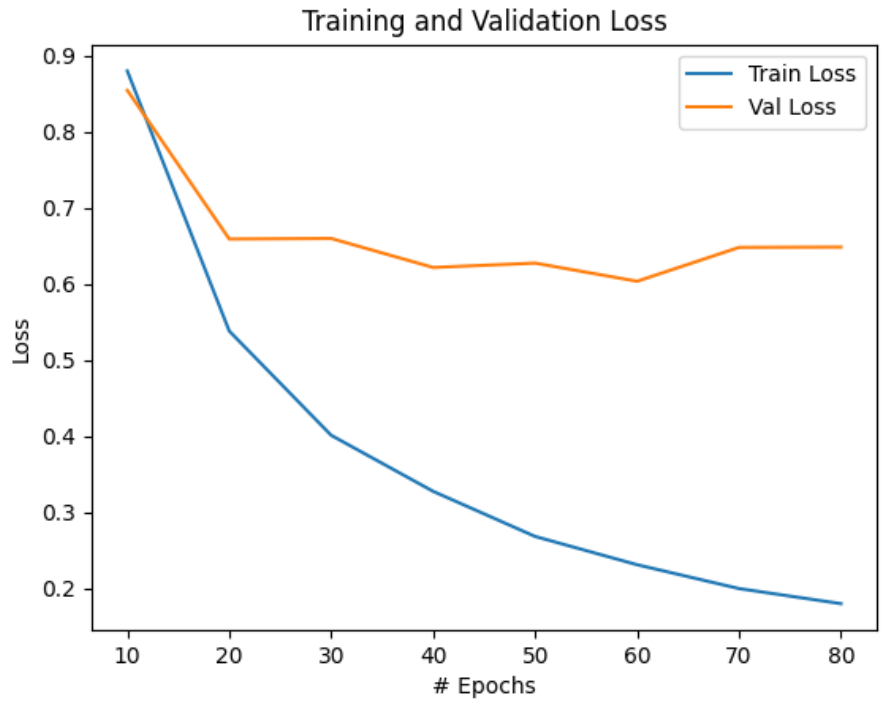
CS579 – Homework 4 Solutions

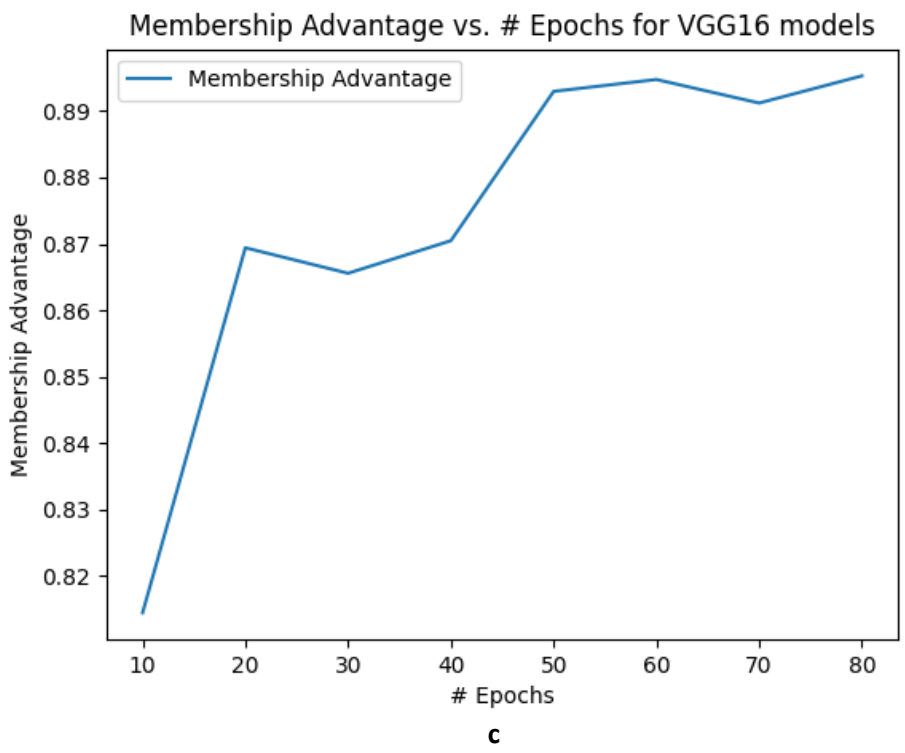
Opeyemi Ajibuwa

Task 1: The plots of the {# epochs} vs. {train and test loss} for FashionMNIST and CIFAR-10 are shown in Fig. 1(a) and (b). The plots of {# epochs} vs. {membership advantage} for FashionMNIST and CIFAR-10 are shown in Fig. 1(c) and (d).

