Advanced Machine Learning

Assignment 1

Roudranil Das
MDS202227
roudranil@cmi.ac.in

Saikat Bera
MDS202228
saikatb@cmi.ac.in

Shreyan Chakraborty
MDS202237
shreyanc@cmi.ac.in

Contents

1 About the datasets

In the given datasets we have images consisting of hand gestures corresponding to the english sign language and images consisting of dressing items. The images are 28×28 in resolution.

2 Processing the dataset

The images were loaded from the corresponding csv file, where individual images are stored as rows, the columns being the pixel values. The images were then loaded as tensors and then normalised -

- for fashion mnist, we normalised the pixel values to 0 mean and 1 standard deviation
- for sign language mnist we divided all pixel values by 256 to bring the values between 0 and 1

3 Fully connected NN vs CNN

We designed, trained and compared the performance of 2 basic DNN and CNN architectures on both the datasets. However even though the architectures are simplistic, when comparing the performances of the 2 models on the datasets, it is evident that CNN's are the better fit for image data.

CNN's are far more able to capture the local patterns and the hierarchies in the images - which implies that they are able to extract better features from the images.

4 Comparing results

To compare results we trained both models for 10-30 epochs, and compared the results.

	Loss		Accuracy		
Model	Best	Last	Best	Last	Time/Epoch.
DNN	0.3045	0.3154	89	89	18s
CNN	0.2136	0.26	93	92	24s

Table 1: Fashion mnist

	Loss		Accuracy		
Model	Best	Last	Best	Last	Time/Epoch.
DNN	1.193	1.358	68.12	66.13	60s
CNN	0.786	1.363	80.73	79.48	61s
CNN with dropout	0.377	0.419	90.31	89.3	62s

Table 2: Sign language mnist

5 Dependence of CNN on visual information

CNN's rely on the visual information and patterns present in images to a varying degree. We can judge this by seeing if the models trained on a permutation of the pixels of the original images performs the same as the model trained on the original images.

For fashion mnist, which contains fairly simple images, there was not much difference in this scrambling - accuracy reduced from 93% to 87%.

However for sign language mnist, there was a noticeable drop - from 90% to 52%.

6 FGSM attack on the networks

Here is the performance of the FGSM attack on the 2 datasets

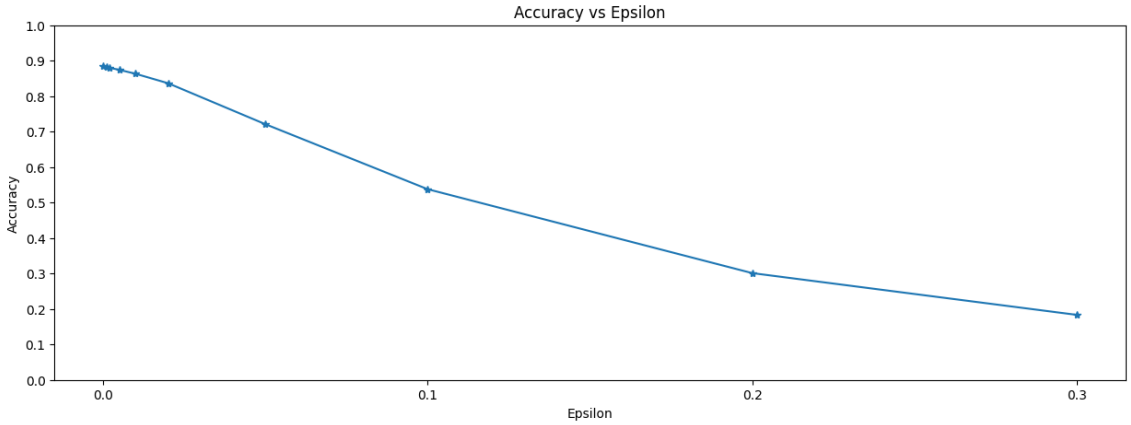

Figure 1: Fashion mnist CNN

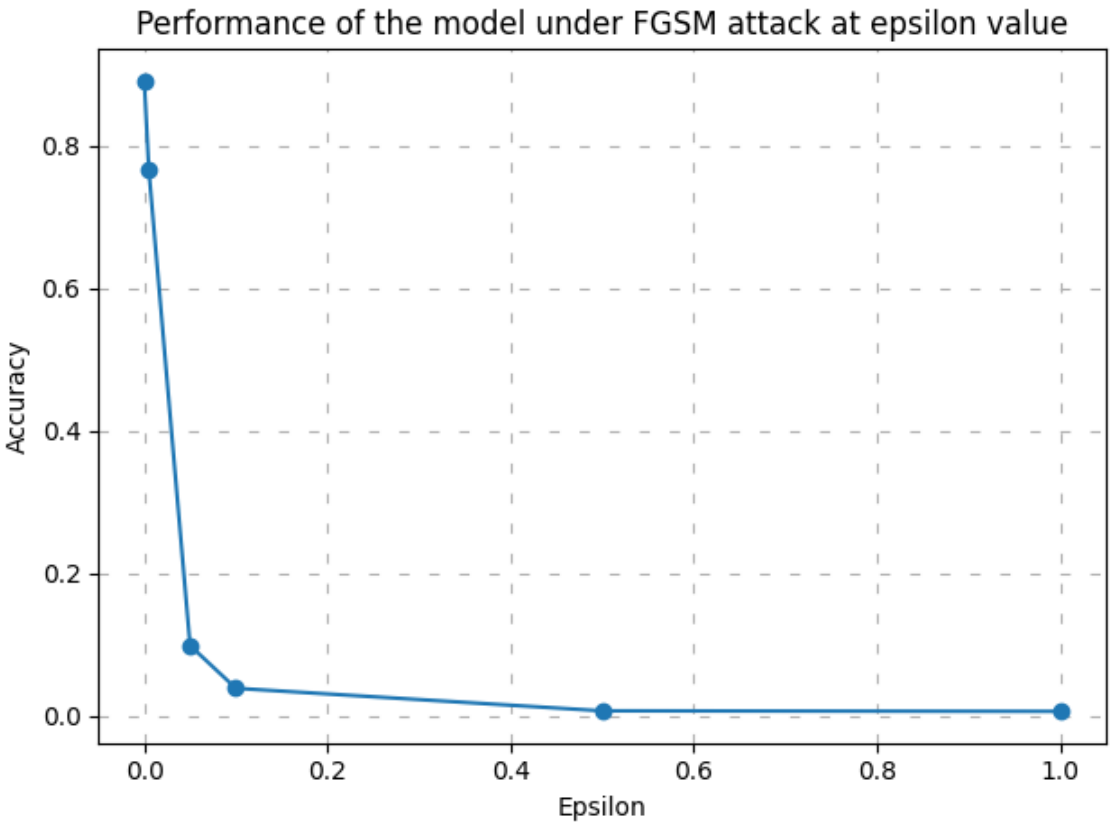

Figure 2: Sign language mnist CNN