IIT Stuart School of Business
ILLINOIS INSTITUTE OF TECHNOLOGY

Deep Neural Networks for Market Prediction on the Xeon Phi*

Matthew Dixon[1], Diego Klabjan[2] and Jin Hoon Bang[3]

* The support of Intel Corporation is gratefully acknowledged

1 Department of Finance, Stuart School of Business, Illinois Institute of Technology
2 Department of Industrial Engineering and Management Sciences, Northwestern University
3 Department of Computer Science, Northwestern University

Overview

- Background on engineering automated trading decisions
- Early machine learning research and what's changed since then
- Overview of deep learning
- Training and backtesting on advanced computer architectures

Algo Trading Strategies

IIT Stuart School of Business
ILLINOIS INSTITUTE OF TECHNOLOGY

Market data

Rules

Trading decision

Symbol	Pct	Last	Change	Prev	Bid	Ask	High	Low	EVol	Volume
ES #F	+0.11%	136400	+150	136250	136400	136425	136625	134250	1084656	2076198
SP #F	+0.13%	136430	+180	136250			136620	134210	2111	10585
$SPX	-0.05%	1368.39	-0.71	1369.10			1370.08	1363.94		
NQ #F	+0.25%	263200	+650	262550	263175	263200	264150	258650	160303	320105
$NDX D	+0.01%	2638.09	+ 0.17	2637.92			2644.37	2620.96		
$COMP D	-0.04%	2955.09	-1.25	2956.34			2960.67	2939.21		
YM #F	-0.08%	1294600	-1100	1295700	1294500	294600	1296200	1277900	78269	139189
DJ M2	-0.02%	1295500	-200	1295700	1295000		1296000	1291000	0	131
DJ #F	-0.02%	1295500	-200	1295700	1295000		1296000	1291000	0	131
ES U2	+0.17%	135825	+225	135600	135750	135800	135975	133750	908	1679
$INDU D	-0.27%	13003.04	-35.23	13038.27			13036.15	12970.00		43960296
TF #F	+0.23%	78840	+180	78660	78830	78840	79140	77280	66434	153710
$RUT D	-0.06%	791.34	-0.50	791.84			792.67	788.07		
MC #F	+0.29%	96330	+280	96050	96320	96340	96630	94660	15015	26228
NK #F	-1.13%	9155	-105	9260			9155	9155	0	3315
$TRAN D	+0.57%	5257.48	+29.84	5227.64			5257.62	5198.00		4747289
ZB #F	+0.04%	14324	+2	14322	14324	14325	14419	14321	142139	323844
US M2	+0.07%	14325	+3	14322			14325	14325	0	323844
$TYX	-0.20%	3.065	-0.006	3.071			3.070	3.052		
ZN M2	+0.01%	132210	+5	132205	132210	132215	133020	132195	484043	1299349
TY M2	+0.34%	132205	+145	132060			132210	132210	0	1299349
$TNX	-0.37%	1.873	-0.007	1.880			1.880	1.859		
ZF M2	+0.02%	123307	+7	123300	123307	123310	124032	123297	170942	405572
ZT M2	-0.01%	110087	-2	110090	110087	110090	110092	110082	59470	122419
6E M2	-0.28%	13052	-37	13089	13051	13052	13062	12957	190719	237586
6S M2	-0.30%	10868	-33	10901	10867	10868	10877	10778	35792	39567
6B M2	+0.26%	16183	+42	16141	16182	16184	16193	16110	64892	99435
6J M2	+0.01%	12526	+1	12525	12525	12526	12564	12504	48836	78990
6A M2	+0.20%	10160	+20	10140	10159	10160	10165	10065	83072	160266
6C M2	+0.10%	10048	+10	10038	10047	10048	10060	10002	60345	103796
6N M2	+0.05%	7934	+4	7930	7934	7936	7945	7886	11427	18113
DX M2	+0.15%	79705	+118	79587	79705	79710	80135	79660	17688	24757

