MapReduce for Neural Networks

...

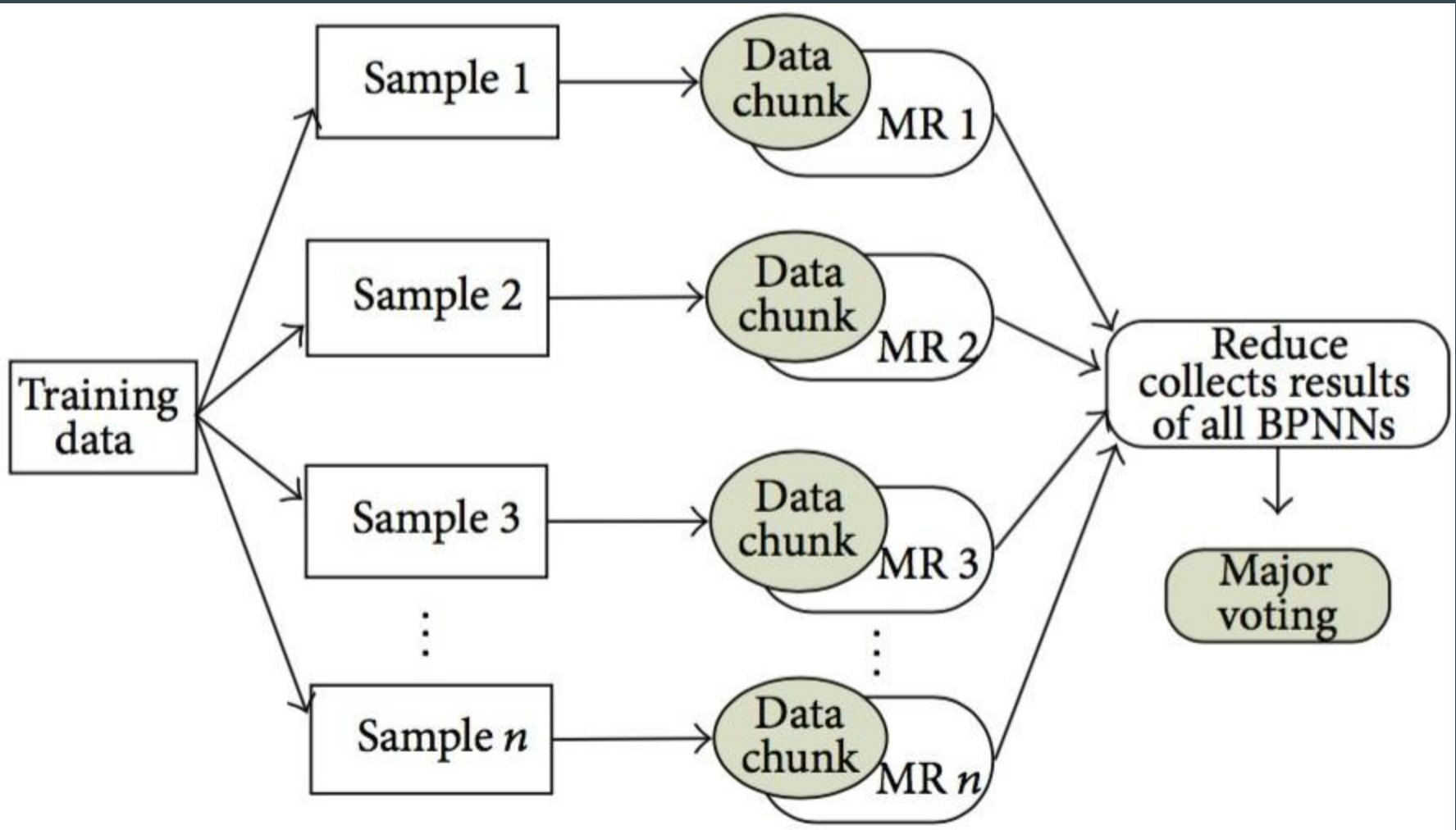
Olivier Dutfoy A20364492
Nicolas Lassaux A20365397

