LEARNING with LIMITED LABELED DATA

Shioulin Sam @shioulin_sam @FastForwardLabs
Learning with Limited Labeled Data
Pneumonia

No Pneumonia
Machine Learning Model
Labeled Data
Labeled Data \hspace{2cm} \rightarrow \hspace{2cm} \text{Machine Learning Model}
Labeled Data → Machine Learning Model → Pool of Unlabeled Data
Labeled Data → Machine Learning Model → Pool of Unlabeled Data

Point That is Difficult for Machine
Labeled Data $\rightarrow$ Machine Learning Model $\rightarrow$ Pool of Unlabeled Data $\downarrow$ Point That is Difficult for Machine
Labeled Data → Machine Learning Model → Pool of Unlabeled Data

Label for Difficult Point ← Point That is Difficult for Machine
Labeled Data → Machine Learning Model → Pool of Unlabeled Data

Label for Difficult Point ← Point That is Difficult for Machine
Accuracy vs Rounds on Active Learning Iterations
Labeling Smartly
Random Sampling
Uncertainty - Margin Sampling
Uncertainty - Margin Sampling
Uncertainty - Entropy
Extending to deep learning
Distance from the decision boundary

Computing the distance between the datapoint and the closest neighbor of a different class
Distance from the decision boundary

Using adversarial perturbations to estimate the distance

Image from https://openai.com/blog/adversarial-example-research/
Distance from the decision boundary

Computing the distance between the datapoint and the decision boundary using perturbation magnitude
Bayesian Framework
Dropout
Ensemble
Resources

● Prototype - Active Learner (activelearner.fastforwardlabs.com)
● Blogs (blog.fastforwardlabs.com)
  ○ Guide to learning with limited labeled data
  ○ An invitation to Active Learning
  ○ Visualizing Active Learning
  ○ Meta Learning
  ○ Weak Supervision using Snorkel
● TWiML & AI  [link](https://twimlai.com/twiml-talk-255-learning-with-limited-labeled-data-with-shioulin-sam/)
THANK YOU