Deep Active Learning using Monte Carlo Dropout

Lucas Moura
https://github.com/lucasmoura
lmoura@ime.usp.br

December 6, 2017

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

Introduction

Motivation

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

However...

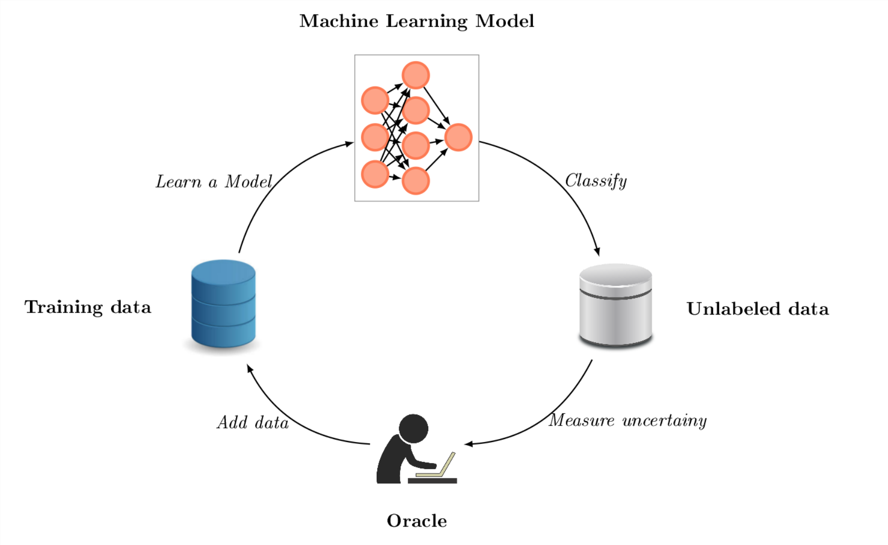
- Training Deep Learning models require a huge amount of labeled data
- For the task of image classification on the ImageNet database, 1.2 million labeled images were used [1]
- This restriction causes huge difficulties on applying Deep Learning techniques to a wide range of problems, such as Sentiment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

Sentiment Analysis

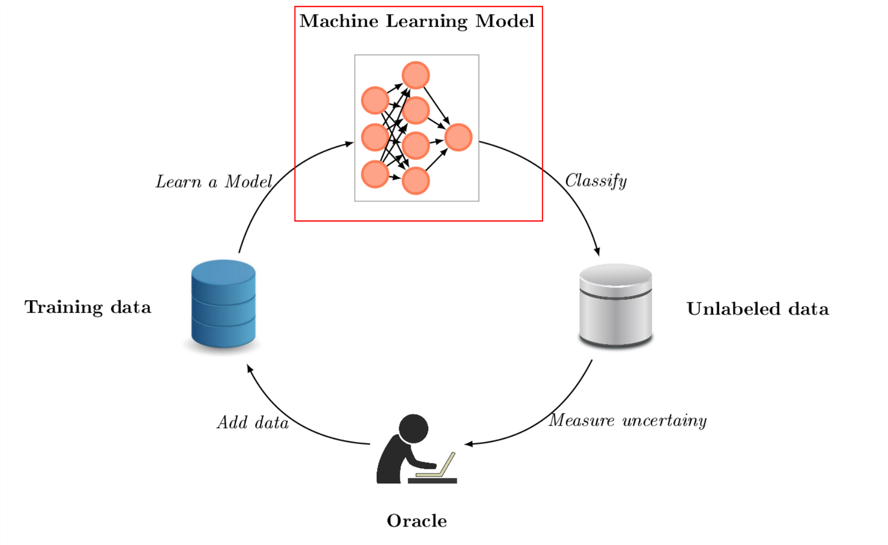
- Verify if a text is expressing negative or positive feelings.
- Huge amount of data, but few labeled.

