

Deep Active Learning using Monte Carlo Dropout

Lucas Moura

<https://github.com/lucasmoura>

lmoura@ime.usp.br

November 11, 2018

IME-USP: Institute of Mathematics and Statistics, University of São Paulo

Introduction

- **Deep Learning** is a growing field with state-of-the-art results in several areas.
- Image Classification, Machine Translation

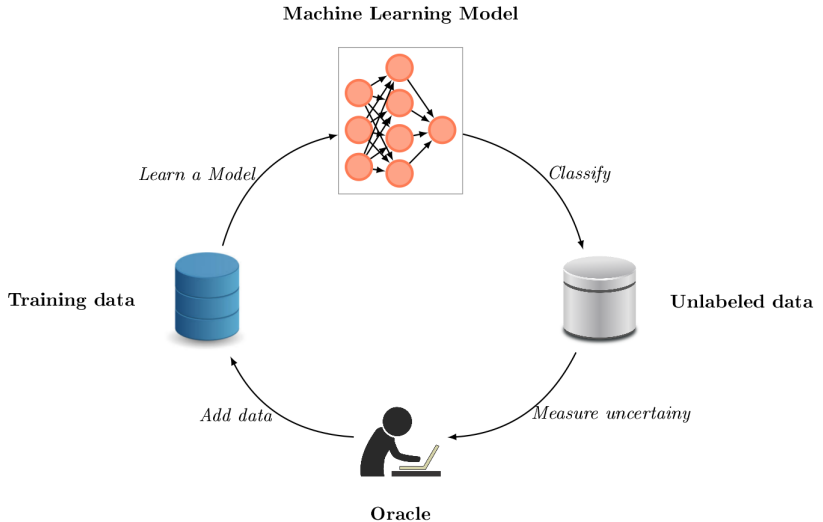
However...

- Training **Deep Learning** models require a huge amount of labeled data
- For the task of image classification on the ImageNet database, 1.2 million labeled images were used [1]
- This restriction causes huge difficulties on applying Deep Learning techniques to a wide range of problems, such as **Sentiment Analysis**

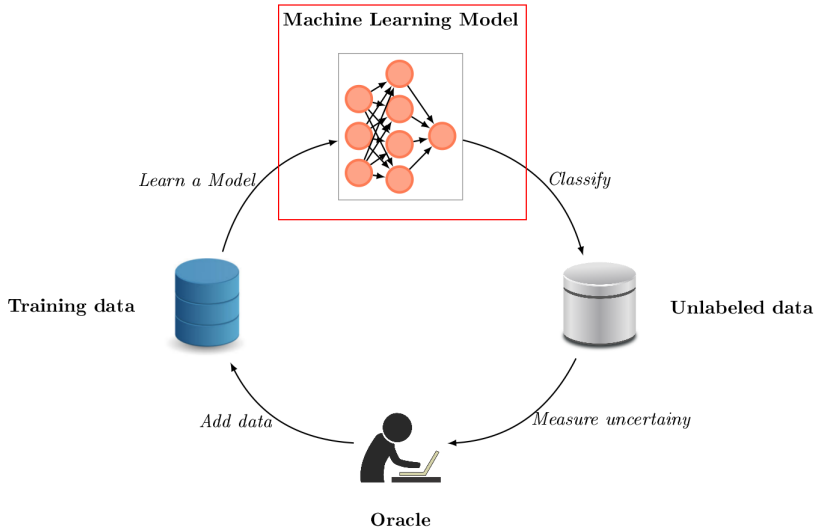
- Verify if a text is expressing negative or positive feelings.
- Huge amount of data, but few labeled.