Mathematical operations

Input

Output

Configurations

Example

Buy signal

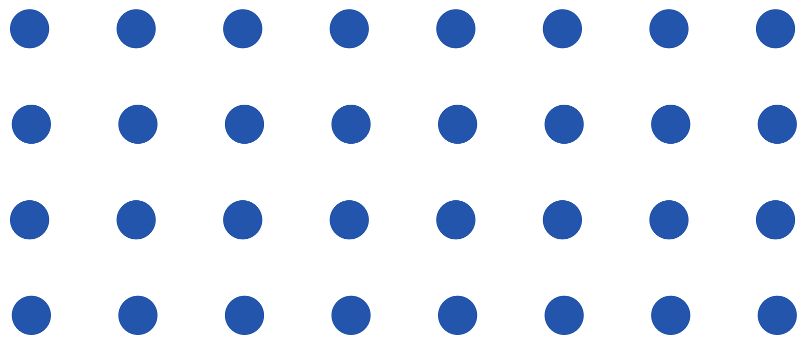
Sell signal

Markets

- Futures market
  - CME 50 liquid futures
  - Other exchanges
- Equity markets
- FX markets
  - 10 major currency pairs
  - 30 alternative currency pairs
- Options markets

Strategy Universe

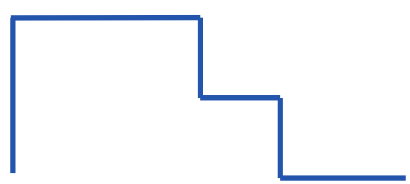
Strategy configurations



Configuration c

Trading decision s(c)

Utility U(s(c))



P&L / drawdown

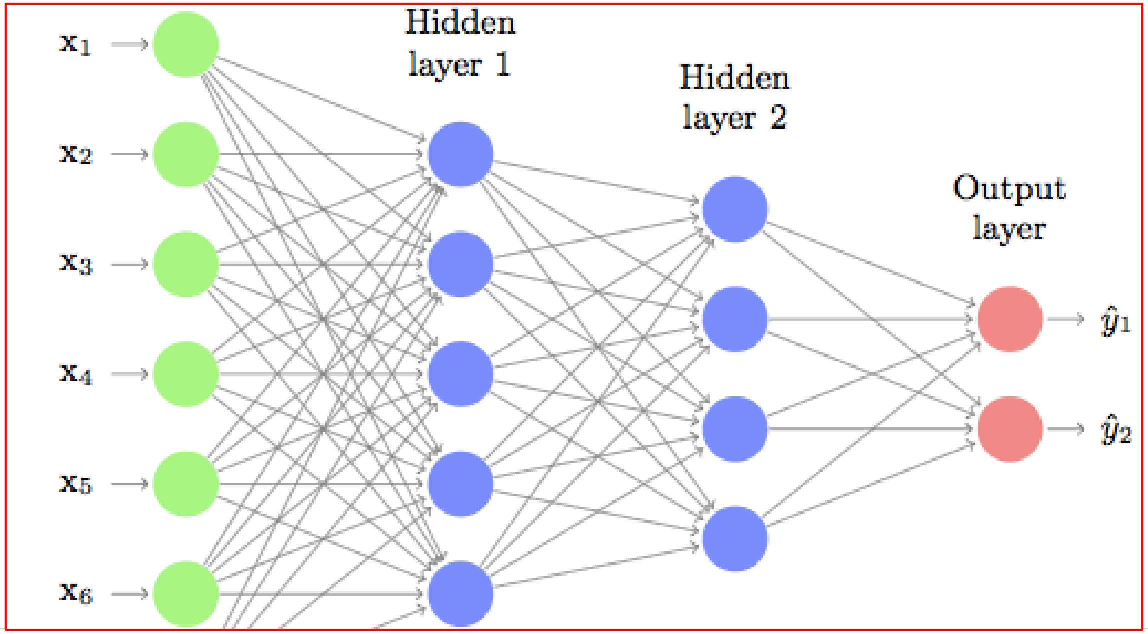
How can we engineer a strategy producing buy / sell decisions ?

Finance Research Literature

- J. Chen, J. F. Diaz, and Y. F. Huang. High technology ETF forecasting: Application of Grey Relational Analysis and Artificial Neural Networks. Frontiers in Finance and Economics, 10(2):129–155, 2013.
- S. Niaki and S. Hoseinzade. Forecasting S&P 500 index using artificial neural networks and design of experiments. Journal of Industrial Engineering International, 9(1):1, 2013.
- M. T. Leung, H. Daouk, and A.-S. Chen. Forecasting stock indices: a comparison of classification and level estimation models. International Journal of Forecasting, 16(2):173–190, 2000.
- A. N. Refenes, A. Zapranis, and G. Francis. Stock performance modeling using neural networks: A comparative study with regression models. Neural Networks, 7(2):375 – 388, 1994.
- R. R. Trippi and D. DeSieno. Trading equity index futures with a neural network. The Journal of Portfolio Management, 19(1):27–33, 1992.