Active Learning

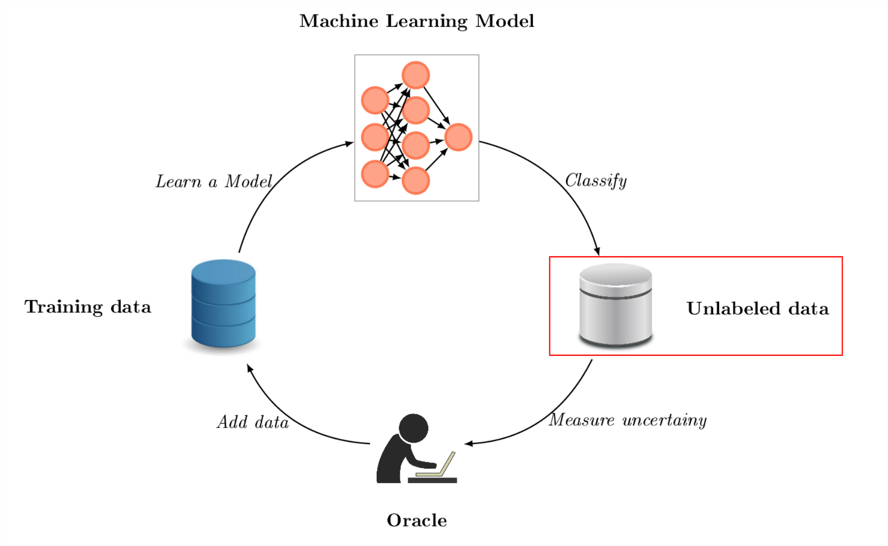
Active Learning



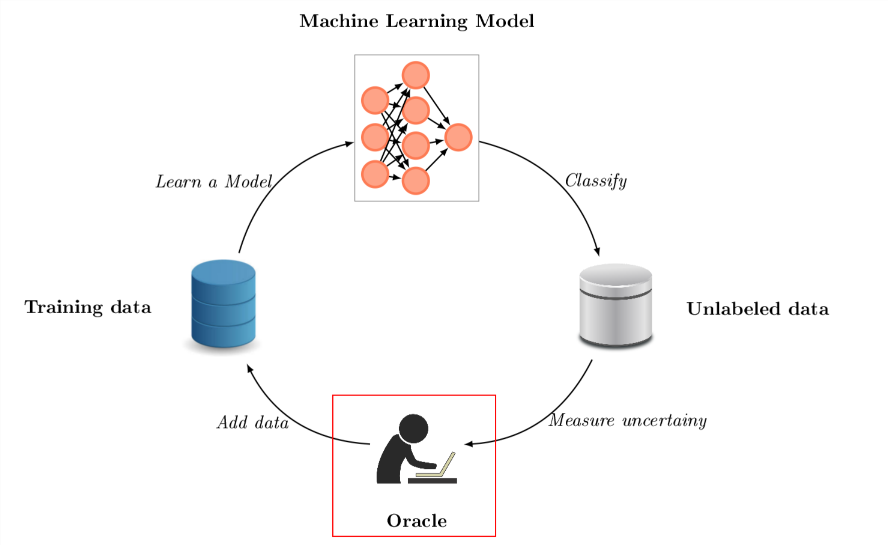
Active Learning



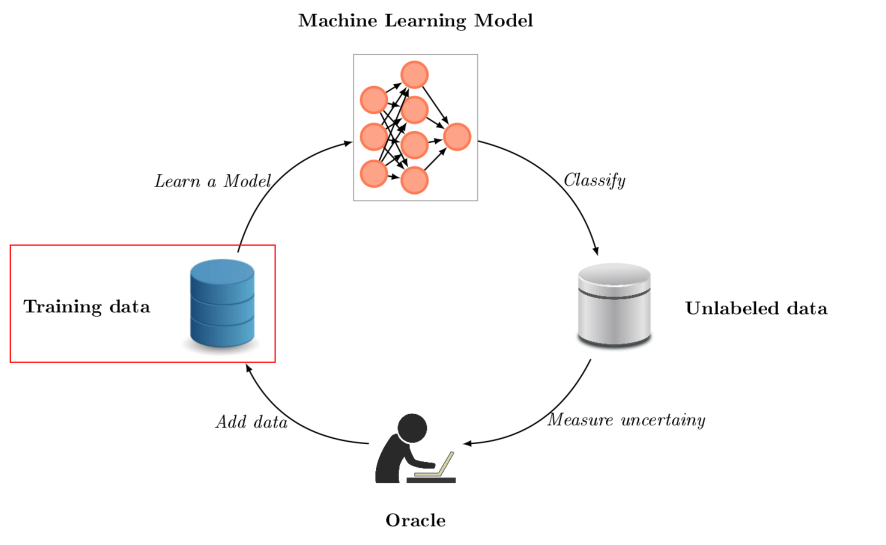
Active Learning



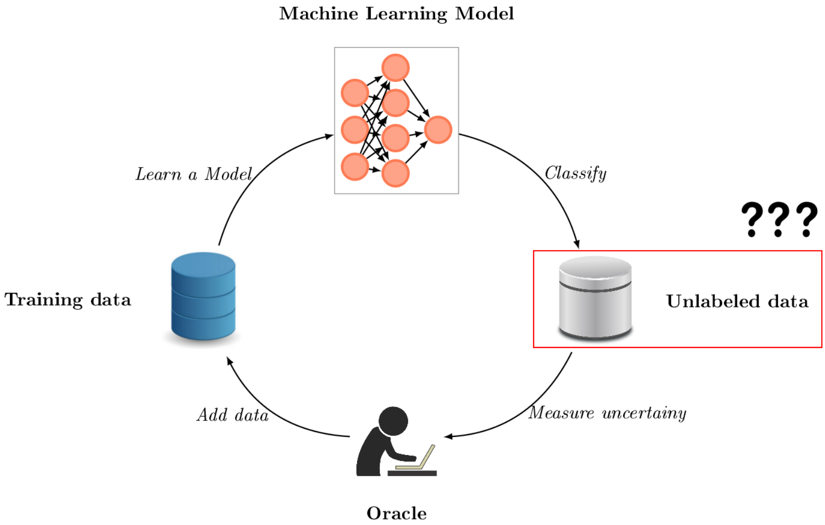
Active Learning



Active Learning



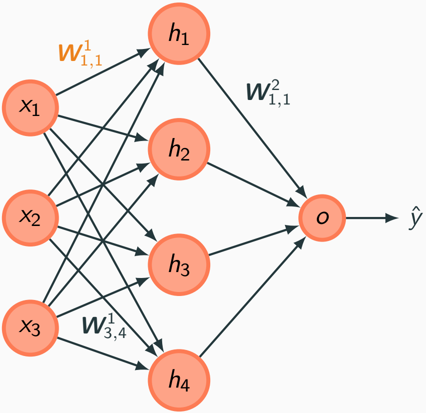
Active Learning



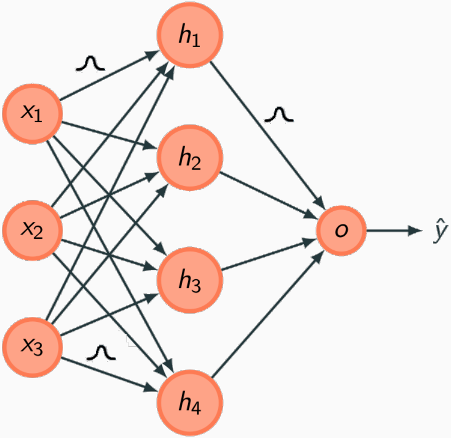
Uncertainty measurement

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

Neural Network



Bayesian Neural Network

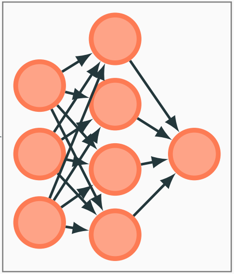


Bayesian Neural Network

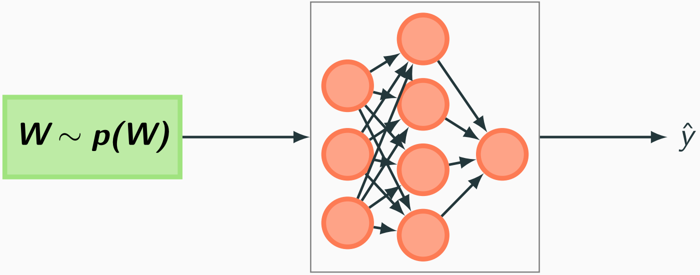
$$\boxed{W \sim p(W)}$$

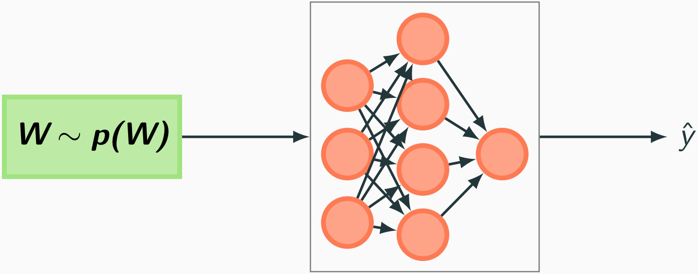