Active Learning

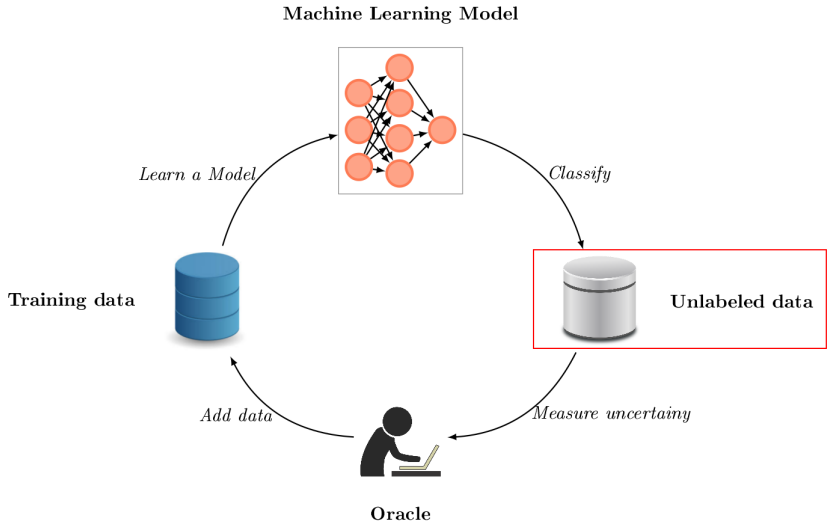
Active Learning



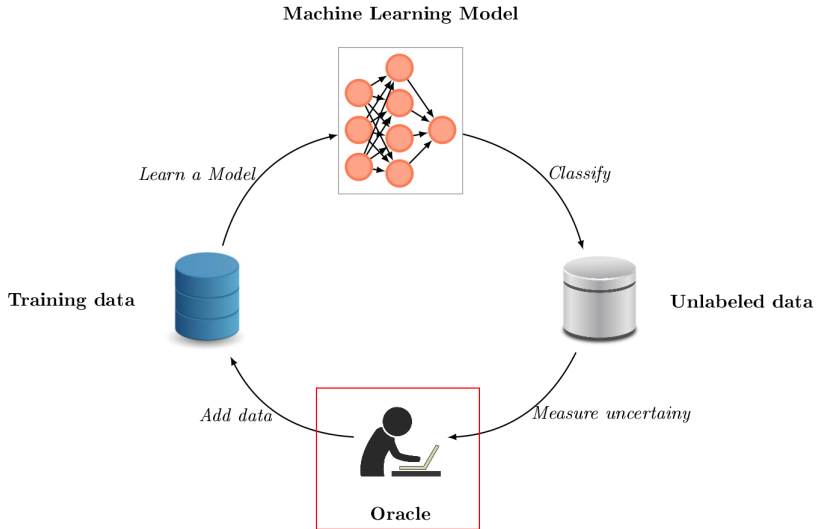
Active Learning



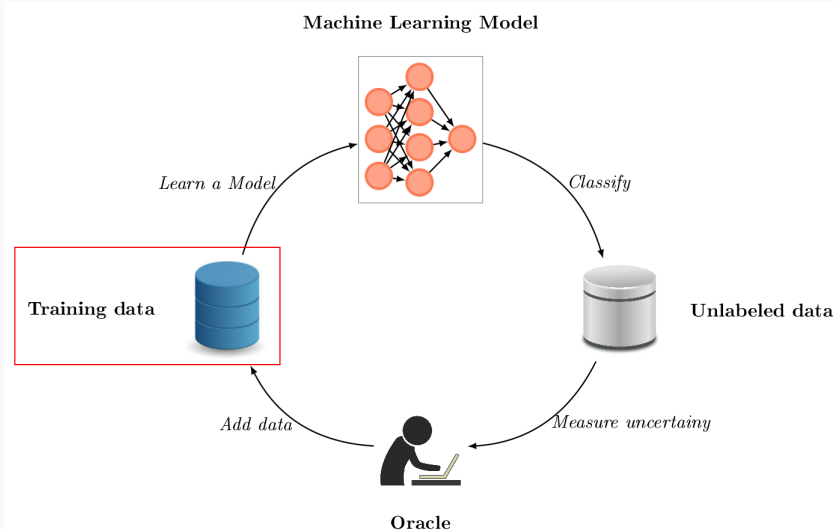
Active Learning



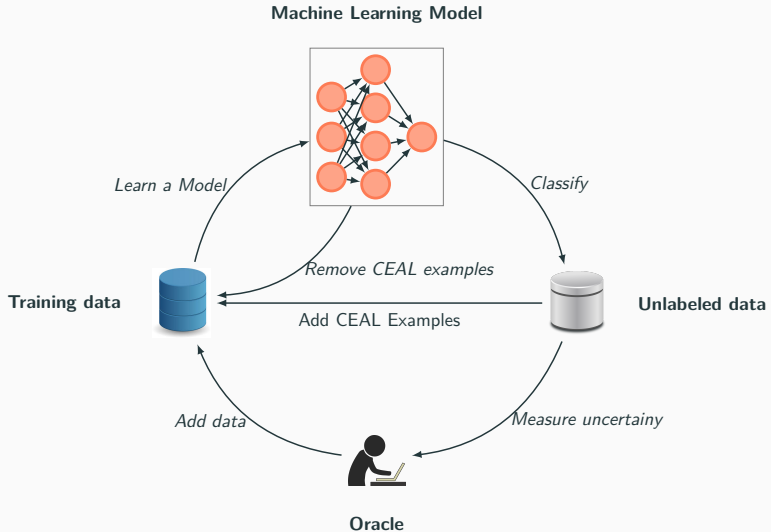
Active Learning



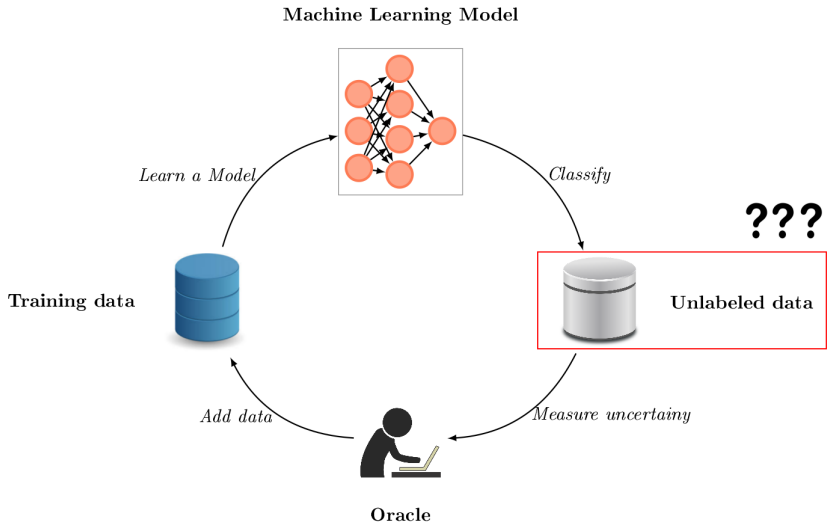
Active Learning



Active Learning - CEAL

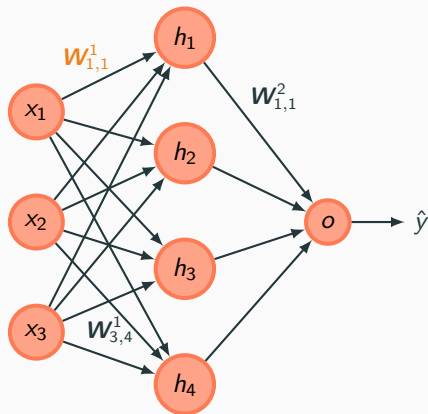


Active Learning

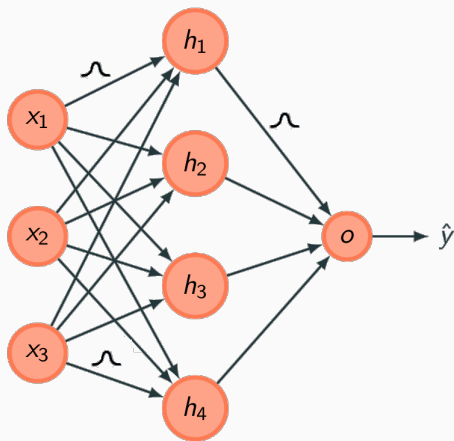


- To select informative samples, it is necessary to measure the **uncertainty** of the model prediction.