Deep Neural Networks Literature

- Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
- R. Rojas. Neural Networks: A Systematic Introduction. Springer-Verlag New York, Inc., New York, NY, USA, 1996.



Classifiers

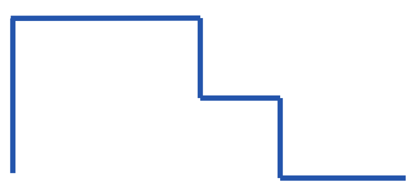
Strategy configuration c

Features

Classifier

E.g.
-Historic price returns at various lags
-Moving averages
-Oscillators

Trading decision s(c)



1 **buy**
0 hold
-1 sell

What's changed since the 90's?

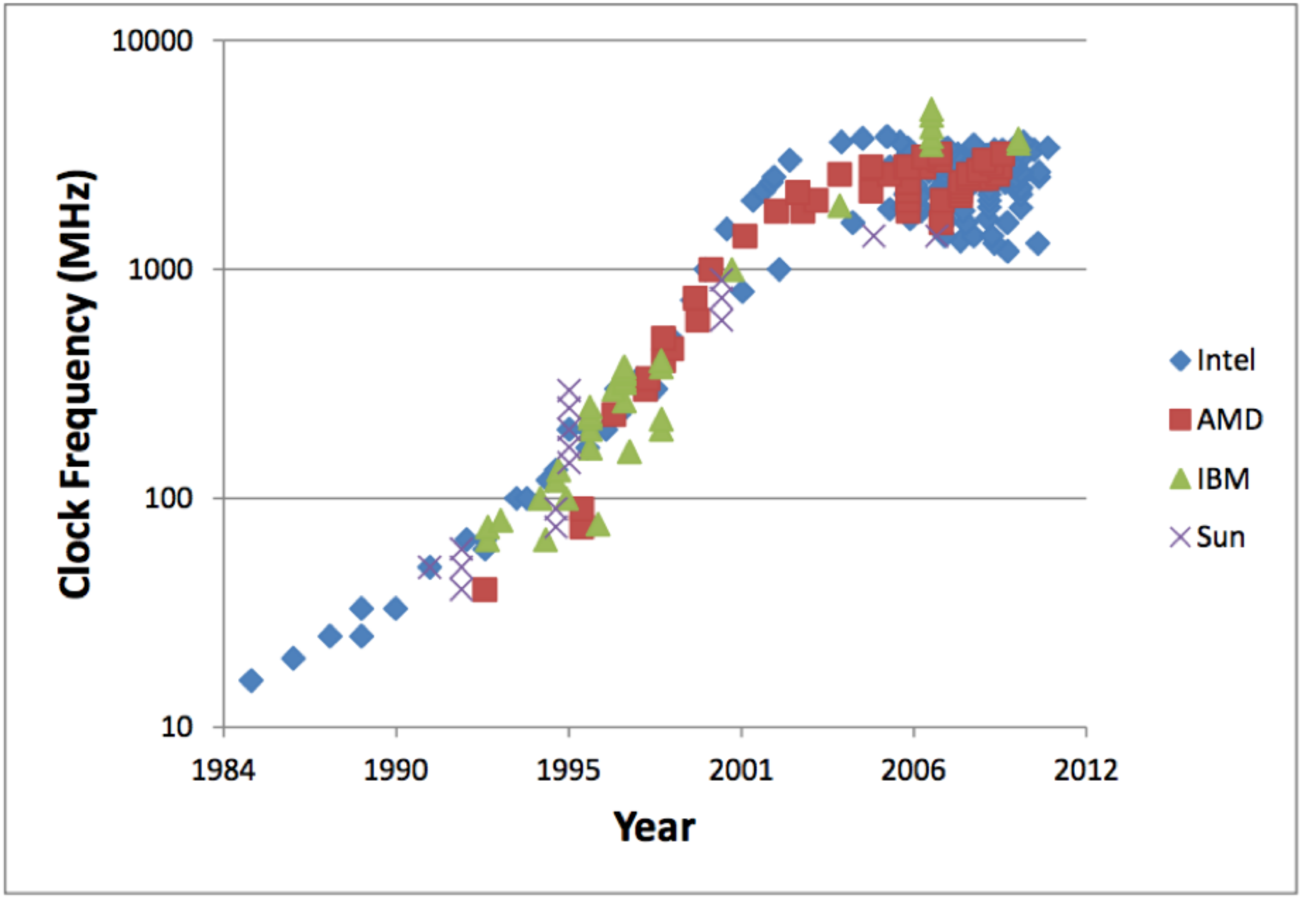
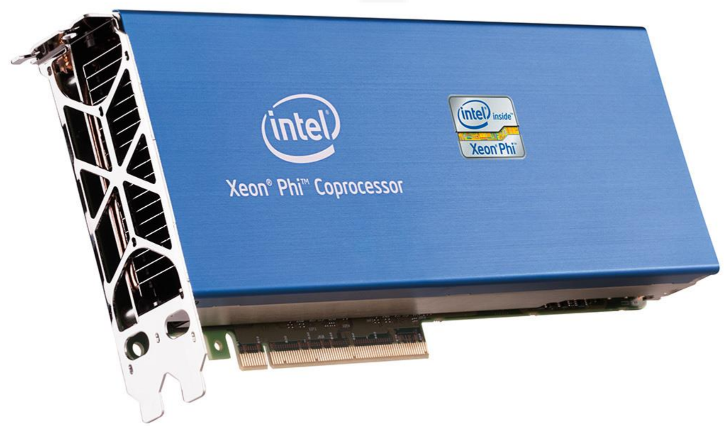