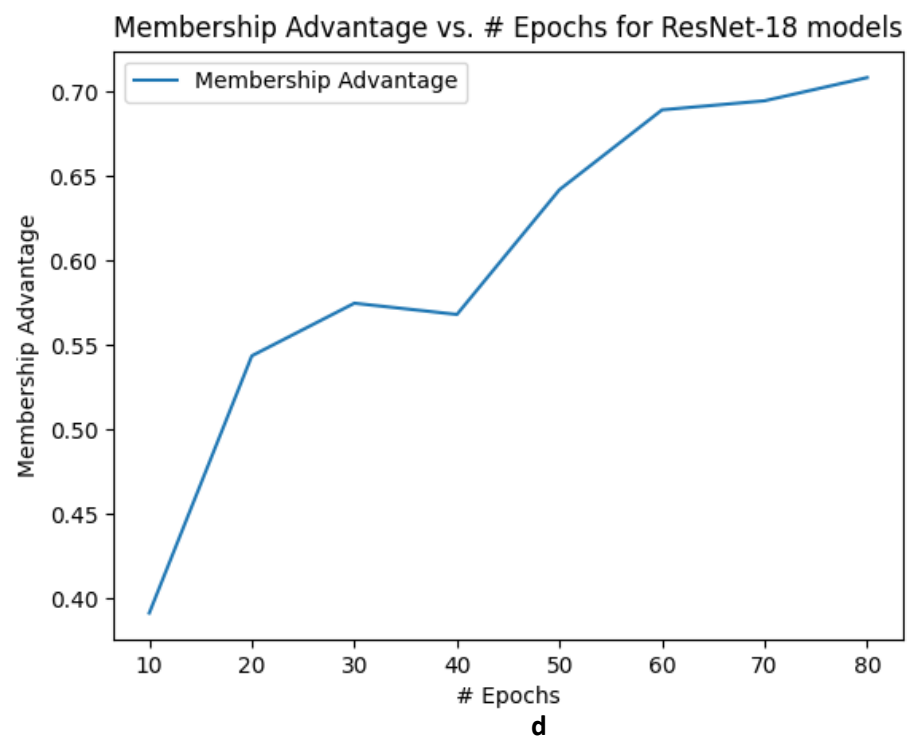
a



b



c

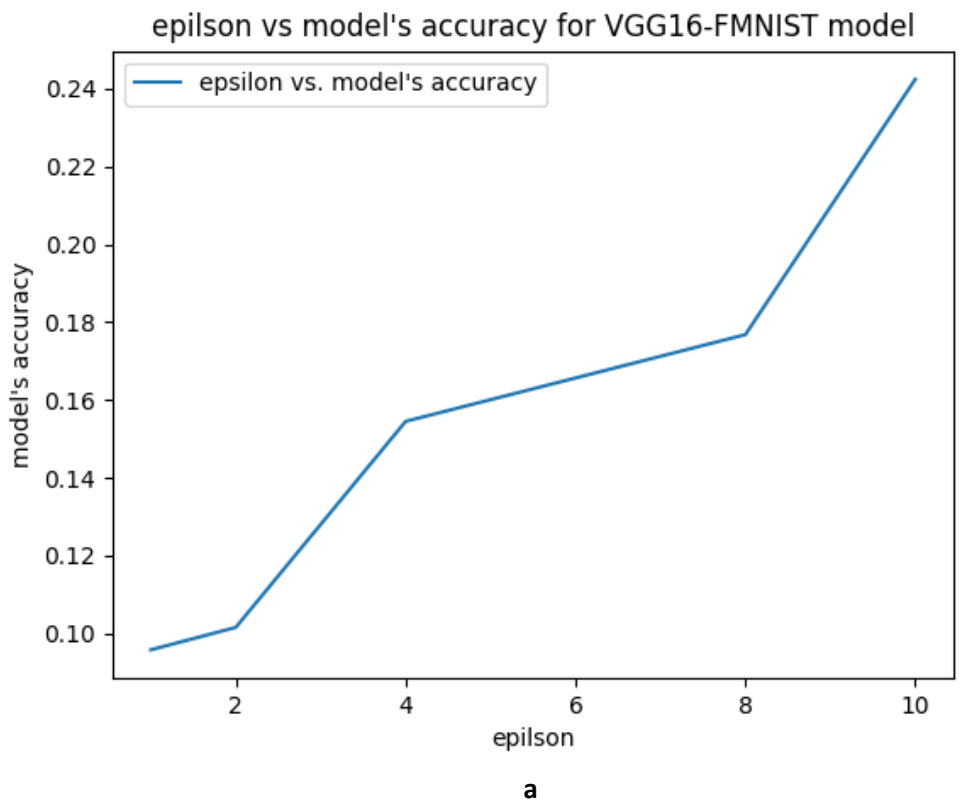


d

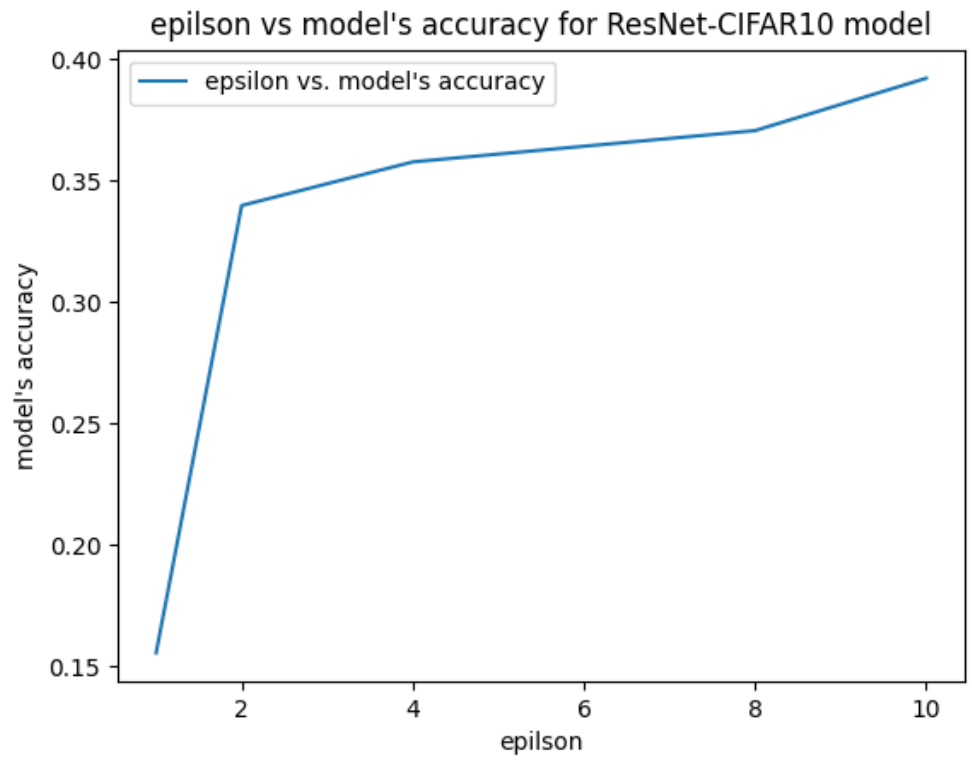
Figure 1

Results analysis: As seen in figure 1a, the validation loss diverges quickly from the training loss for the VGG16-FMNIST model. As the number of training epochs increases, the validation loss increases while the training loss decreases asymptotically. In contrast, as figure 1b shows, the ResNet-CIFAR-10 model's training and validation loss tends to decrease with more epochs which suggests the model is not very overfitted to