Neural Network

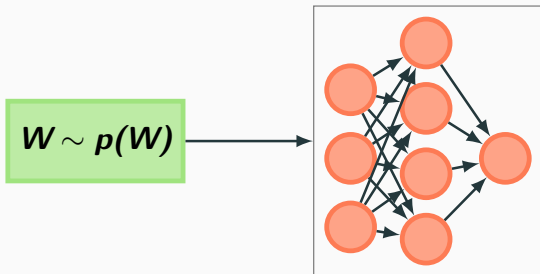


Bayesian Neural Network

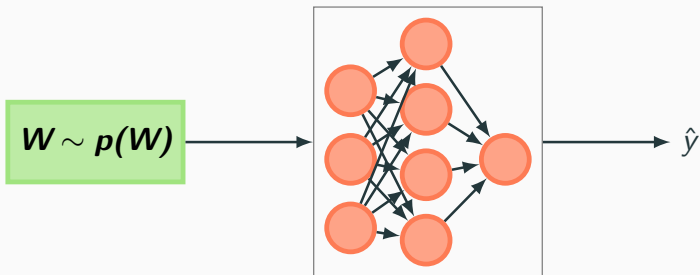


$$W \sim p(W)$$

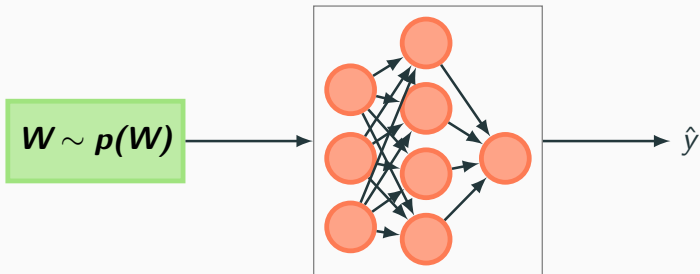
Bayesian Neural Network



Bayesian Neural Network

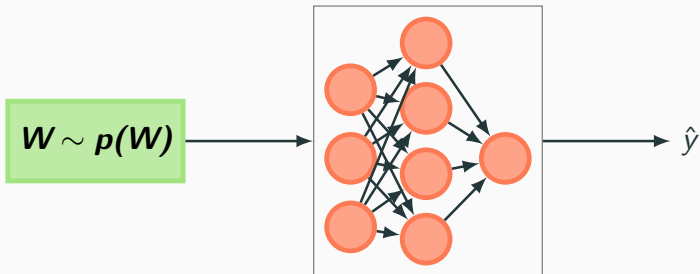


Bayesian Neural Network



Get T Classifications

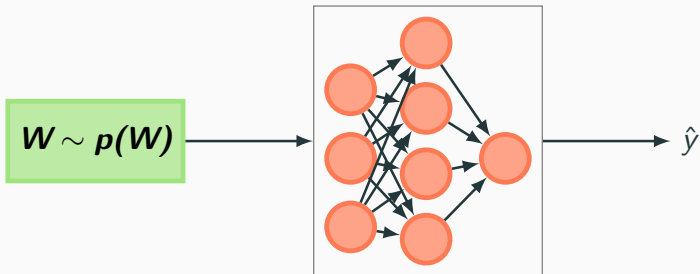
Bayesian Neural Network



Get T Classifications

$$\textit{Classifications} = [\hat{y}_1]$$

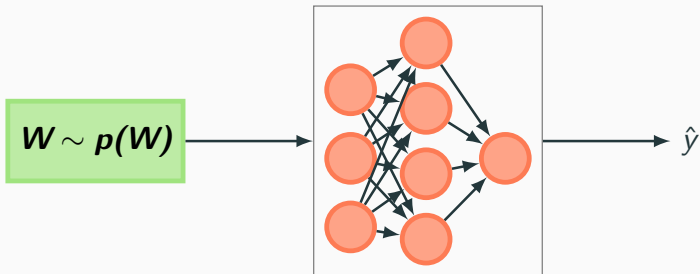
Bayesian Neural Network



Get T Classifications

$$\textit{Classifications} = [\hat{y}_1 \quad \hat{y}_2]$$

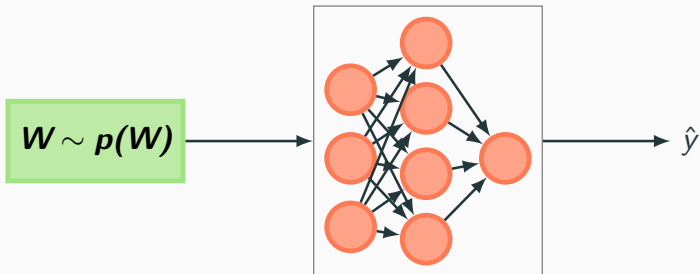
Bayesian Neural Network



Get T Classifications

$$\textit{Classifications} = [\hat{y}_1 \quad \hat{y}_2 \quad \hat{y}_3]$$

Bayesian Neural Network



Get T Classifications

$$\textit{Classifications} = [\hat{y}_1 \quad \hat{y}_2 \quad \hat{y}_3 \quad \dots \quad \hat{y}_T]$$

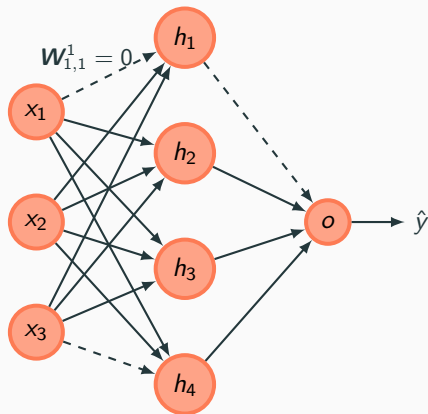
- Training Bayesian networks is a costly process
- Use techniques such as Variational Inference

- What if we could extract uncertainty measurements from current Deep Learning models if they use stochastic regularization techniques such as Dropout ?
- Uncertainty in Deep Learning (Yarin Gal, 2017)

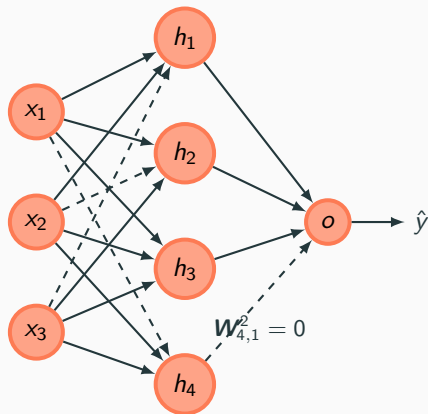
Dropout

- During training some weights are dropped from the network

Dropout



Dropout



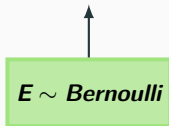
- The optimization function of Neural Networks using Dropout is practically the same as the optimization function of a Network trained with Variational Inference.
- Therefore it is possible to extract uncertainty measures from these networks, a technique called Monte Carlo Dropout.