Bayesian Neural Network

$\boldsymbol{W \sim p(W)}$



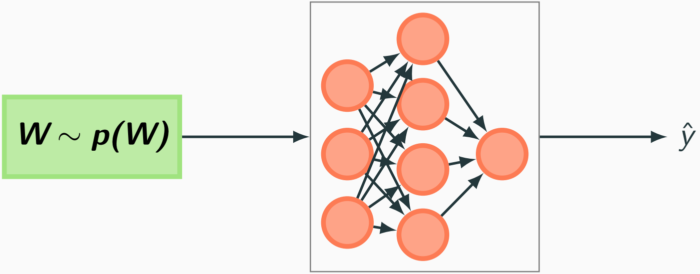
Bayesian Neural Network





Get T Classifications

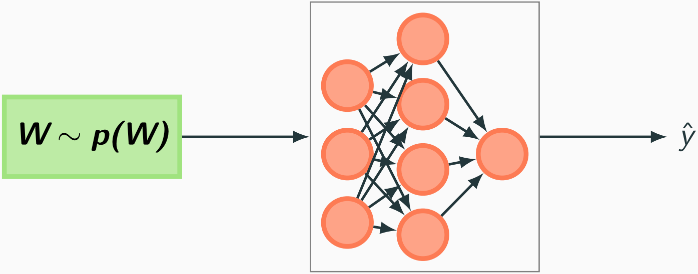
Bayesian Neural Network



Get T Classifications

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

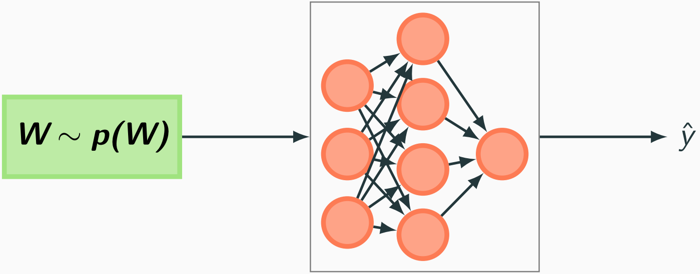
Bayesian Neural Network



Get T Classifications

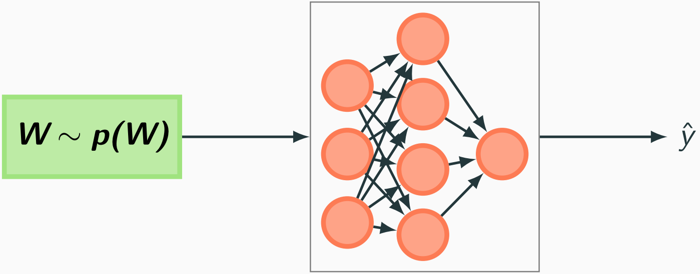
$$Classifications = [\quad \hat{y}_1 \quad \hat{y}_2 \quad]$$

Bayesian Neural Network



Get T Classifications

$$Classifications = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 & \hat{y}_3 \end{bmatrix}$$



