Machine learning ||: Bayesian & Unsupervised Methods

Research Paper presentation

Title: Deep Bayesian Active Learning with Image Data

Presented by:

Alexander poppe
Pramod kumar Nagaraj

Research paper author:

Yarin Gal
Riashat Islam
Zoubin Ghahramani

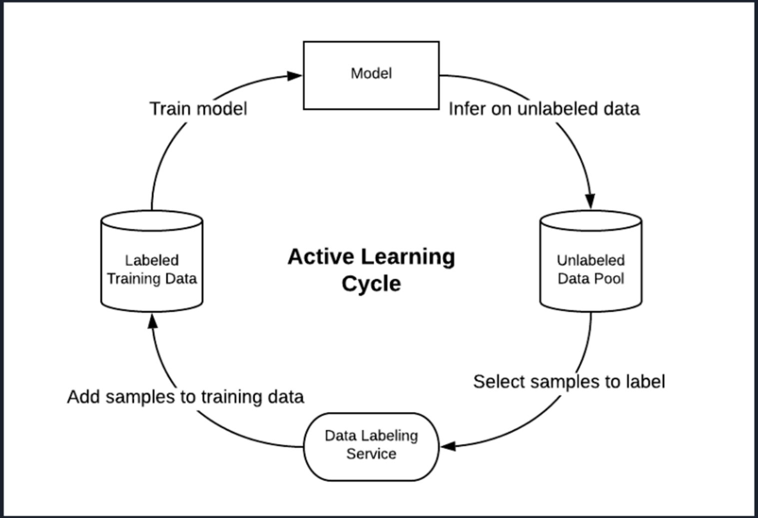
Context

- Majority of machine learning models built under supervised
- Data hand labelled and Collect the labels via feedback mechanism
- Samples to be labelled
    - Label all of them - get very expensive
    - Label randomly - cause redundant
    - Concept of active learning

Implementing an Active learning based approach to choose samples from unlabelled dataset provides the most value. Recent advance in Bayesian deep learning regarding extracting reliable uncertainty estimates from Neural network into Active learning

Active learning

- Proactively select a subset of sample
- Better performance with fewer labelled samples



<u>*Active learning cycle*</u>

1. Small sample get labelled
2. Train a model on labelled
3. Use train model to select unlabelled samples
4. Add label samples to training dataset
5. Repeat

- Step : Picking sample to be labelled ? How ?
- Results uncertainty. How to measure ?

Measuring Uncertainty and Gaussian process

- Simple way, just look at the output of model
- Neural network use softmax function
- Using softmax function model sometimes can be uncertain about prediction
- Other option: Gaussian process
- Gaussian process uses a non-parametric approach
- non -parametric approach results high dimensional spaces

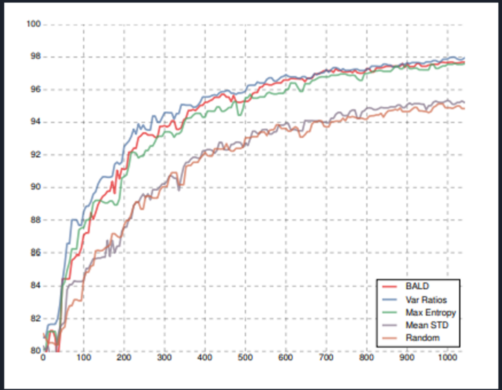
To overcome these we use <u>Deep Gaussian process</u>

Bayesian Deep learning / CNN

- Concept of Dropout
- Dropout applied only during training section
- Garin et al showed Dropout in NN is identical to variational inference in Gaussian process
- Method of Monte carlo dropout
- Example : Categorizing cat and dog result 40% cat, 30% dog. 30% neither

Connecting ….! Active learning

- Recall
- Using monte carlo dropout we can make multiple prediction for each unlabeled samples and use them to extract uncertainty estimates.
- Regression: Choose samples with high predictive variance to be labelled
- Classification: Use acquisition function. BALD gives best score.



Next code…………!