Monte Carlo Dropout - Prediction

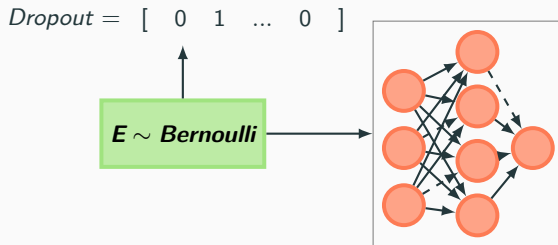
$$E \sim \text{Bernoulli}$$

Monte Carlo Dropout - Prediction

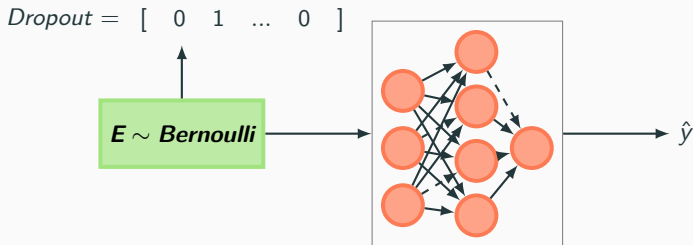
$$\text{Dropout} = [\ 0 \ 1 \ \dots \ 0 \]$$



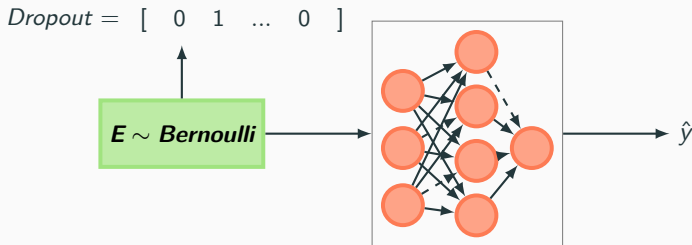
Monte Carlo Dropout - Prediction



Monte Carlo Dropout - Prediction

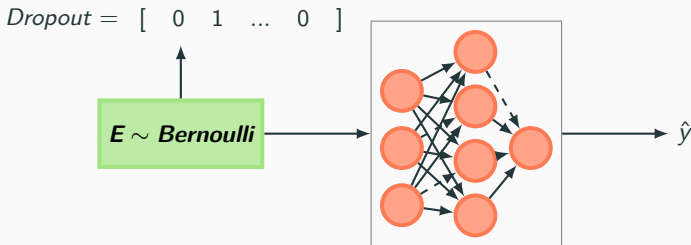


Monte Carlo Dropout - Prediction



Get T Classifications

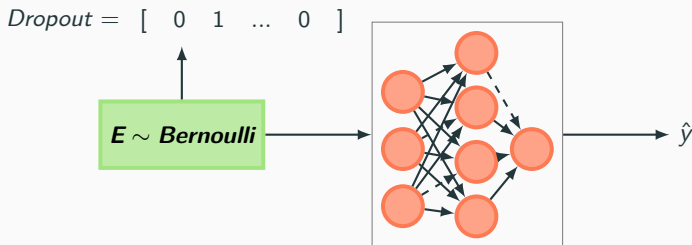
Monte Carlo Dropout - Prediction



Get T Classifications

$$Classifications = [\hat{y}_1]$$

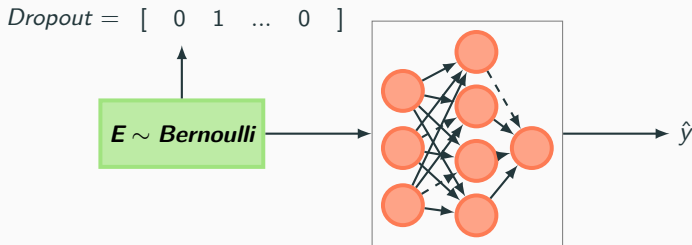
Monte Carlo Dropout - Prediction



Get T Classifications

$$Classifications = [\hat{y}_1 \ \hat{y}_2]$$

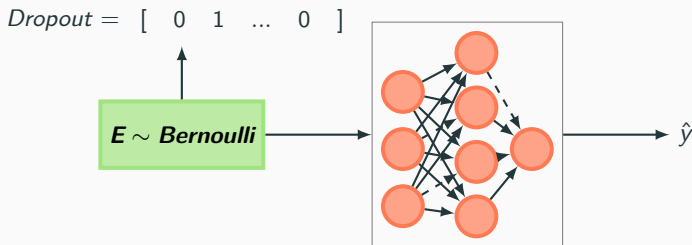
Monte Carlo Dropout - Prediction



Get T Classifications

$$Classifications = [\hat{y}_1 \ \hat{y}_2 \ \hat{y}_3]$$

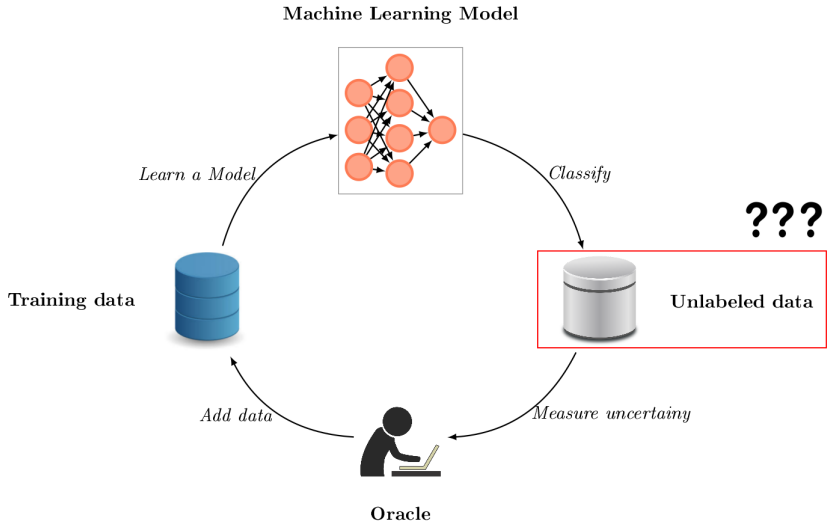
Monte Carlo Dropout - Prediction



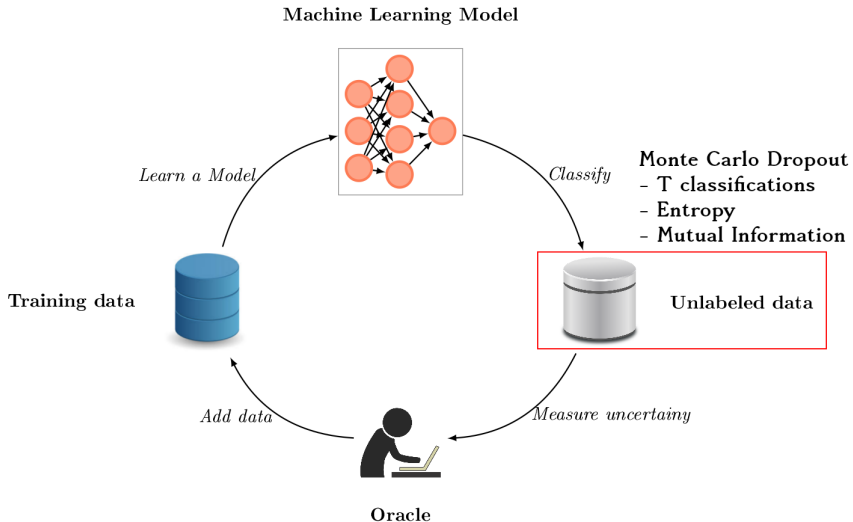
Get T Classifications

$$Classifications = [\hat{y}_1 \ \hat{y}_2 \ \hat{y}_3 \ \dots \ \hat{y}_T]$$

Active Learning



Active Learning



