Week 8: Training CNNs to Differentiate Dogs and Cats

Background and Motivation

This analysis was performed on the Dogs and Cats dataset, made available via the "Dogs vs Cats Redux: Kernels Edition" Kaggle competition (Kaggle 2022). The dataset contains 37,500 images, with 25,000 "training" observations and 12,500 "test" observations. Half of the images are of cats, the other half of dogs. Successful work with this dataset allows for the application of computer vision to characterize images, an important foundational step for automated scene identification, self-driving cars, and other applications.

Approaches to Pre-Processing the Data

Each image in the original dataset contained 22,500 pixels (150x150 square) across three channels, scored from 0-255 representing the depth of shading. All metrics were MinMax scaled from 0-1. A 20% holdout was used for validation. Given the computation complexity of the dataset, the 80% training set was not further enriched (e.g., via image scaling or rotation).

Modeling Approach

The focus this week was on developing convolutional neural networks, an approach where images are convolved to define fundamental features. These features ultimately are pushed into a traditional dense ANN architecture for final processing. Due to the training time required, model training used only the 80% training set, and early-stopped with a 2-epoch patience. Three models - with similar structures - were tested:

- Model 1: a three-layer convolution, where each layer doubled the number of channels (from 3 to 32 to 64 to 128) while halving the image dimensions. After the convolutions, pass results through two dense hidden layers, each with 256 nodes
- Model 2: for each layer in Model 1, add an additional convolution

- Model 3: adjust Model 1 to only have a single dense hidden layer.

Results:

	Time per epoch (min)	Loss (Val)	Loss (Test)
Model 1	18	0.4	1.88
Model 2	46	0.42	2.08
Model 3	16	0.38	1.73

Implications

- More complex models were not necessarily better. Model 3 was the simplest structure but performed the best of the three models tested. Importantly, the training time was abridged across two dimensions: a conservative patience (only two epochs) and no adjustment to model learning rate. It is likely that the other models would have performed better given adequate space to train
- Kaggle leaderboards appear to be dominated by transfer learning. The three models developed were all built from scratch, and underperformed the Kaggle leaderboards. Leveraging off-the-shelf CNNs likely would have materially improved performance
- Using the right hardware is critical. The models were developed in Google Colab using TPUs, typically run early in the morning. As a test, Model 2 was initially launched using CPUs (~200 minutes / epoch) and again using GPUs (~75 minutes / epoch).
- Not shown in the results above, but using an appropriate loss function is critical. I tested the submission of Model 1 scores translating predicted probability into categorical predictions, and the loss function was notably higher (>17 versus 1.88 when predicted probabilities were submitted). The value of the log-loss penalty structure is that it over-weights being very wrong. Since there was such a large difference between submissions of probabilities and submissions of categories, it implies that many of the failed classifications were in border / edge cases.

References

"Dogs vs Cats Redux: Kernels Edition." Kaggle. Accessed February 20, 2022. https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/overview.