Machine Learning Engineer Nanodegree Capstone Proposal

Cats vs. Dogs
(from a recent Kaggle competition)

Resmi Arjunanpillai
Sep 14, 2017

Domain Background

Computer vision as a space has seen rapid advances in just the last few years. Deep learning has created very high accuracy in recognizing images, leading to applications in multiple areas like health, self driving cars, robotics etc.

A few of the key papers in the space:

- Alexnet is the paper that jump started the current interest in CNNs. This paper by Geoffrey Hinton and others, describes a deep convolutional neural network that won the 2012 Imagenet competition.
  https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
- VGGnet was another influential paper that demonstrated the use of smaller 3x3 filters and promoted deeper networks.
  https://arxiv.org/pdf/1409.1556v6.pdf

Considerable research has also gone into optimizing hyper parameters. A few examples

- Momentum: https://arxiv.org/pdf/1412.6980.pdf and .
- Adam optimizer: https://arxiv.org/pdf/1412.6980.pdf

Problem Statement

https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition
The Dogs vs. Cats Redux: Kernels Edition on Kaggle is a competition to differentiate images of cats from images of dogs.

Though the problem is a well researched one, I think trying out many different approaches to achieve a very high level of accuracy to solve the problem can be challenging. There are multiple ways to approach the problem from building a network from scratch to transfer learning using pre-trained networks. Optimizing the learning parameters, using batch normalization, data augmentation etc would be important.

Datasets and Inputs

Data: https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data

The train folder contains 25,000 images of dogs and cats. Each image in this folder has the label as part of the filename. The test folder contains 12,500 images, named according to a numeric id. For each image in the test set, the goal is to predict a probability that the image is a dog (1 = dog, 0 = cat).

Solution Statement

My goal would be to achieve a logloss of 0.1 or lower on the test set (which will put me in the top 1/3rd of the competitors)

Benchmark Model

Benchmark for the model is a log loss of 0.03 achieved by the winner of the competition

Evaluation Metrics

The contest uses log-loss as the evaluation metric

$$\text{LogLoss}=-1n\sum i=1n[yilog(\hat{y}\ i)+(1-yi)log(1-\hat{y}i)],$$

where

- n is the number of images in the test set
- ŷ iy^i is the predicted probability of the image being a dog
- yiyi is 1 if the image is a dog, 0 if cat
- log()log() is the natural (base e) logarithm

A smaller log loss is better.

LogLoss is related to cross-entropy and measures the performance of a classification model where the prediction is a probability value between 0 and 1. LogLoss takes into account the uncertainty of the prediction based on how much it varies from the actual model.

Project Design

- Start with a small CNN, optimize learning rate, dropout rate etc. Try batch normalization and multiple forms of data augmentation for better results. Add additional layers as needed.
- Use one or two pretrained networks and optimize with fine-tuning.
- (optional) Ensemble results from multiple models

A good solution to this problem can be used to tackle other computer vision problems as well.