Experimental Design

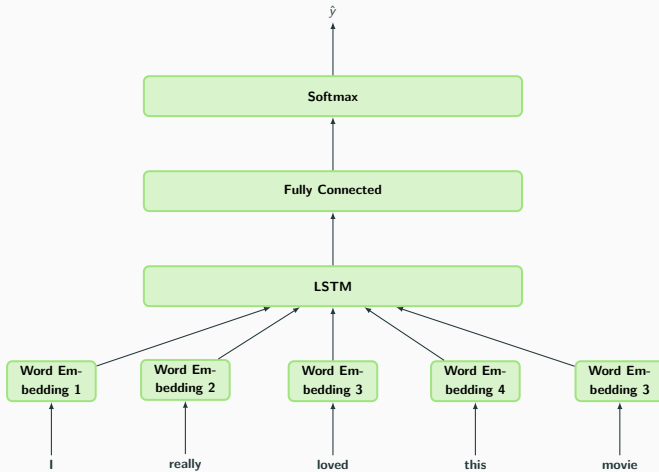
- Evaluation of the use of Active Learning together with Deep Learning for the task of sentiment analysis.
- Study of how Monte Carlo Dropout can provide uncertainty measurements for the LSTM Deep Learning architecture.
- Practical considerations when applying Deep Learning with Active Learning

Research Questions

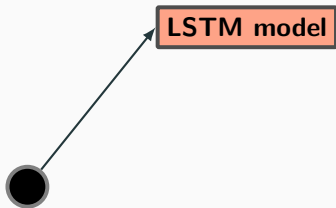
- **Q1:** On the task of sentiment analysis, does modelling the uncertainty measurement of the model using the Monte Carlo Dropout technique help us achieve a better accuracy value in the Active Learning context ?
- **Q2:** Does Monte Carlo Dropout provides best uncertainty measurements then using the soft- max output as a uncertainty measurement, the classical approach used in DL models ?

- Large Movie Review Dataset
 - 25000 train reviews and 25000 test reviews
 - Both train and test datasets have an equal number of positive and negative reviews
- Subjectivity Dataset
 - 10000 movie reviews divided into subjective and objective text.
 - Dataset perfectly balanced
 - Reviews have an average size of 20 words