Our code






Acquisition functions – Var ratio

Labelled Set L, Unlabelled set U
Train Model

For X ***do:***

    take a sample from the unlabelled pool

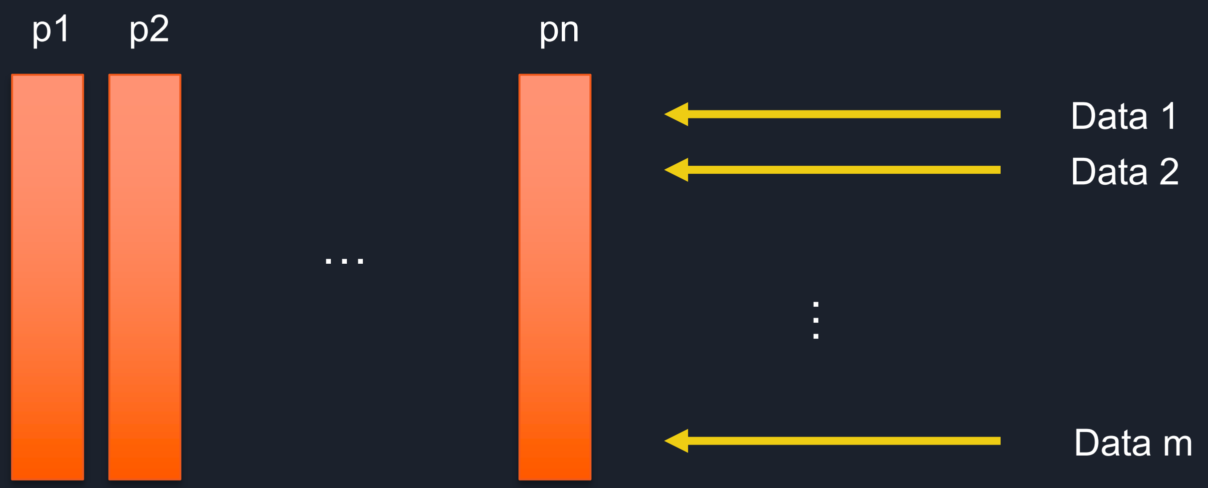
    For Number of predictions ***do***

        predict values of unlabelled sample
        gather these samples

    Predict var ratio’s
    add the 10 best data points to labelled data set

    train on labelled data set

Acquisition functions – Var ratio



Acquisition function - BALD

$$U(x) = H[p(\theta|D)] - \mathbb{E}_{p(y|x,D)} H[p(\theta|D,x,y)]$$

Acquisition function - BALD

$$U(x) = H[p(\theta|D)] - \mathbb{E}_{p(y|x,D)} H[p(\theta|D,x,y)]$$

Comparison to current active learning technique

- Compare to sparse existing literature of active learning with image data, which relies of the kernel method and further leverages the unlabelled datasets.
- Evaluate RBF (Radial basis function) kernel over the raw images to get a similarity graph which can be used to share information about unlabeled data
- Result expected to get minimised classification error which is referred as MBR (Minimum Bayes Risk)
- MBR is for binary classification case. So they compare to acquisition functions and they experimented with CNN version for MBR and the was not good.

Compared to semi-supervised learning

- In semi-supervised learning a model is given with a fixed set of labelled and unlabelled datasets.
- The model can use the unlabelled dataset to learn about distribution of input that might help with mapping to the output
- Several semi-supervised model which have set benchmark on MNIST given small number of labelled data (1k) and validation dataset (5K) and set of unlabelled data (huge) tuning hyperparameters and model structure. They compare to semi-supervised models which use a similar model structure of AL

Technique	Test error
Semi-supervised:	
Semi-sup. Embedding (Weston et al., 2012)	5.73%
Transductive SVM (Weston et al., 2012)	5.38%
MTC (Rifai et al., 2011)	3.64%
Pseudo-label (Lee, 2013)	3.46%
AtlasRBF (Pitelis et al., 2014)	3.68%
DGN (Kingma et al., 2014)	2.40%
Ladder Network (Γ-model) (Rasmus et al., 2015)	1.53%
Virtual Adversarial (Miyato et al., 2015)	1.32%
Active learning with various acquisitions:	
Random	4.66%
BALD	1.80%
Max Entropy	1.74%
Var Ratios	1.64%

Future….!

- What is interesting data to label ? (When model is uncertain)
- Active learning in real-world medical applications
- Much more


thank you
&