Get T Classifications

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

Bayesian Neural Network

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

Bayesian Neural Network

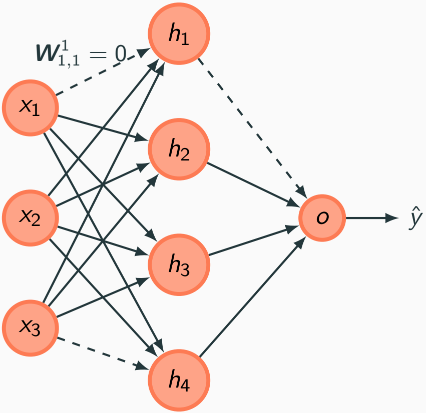
- What if we could extract uncertainty measurements from current Deep Learning models if they use stochastic regularization techniques such as Dropout ?
- Uncertainty in Deep Learning (Yarin Gal, 2017)

Dropout

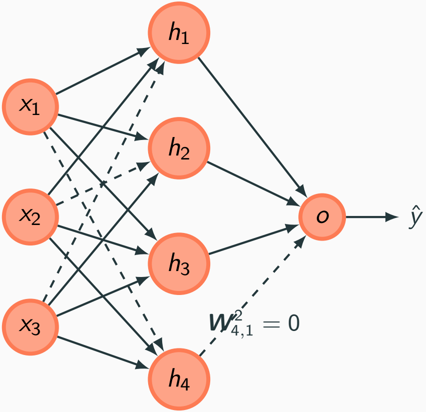
Dropout

- During training some weights are dropped from the network

Dropout



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

$\boldsymbol{E \sim \textit{Bernoulli}}$

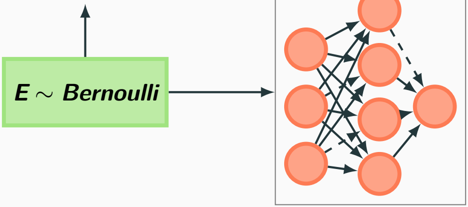
Monte Carlo Dropout

$$Dropout = \begin{bmatrix} 0 & 1 & \dots & 0 \end{bmatrix}$$

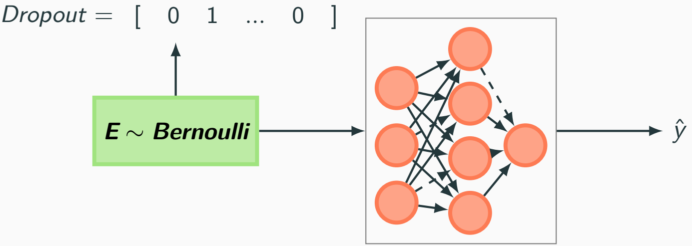
$$\boxed{\boldsymbol{E \sim Bernoulli}}$$

Monte Carlo Dropout

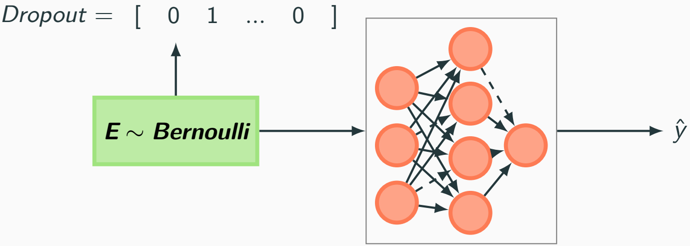
$$Dropout = \begin{bmatrix} 0 & 1 & \dots & 0 \end{bmatrix}$$



Monte Carlo Dropout



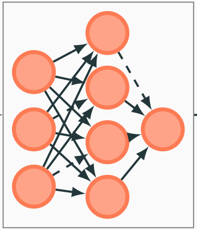
Monte Carlo Dropout



Get T Classifications

$$Dropout = [\quad 0 \quad 1 \quad \dots \quad 0 \quad]$$

$\boldsymbol{E \sim Bernoulli}$



$\hat{y}$

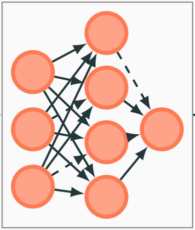
Get T Classifications

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

Monte Carlo Dropout

$$Dropout = \begin{bmatrix} 0 & 1 & \dots & 0 \end{bmatrix}$$

$\boldsymbol{E \sim Bernoulli}$



$\hat{y}$

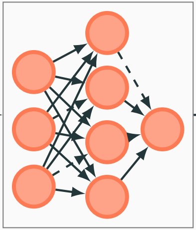
Get T Classifications

$$Classifications = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 \end{bmatrix}$$