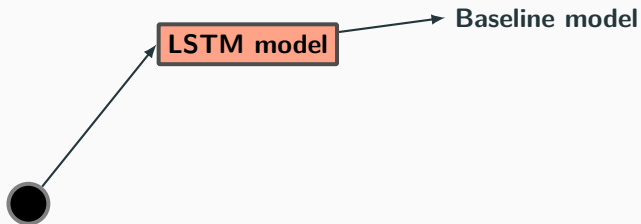
Network Architecture



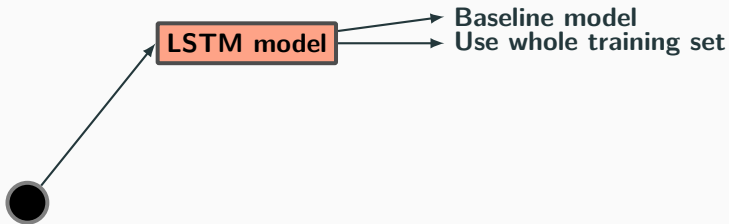




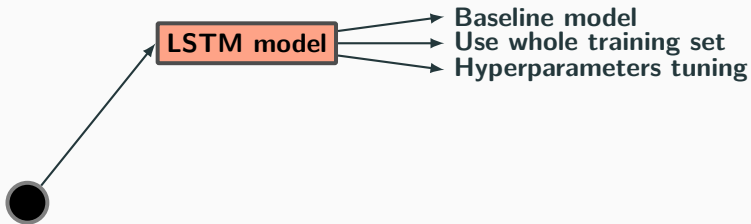
Experimental Design



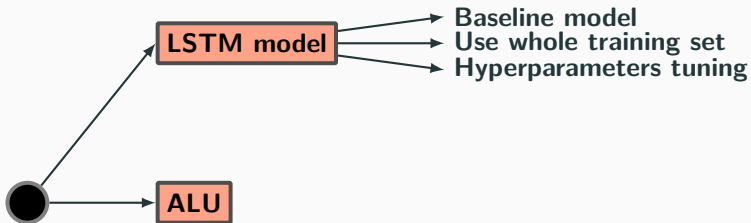
Experimental Design



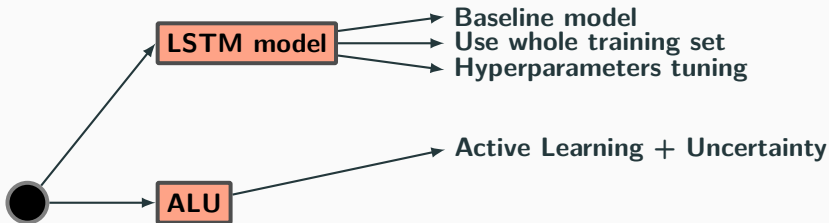
Experimental Design



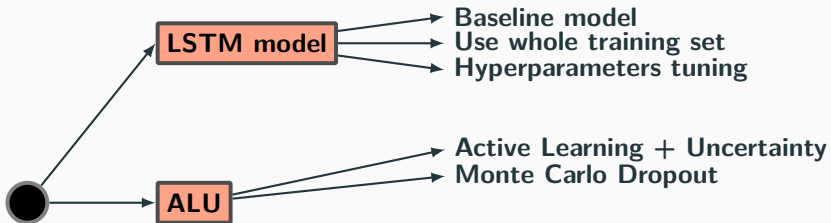
Experimental Design



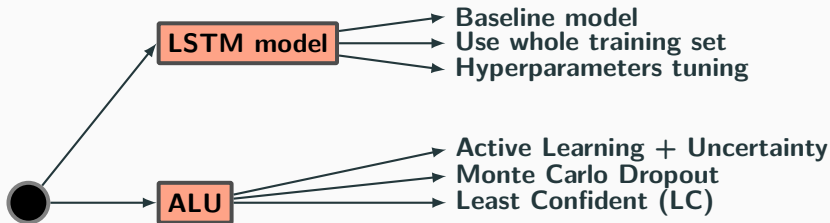
Experimental Design



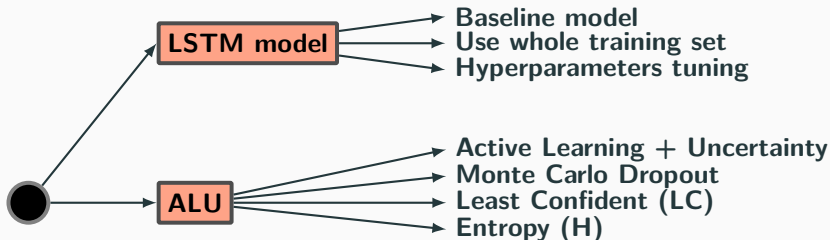
Experimental Design



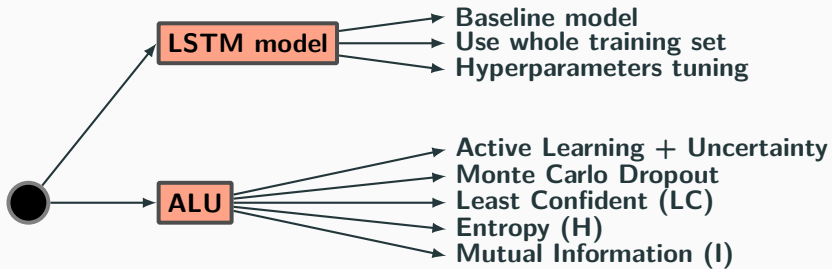
Experimental Design



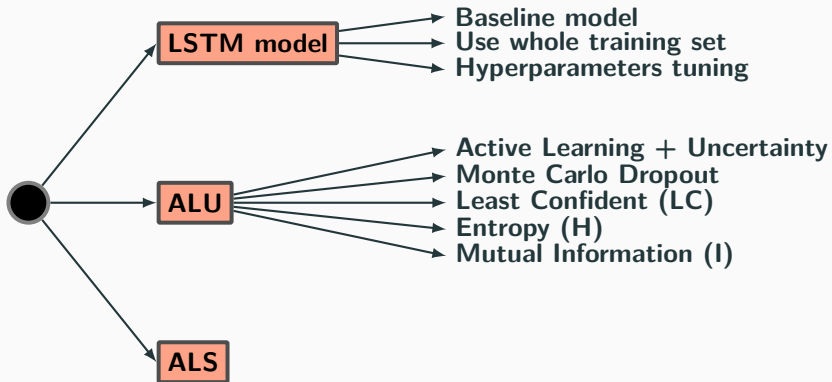
Experimental Design



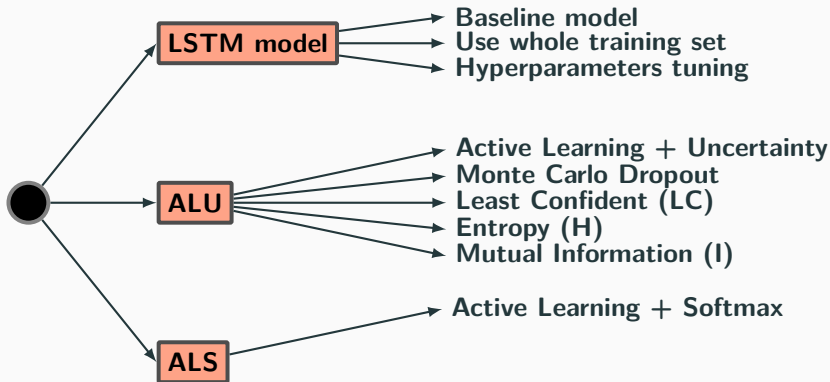
Experimental Design



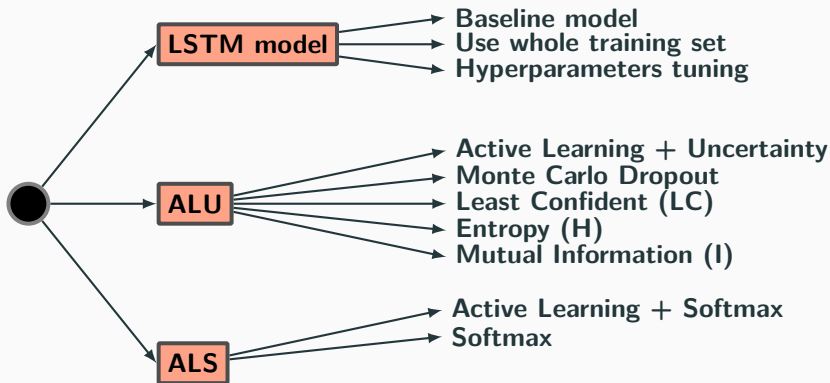
Experimental Design



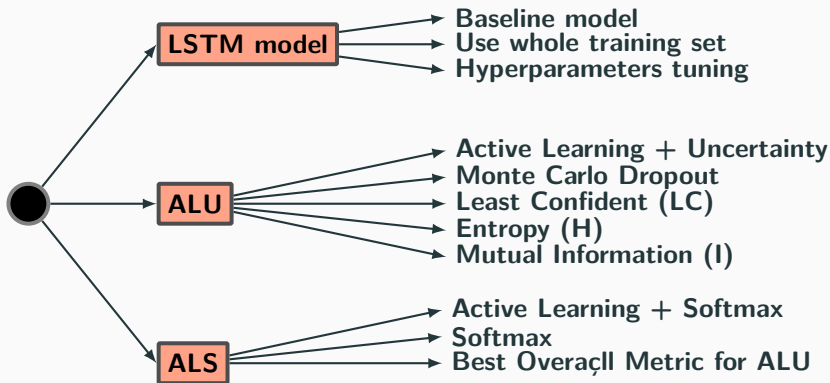
Experimental Design



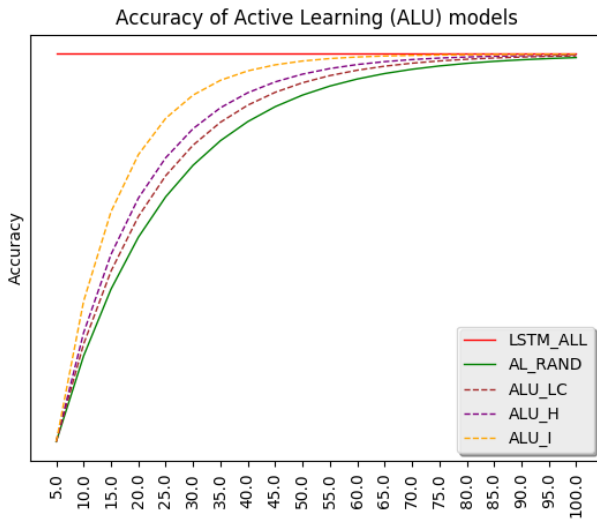
Experimental Design



Experimental Design



Active Learning



Active Learning Experiments Parameters

Dataset	Labeled Group	Unlabeled Group
Large Movie Review Dataset	225	22275
Subjectivity Dataset	10	8090

