4 types.
2)for 30 epochs, learning rate 0.001, Adam, training for all 4 types.
3)for 30 epochs, learning rate 0.001, Adam, training for only the PGD attack.

SGD is faster than Adam at initial stages so used SGD at the beginning, Then shifted to Adam because it's better for fine-tuning
The PGD attack accuracy was relatively lesser than other 3 hence the last step was to exclusively improve this. But surprisingly it pushed the other accuracies too.

We wanted to experiment a bit more on this problem but were not able to because of mid-sem exams But I would like to highlight that adding 2-3 more fully connected layers will improve the accuracy.