Monte Carlo Dropout

$$Dropout = [\ 0 \quad 1 \quad \ldots \quad 0\ ]$$

$\boldsymbol{E \sim Bernoulli}$



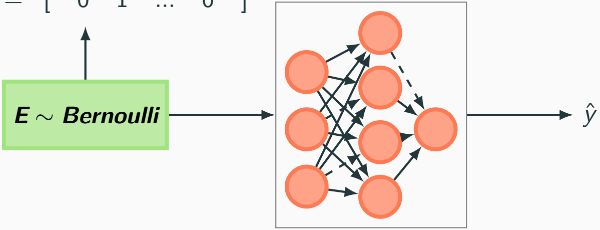
$\hat{y}$

Get T Classifications

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

Monte Carlo Dropout

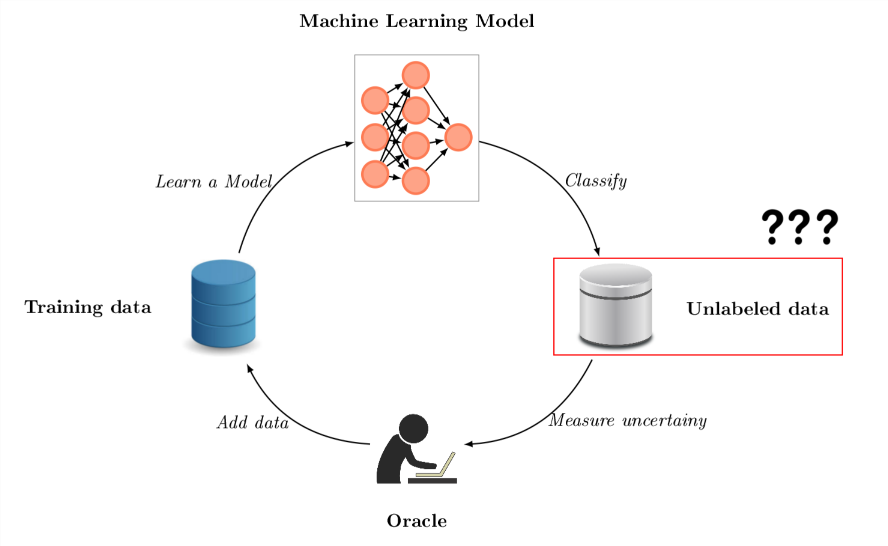
$$Dropout = \begin{bmatrix} 0 & 1 & \dots & 0 \end{bmatrix}$$



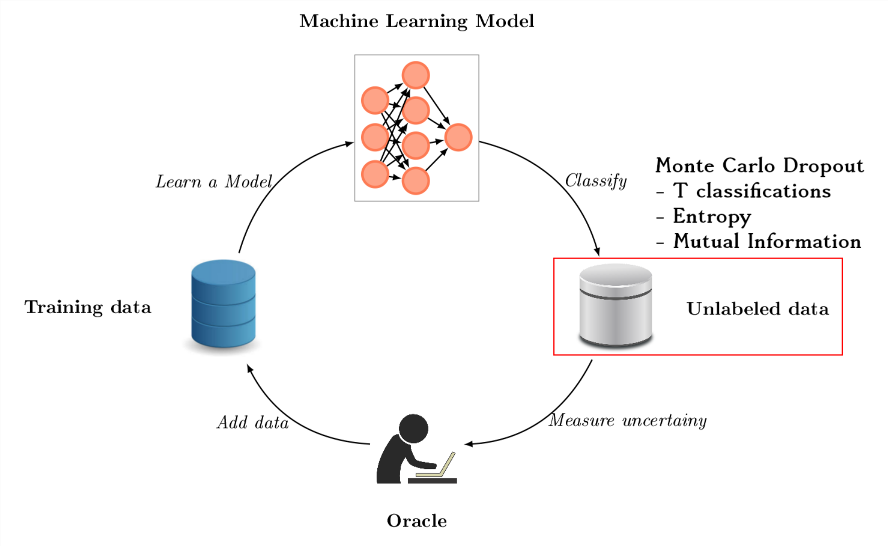
Get T Classifications

$$Classifications = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 & \hat{y}_3 & \dots & \hat{y}_T \end{bmatrix}$$

Active Learning



Active Learning



Experimental Design

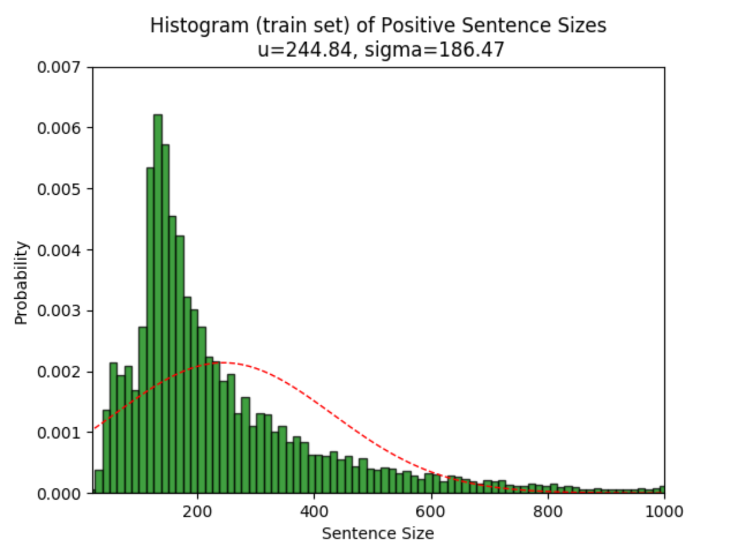
Research Questions

- Q1: On the task of sentiment analysis, can we achieve the same accuracy of a standard Deep Learning model by using Active Learning with uncertainty measurements, but with fewer labeled data ?
- Q2: Does modelling uncertainty in a Deep Learning model helps achieving a better result when using Active Learning ?

Dataset

- 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

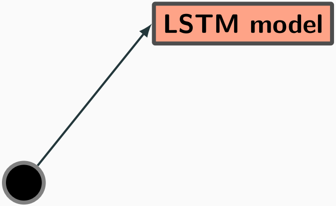
Dataset



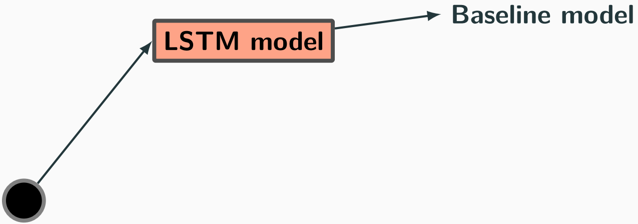
Histogram (train set) of Positive Sentence Sizes
u=244.84, sigma=186.47
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
Probability
200
400
600
800
1000
Sentence Size