Active Learning Experiments Parameters

Parameter	Description
Unlabeled Data Queries (Q)	The number of example we will select from the unlabeled group to be labeled by the oracle.
Number of epochs (EPO)	At each AL cycle, we will train our model for a given number of epochs. This variable defines this quantity.
Dropout Values (DROP)	The dropout probability for the weights in our network.
Number of Active Learning Cycles (NC)	The number of AL cycles we have run for a given experiment.

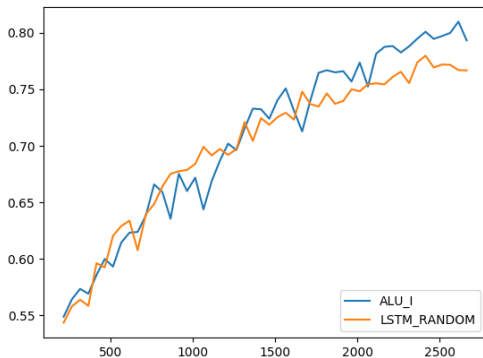
Experiments Evaluation

First Iteration

- Use Large Movie Review Dataset
- Train only a Random and ALS_I model (Mutual Information)
- Use Hyperparameters found with our baseline model
- Parameters
 - $Q = 50$
 - $EPO = 16$
 - $NC = 50$
- Random and ALS_I curve are almost identical

- Increase number of epochs in experiments

Second Iteration



• **Q** = 50

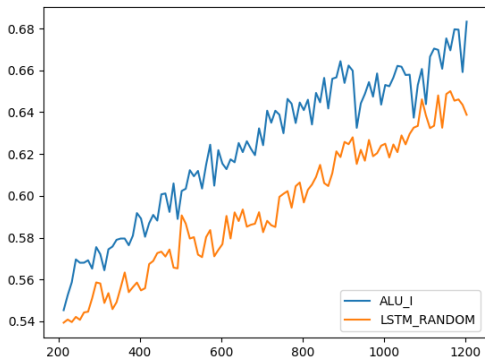
• **EPO** = 200

• **NC** = 50

Third Iteration

- Decrease number of unlabeled examples added to training dataset
- Decrease number of epochs