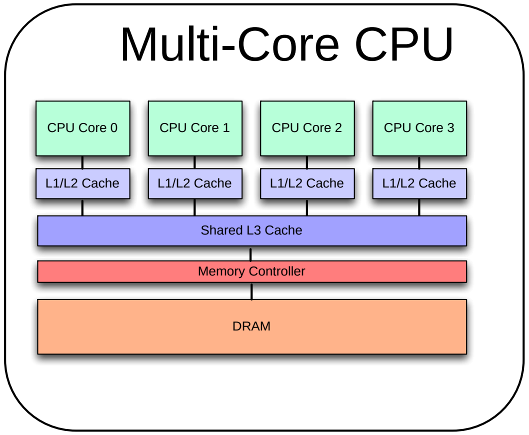
Figure 1.1: Scaling of the processor clock speeds.

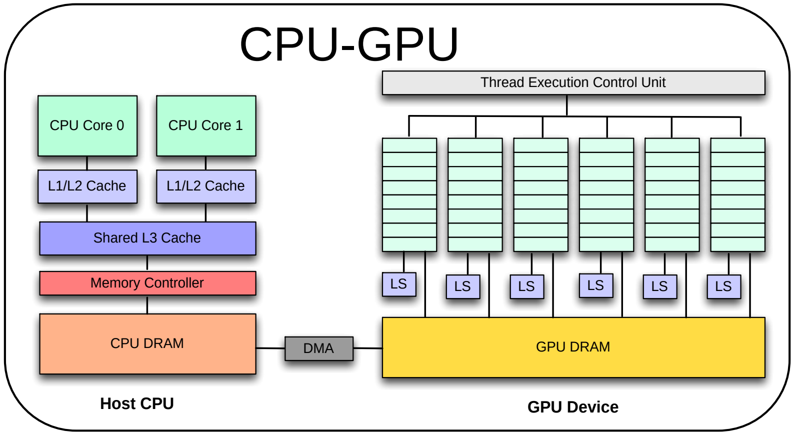
Computer architecture transformation in half a century