Experimental Design

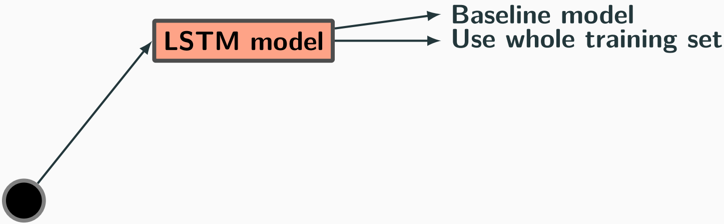
Experimental Design



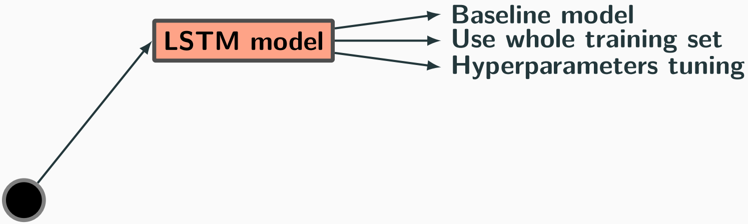
Experimental Design



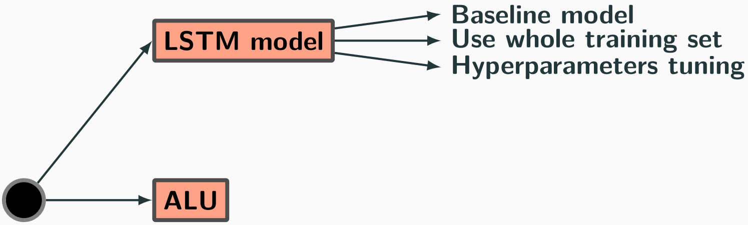
Experimental Design



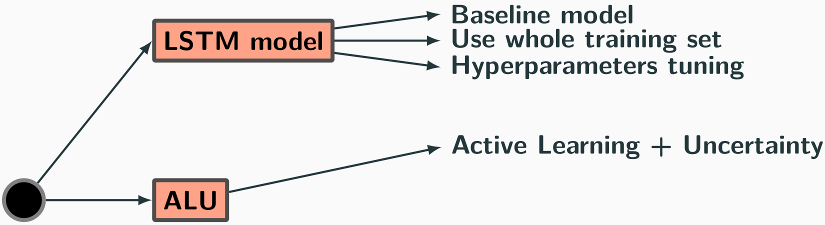
Experimental Design



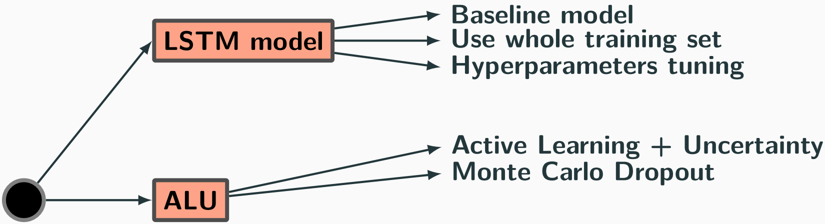
Experimental Design



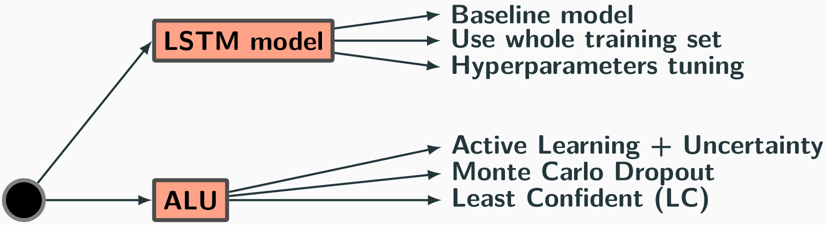
Experimental Design



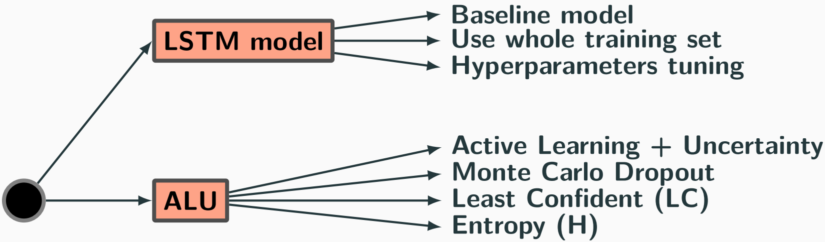
Experimental Design



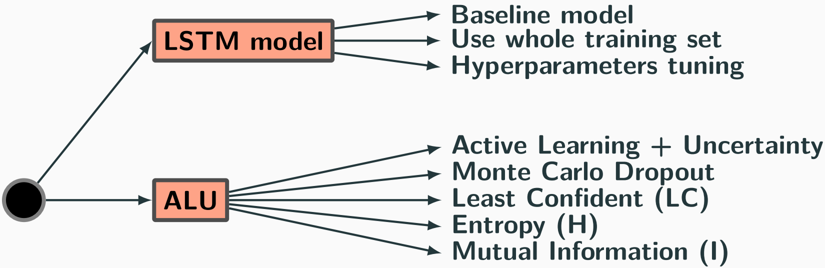
Experimental Design



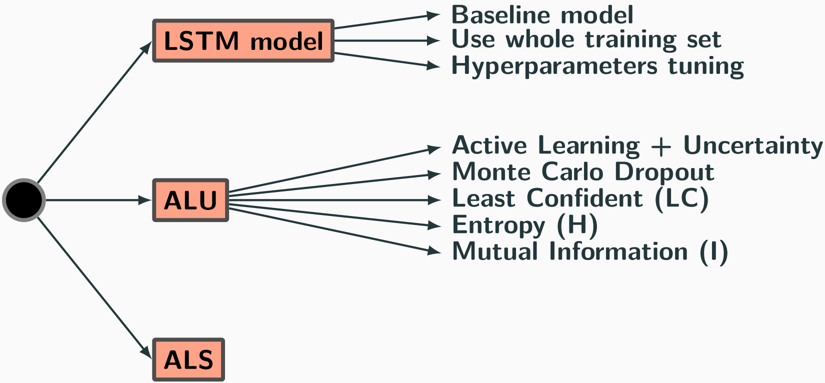
Experimental Design



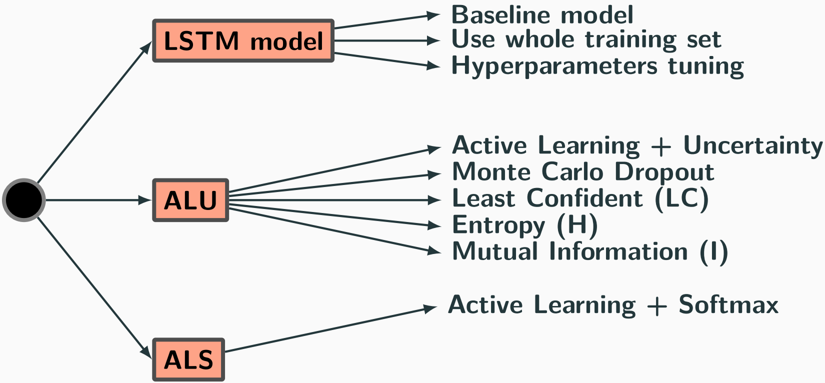
Experimental Design