Third Iteration



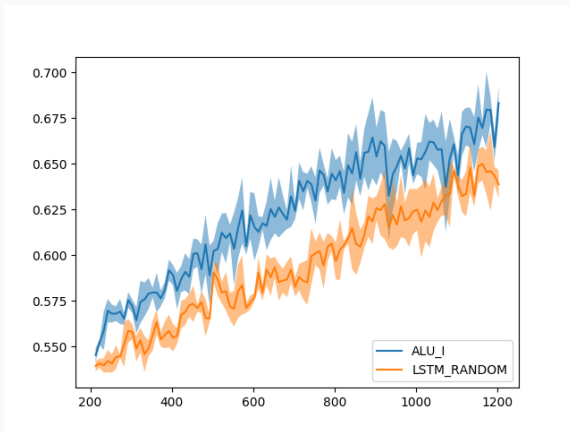
• **Q** = 10

• **EPO** = 150

• **NC** = 50

DROP = 0.5

Third Iteration



• **Q** = 10

• **EPO** = 150

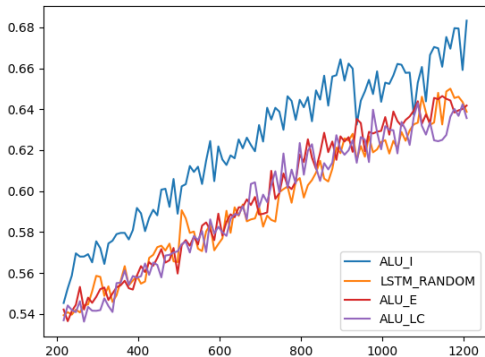
• **NC** = 50

DROP = 0.5

Fourth Iteration

- Compare all metrics
- Use CEAL approach

Fourth Iteration



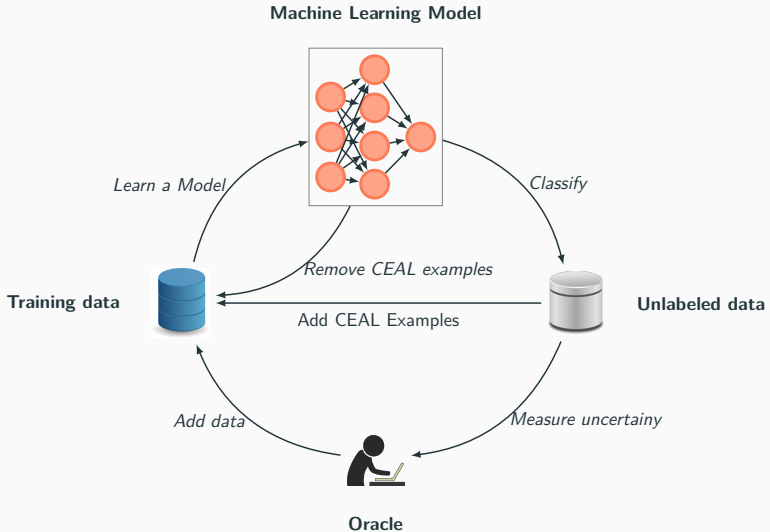
• **Q** = 10

• **EPO** = 150

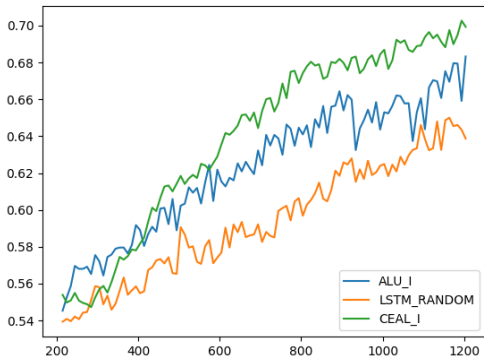
• **NC** = 100

DROP = 0.5

Fourth Iteration

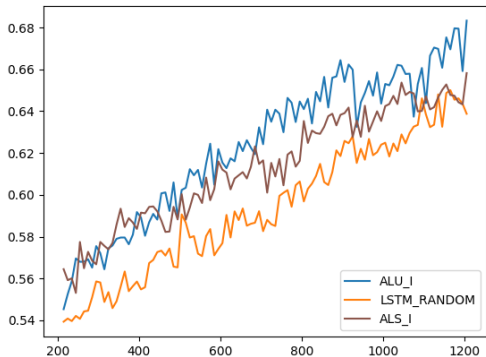


Fourth Iteration



- Compare softmax metric

Fifth Iteration



• **Q** = 10

• **EPO** = 150

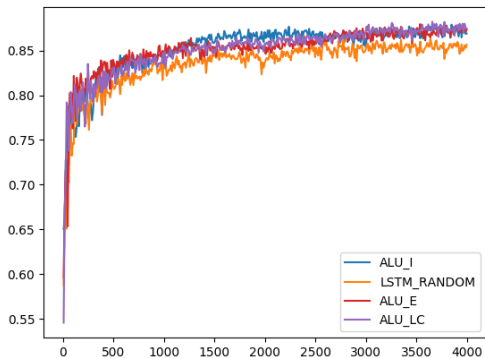
• **NC** = 100

DROP = 0.5

Sixth Iteration

- Use Subjectivity Dataset
- Compare all metrics
- Use CEAL approach
- Make softmax comparison

Sixth Iteration



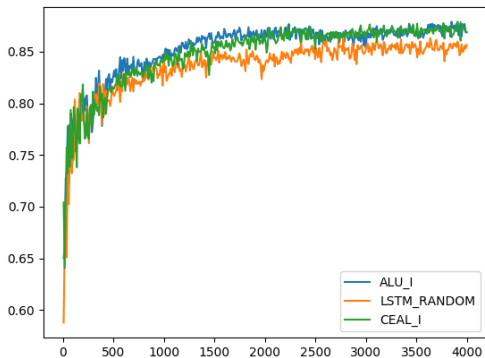
• **Q** = 10

• **EPO** = 150

• **NC** = 400

DROP = 0.5

Sixth Iteration



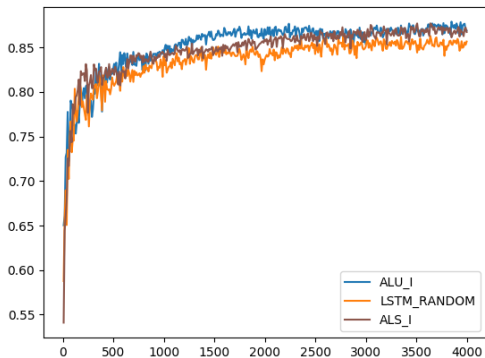
• **Q** = 10

• **EPO** = 150

• **NC** = 400

DROP = 0.5

Sixth Iteration



• **Q** = 10

• **EPO** = 150

• **NC** = 400

DROP = 0.5

Conclusion

Conclusion

- Measuring the uncertainty of sample using the Monte Carlo Dropout has created better accuracy curves than using both random and softmax.
- We have positive results for both of our research questions
- Not consistent result for both datasets

However

Active Learning Problems

- LSTM is a bad architecture for Active Learning
- Retraining Deep Learning algorithms each cycle does not scale
- Active Learning is biased approach
- Feel engineering approaches for Active Learning
- Costly framework and though to apply it to practical problems
- Huge number of variables to monitor

- Explore CEAL framework with Monte Carlo Dropout for other areas, such as image recognition
- Evaluate more Active Learning parameters
- Monitor the types of sentences the model is choosing from the unlabeled group
- Engineering work



A. Krizhevsky, I. Sutskever, and G. E. Hinton.

Imagenet classification with deep convolutional neural networks.

Commun. ACM, 60(6):84–90, May 2017.