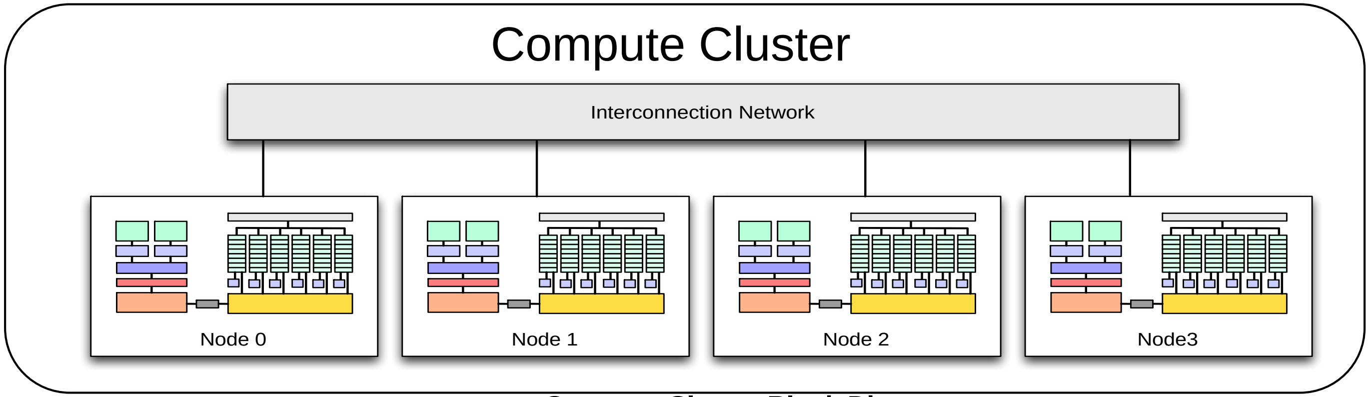
RAND computing facilities in the 1950s-60s



Variety of Parallel Architectures

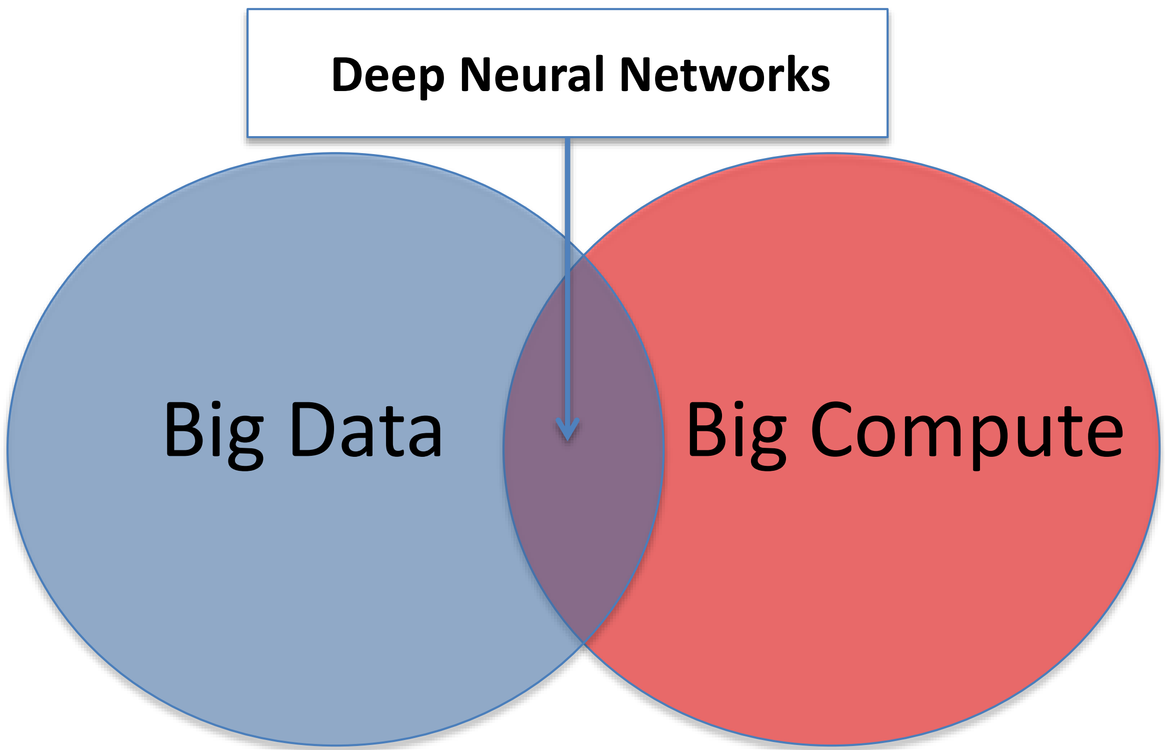






Compute Cluster Block Diagram

Big Data <-> Big Compute



Feature Engineering

Raw features

5 minute mid-prices for 45 CME listed commodity and FX futures over the last 15 years