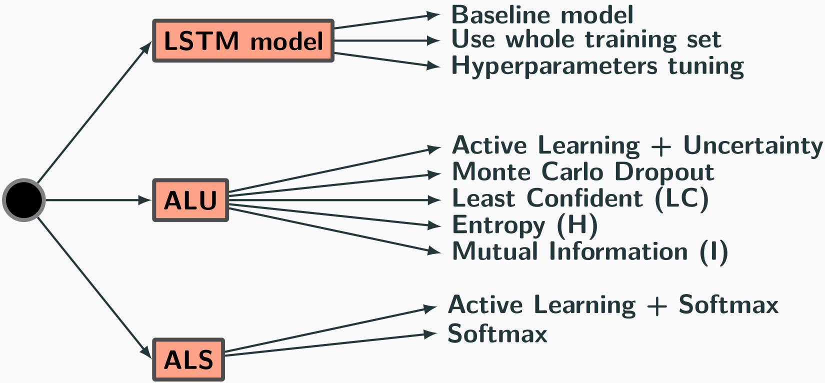
Experimental Design



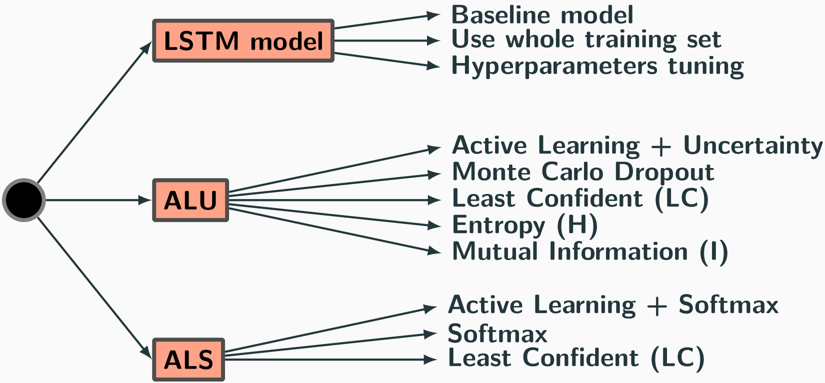
Experimental Design



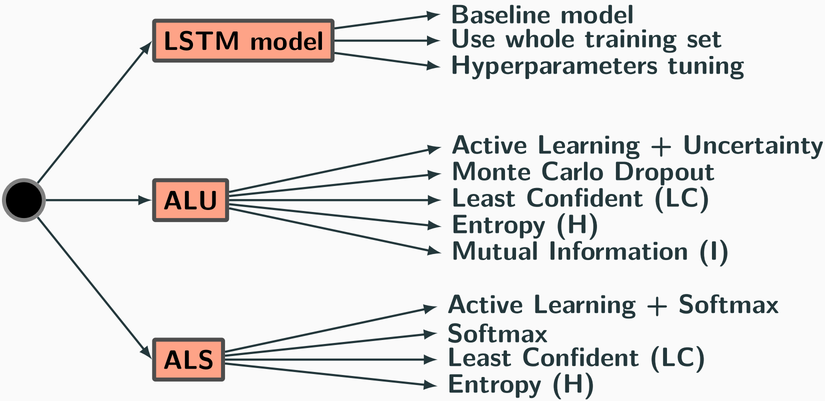
Experimental Design



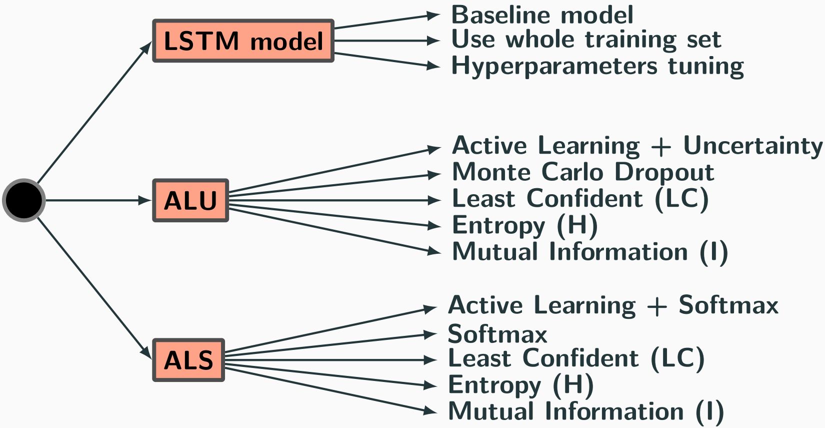
Experimental Design



Experimental Design



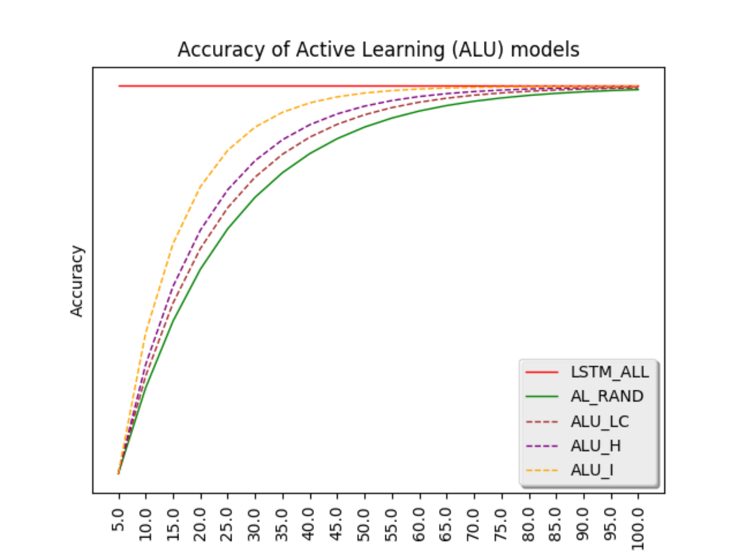
Experimental Design



Active Learning

- Q1: On the task of sentiment analysis, can we achieve the same accuracy of a standard Deep Learning model by using Active Learning with uncertainty measurements, but with fewer labeled data ?

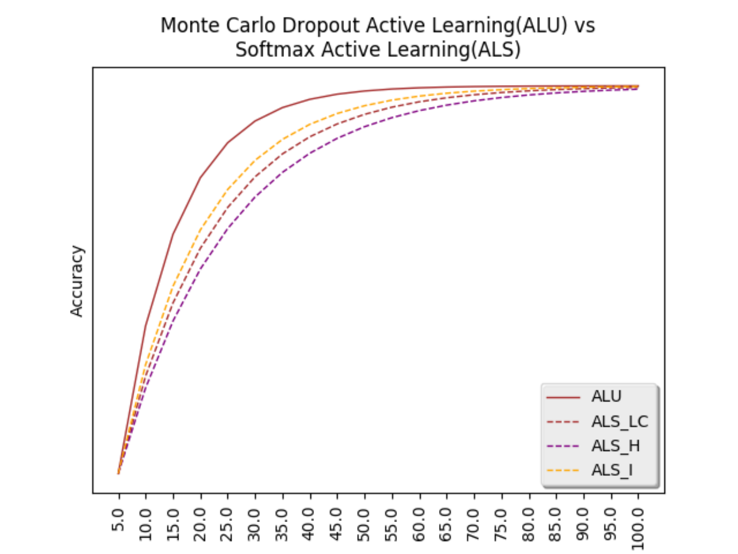
Active Learning



Active Learning

- Q2: Does modelling uncertainty in a Deep Learning model helps achieving a better result when using Active Learning ?