Problem Statement

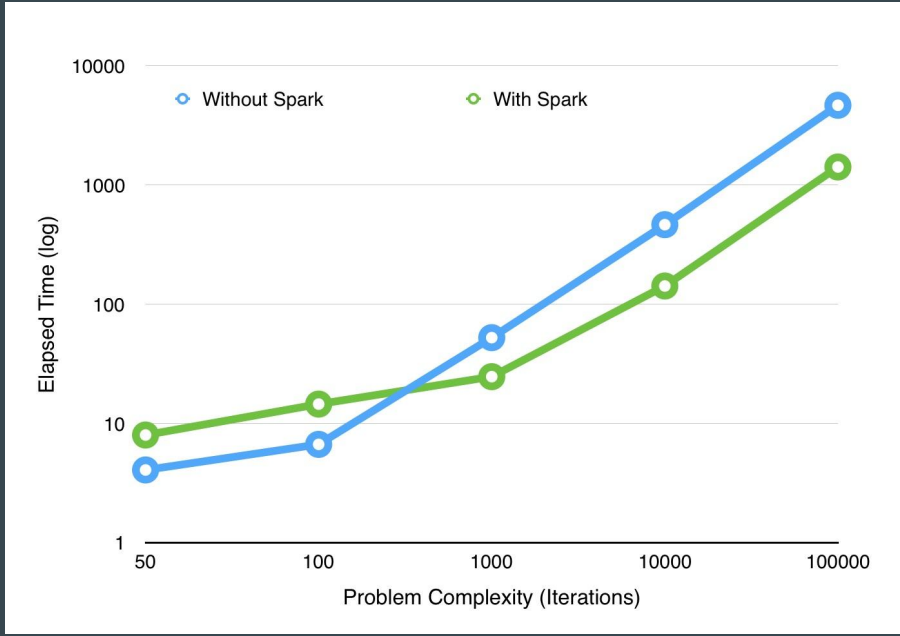
- Neural Networks are used quite frequently in Machine Learning.
- However due to today's large datasets, the computation time on a single machine may be significant.
- On the other hand MapReduce algorithms are fairly popular when considerable amount of resources are available.
- We therefore decided to Implement a MapReduce algorithm for Neural Networks.
- We used Apache Spark which allows cluster computing.
- While it can be used for external servers, it can also run locally using the multiple processors of a machine.

