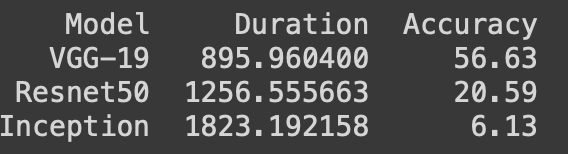
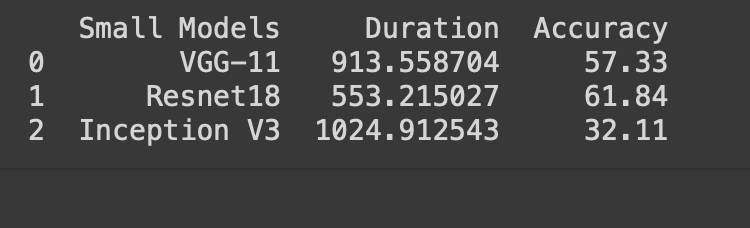
**HW 5 Writeup**

CNN write up:

Of the CNN models I trained on the Tiny ImageNet dataset, the VGG-19, had both the highest accuracy at 56.63% and the shortest training duration of approximately 896 seconds (14.9 minutes). I only trained the models for a single epoch, because it was taking a very long time to train them using the Google Collab free version, so my guess is that the VGG-19 having the highest accuracy could be due to its relative architectural simplicity, which consists of a series of convolutional layers with small filters followed by max-pooling layers, which may have resulted in better feature extraction for the Tiny ImageNet dataset within that single epoch. The Resnet50 and InceptionV4 models had lower accuracies than the VGG-19 of 20.59% and 6.13% respectively. These models also took longer to train, with Resnet50 taking approximately 1257 seconds (20.95 minutes) and Inception taking about 1823 seconds (30.4 minutes). My guess is that the lower accuracy and longer training times could be attributed to their architectural complexity and the fact that I only allowed them to train for a single epoch. These would potentially outperform the VGG-19 if more computing resources were available to train over more epochs within a reasonable time frame. Of the small models, the Resnet18 had the highest accuracy, at 61.84%, followed by the VGG-11 with 57.33%, and the InceptionV3 with 32%. Interestingly, the smaller models (VGG 11, Resnet 18, and InceptionV3) all outperformed the larger models in terms of accuracy. I think this is likely because I only trained each model for a single epoch and larger models require more time to train. So by limiting the training time, they are not converging to their optimal performance, whereas the smaller models are able to reach decent accuracies much quicker.



RNN writeup:

For this part of the homework, I synthetically generated time series data for fictional winter coat sales each month over a period of 10 years, with the winter months reflecting higher sales than the summer months. Overall, it appears that the large models, which were trained on the 3x augmented dataset (30 year timespan), outperformed their smaller versions in terms of RMSE score. Visually inspecting the plots I created from the predicted vs. actual sales, it makes sense that the larger models seem to be a closer fit to the data. Of all the models, the Bidirectional RNN had the lowest RMSE across both the small and large dataset sizes, which is likely due to bidirectional nature capable of identifying patterns that unidirectional models might miss. The model with the biggest change in RMSE between the small and large datasets was the Deep RNN. This makes sense to me because the Deep RNN has the most complex architecture of all the given models, so it is possible that this complex of a model had a hard time with only 10 years of data but then was able to perform drastically better when given a more extensive dataset to learn from. Overall, the ‘large’ models trained on the 3x large dataset took the longest duration to train, which is not surprising because the larger the dataset the longer it will take for a model to train.