Active Learning



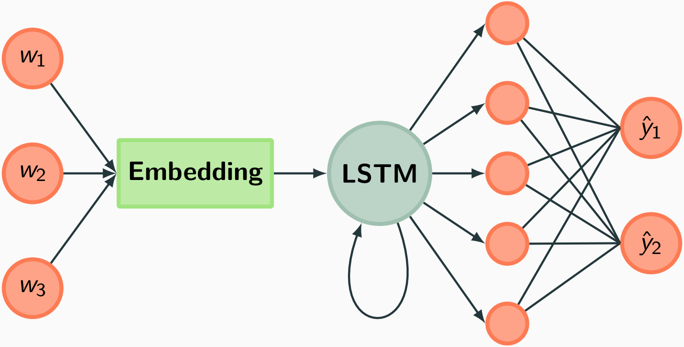
Active Learning

	2017		2018							
	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug
Implement LSTM model	x	x	x							
Implement **ALU** models				x	x					
Perform **ALU** experiments						x				
Implement **ALS** models							x			
Perform **ALS** experiments							x	x		
Update Models and Experiments								x	x	
Write thesis								x	x	
Masters defense										x

A. Krizhevsky, I. Sutskever, and G. E. Hinton.
Imagenet classification with deep convolutional neural networks.
Commun. ACM, 60(6):84–90, May 2017.

Backup Slides

Architecture



Active Learning