the training data as opposed to the VGG16 model. As both figures 1(c) and (d) show, the membership advantage seems to increase with more training epochs in both models. This observation suggests that longer training epochs lead to overfitted models which correspondingly leads to more privacy leakage, thus the attacker has more advantage to infer memberships of the training dataset of the models.

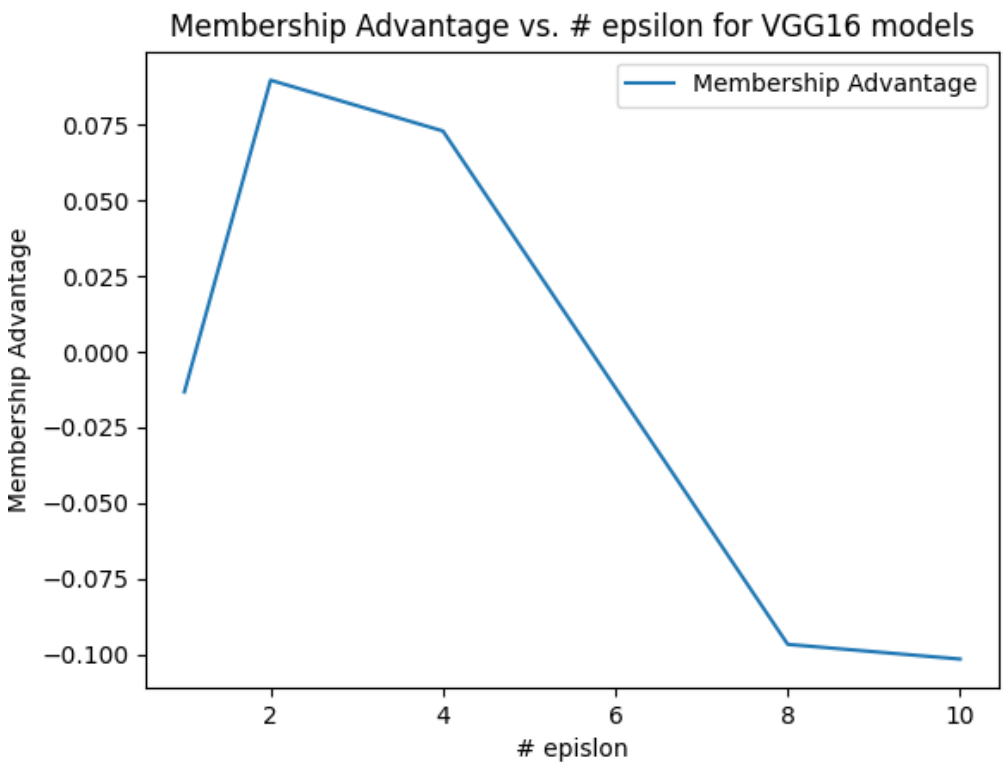
Task II: The plots of the {epsilon} vs. {test acc. and membership advantage} for FashionMNIST and CIFAR-10 are shown in Fig. 2(a-d).



