

DAI ASSIGNMENT 3

Report

Sanyam Jain

Question 1:

Part A: Testing Accuracy: 66.43 with 30 EPOCHS (Modified and customized VGG Model)

Part B: Please refer notebook for visualizations

Part C: Please refer notebook for visualizations

Question 2:

LIME generates local explanations, in other words, explanations for individual instances in a dataset. LIME starts by generating a new dataset of perturbations around the instance to be explained. Then, the trained machine learning classifier is used to predict the class of each instance in the new generated dataset. Finally, a simpler model with intrinsic interpretability, for instance a linear regression model, is fitted and used to explain the prediction of the classifier. Before fitting a simpler model, the instances in the new generated dataset are weighed based on their distance to the original instance being explained. In this way, there is a higher certainty that the model is locally faithful around the explained instance.

References:

1. <https://github.com/Divyanshupy/CLASS-ACTIVATION-MAPS/blob/master/Class%20Activation%20Maps%20with%20MaxPooling.ipynb>
2. https://youtu.be/vz_fkVkoGFM
3. <https://nbviewer.jupyter.org/urls/artegac.github.io/blog/lime.ipynb>
4. <https://github.com/marcotcr/lime/tree/master/doc/notebooks>
5. <https://github.com/marcotcr/lime/blob/master/doc/notebooks/Tutorial%20-%20images.ipynb>