Kaggle Competition

Collaborative Competition
Pneumonia

Background

- Kaggle - Competitions, Training, Datasets, Blogs, etc.
- Pneumonia - Infection in the Lungs - mostly diagnosed by X-Ray plus other observable data (e.g., temperature)
- CNN - Convolutional Neural Network
  - Great for Classification of Images
  - Uses Convolution Functions to transform images to develop features from source images
  - Hidden Layers - Uses repeating sets of layers
    - Convolution
    - Relu - sets < 0 weights to 0
    - Pooling
  - Classification - Finally Flatten, Fully Connected, and Softmax
- Google Colab - Google's Jupyter Notebook Platform

Kaggle Competition

- RSNA Pneumonia Detection - Use Provided Labeled Data and X-Ray DICOM Images
- Data
  - https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
  - Stage One
    - Images: stage_1_train_images.zip and stage_1_test_images.zip.
    - CSVs:
      - stage_1_train_labels.csv - patientId, bounding boxes - **x, y, width, height**, Target (0 or 1)
      - stage_1_sample_submission.csv - provides the IDs for the test set, as well as a sample of what your submission should look like.
      - stage_1_detailed_class_info.csv - detailed information about the positive and negative classes in the training set - possibly other features

Top Kaggle Kernels

- CNN Segmentation (RESNET-Depth 5)
- START HERE: Beginner Intro to Lung Opacity S1
-

CNN Background

- Siraj Videos / Github Repos
  - Convolutional Neural Networks
    - https://www.youtube.com/watch?v=FTr3n7uBIuE
    - https://github.com/llSourcell/Convolutional_neural_network
- Other
  - Adventures In Machine Learning - Implementing Convolutional Neural Network in PyTorch
    - https://github.com/adventuresinML/adventures-in-ml-code/blob/master/convolutional_neural_network_tutorial.py
  - Adventures in Machine Learning - A Pytorch Tutorial - Deep Learning in Python
    - https://github.com/adventuresinML/adventures-in-ml-code/blob/master/pytorch_nn.py
  -

Tonight’s Goal

- Break up into Teams - mix experienced with novices
  - Who has done this before?
- Review Data
  - Summary Stats, explore distributions, etc.
- Review Images
  - Sample Metadata
  - Generate a Heatmap of Bounding Box Locations
- Review Prediction Goals
- Possible Develop Strategies
- Regroup - Share Observations
- Share Strategies

Setup

- Kaggle - Signup / Login
- Join Pneumonia Competition (Possibly Validate account if not done yet)
- Download API Key if you haven’t already
- Go to Google Collab
- Import Notebook
  - https://gist.github.com/divdev/e391d3ccaf26160051c9e50ca50d2efa.js
  - Download
  - Import in Collab
- Run first 2 cells
- Import downloaded API key as **kaggle.json**
- Run rest of cells.

Add Cell to End of Starter Colab Notebook

- Add Cell at end of notebook
- Start building the CNN using Pytorch or other (Keras, Layers, etc.)
- Train using a random sampling
-