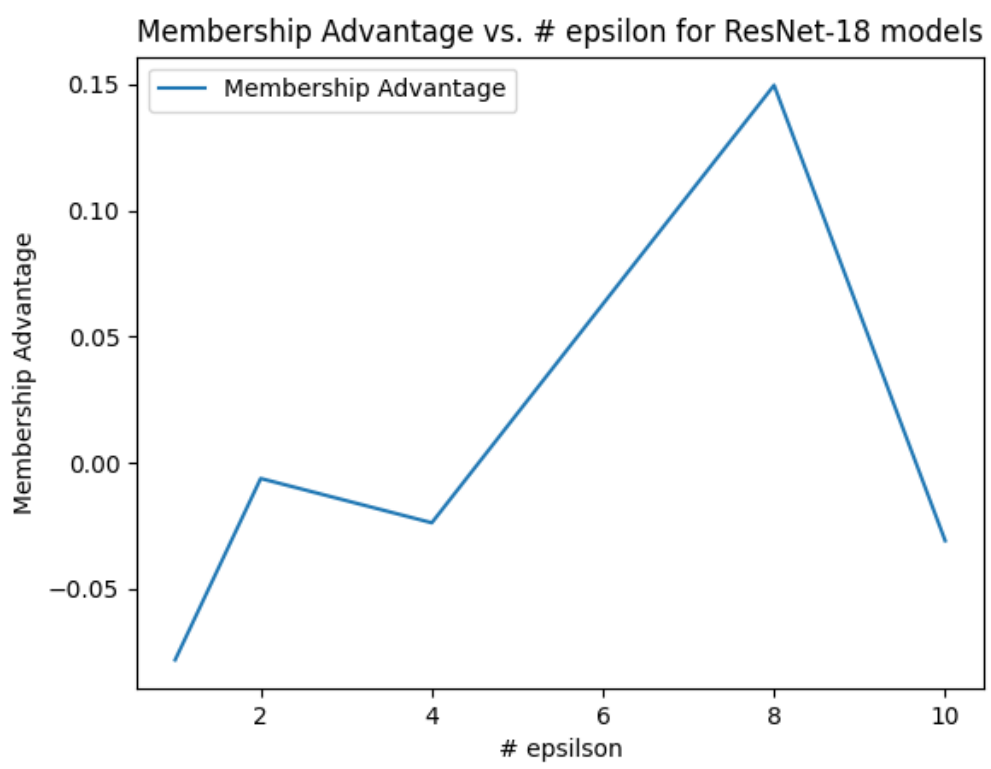
a



b



c



d

Figure 2

Results analysis: Notably in figures(a) and (b), the classification accuracy of both differentially trained models drop significantly from the previous non-differentially trained models. This drop-off is expected as well as the observation that the accuracy of the models increases as epsilon increases, and the lowest classification accuracy is obtained as epsilon gets closer to 0. This behavior suggests that as epsilon, the privacy budget increases, the model classification accuracy increases, and the model's utility is improved. The opposite is also true. The membership advantage of the differentially trained models in fig 2(c) and (d) generally decreases compared to the previous models in task 1. This observation is consistent with our expectation. However, for the VGG16 model, the observed behavior from fig 2(c) conflicts with our expectation that the membership advantage increases with increasing epsilon. This anomaly might be due to inconsistency in the training of the VGG16 models which can be further confirmed with more training iterations.