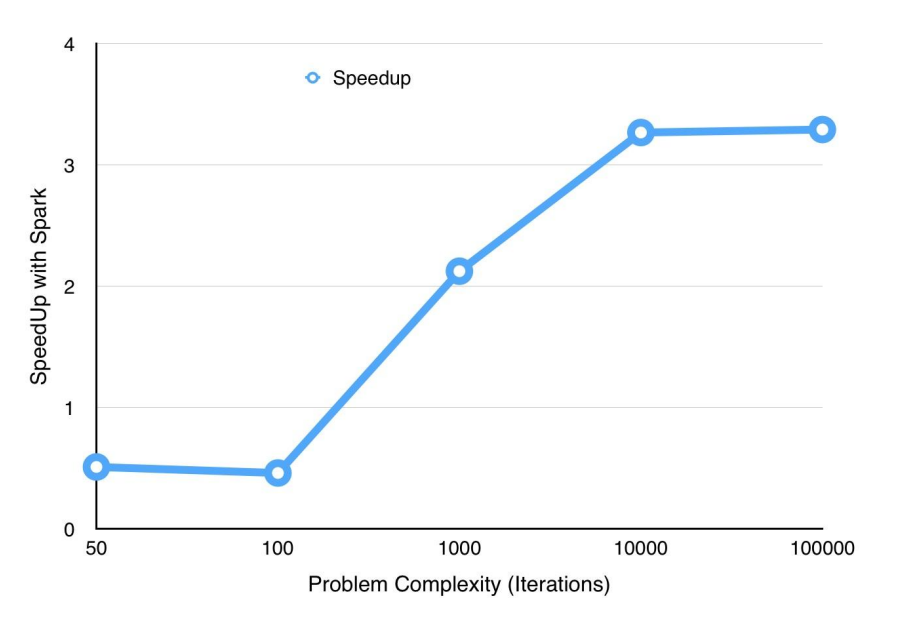
The Algorithm

- The paper we chose had multiple algorithms (one for large testing datasets, one for large training datasets and one for large neural networks.
- We implemented the second one :
  - Choose a number of workers (servers available for example) n
  - For each worker initialize a Neural Network
  - Sample the training data into n clusters
  - Feed each cluster to one different node to train the network
  - Classify each testing instance using each of the n models
  - The final classification is chosen using the majority amongst all models


Training data
Sample 1
Sample 2
Sample 3
Sample n
Data chunk
MR 1
Data chunk
MR 2
Data chunk
MR 3
Data chunk
MR n
Reduce collects results of all BPNNs
Major voting

Results





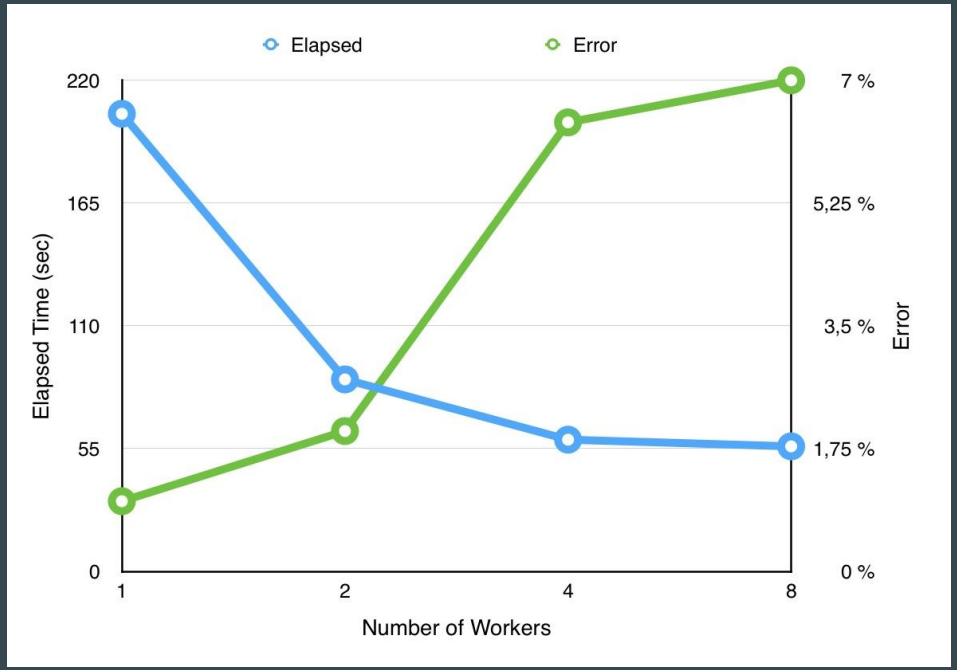
Spark Configuration: 8 workers (locally), on Iris dataset

Results

Accuracy	Iris	Breast Cancer
Spark	76.7%	93.6%
Classic Neural Networks	91.3%	99.0%

Spark Configuration: 4 workers (locally)

Results



On Cancer Detection Dataset.

References and Sources

Main Paper :

Yang Liu, Jie Yang, Yuan Huang, Lixiong Xu, Siguang Li, and Man Qi, MapReduce Based Parallel Neural Networks in Enabling Large Scale Machine Learning, in Computational Intelligence and Neuroscience, August 2015, Volume 2015

DataSets :

https://archive.ics.uci.edu/ml/datasets/Iris

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29