Label

-Lagged price differences from 1 to 100
-moving price averages with window size from 5 to 100
-pair-wise correlation of returns

Labels from positive, neutral or negative market returns

Engineered features → Normalized features → Labeled feature set

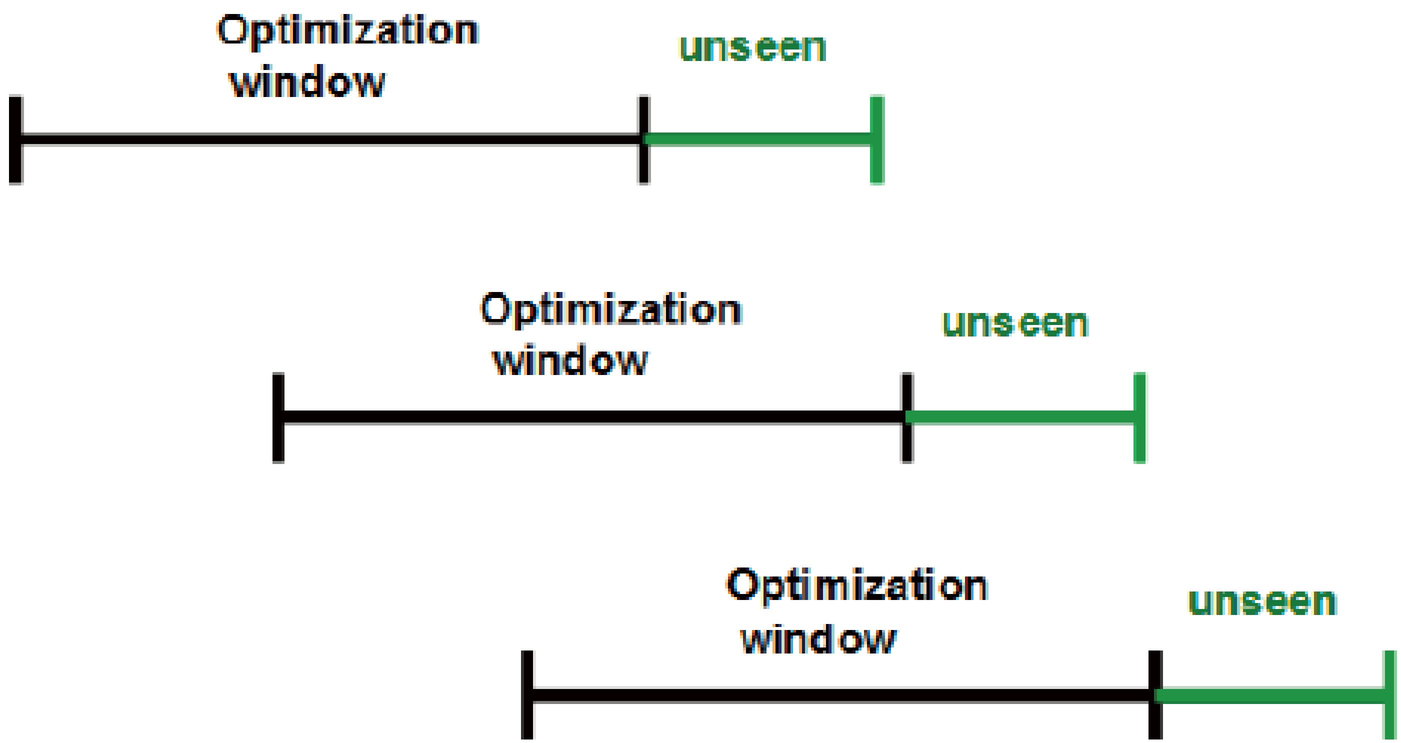
Final feature set consists of 9895 features and 50,000 consecutive observations

Deep Learning Algorithm

Algorithm 2 DEEP LEARNING METHODOLOGY

for $\gamma := 0.1, 0.2, \ldots, 1$ **do**
<Initialize all weights>
$w_{i,j}^{(l)} \leftarrow r, \; r \in \mathcal{N}(\mu, \sigma), \; \forall i, j, l$
<Iterate over epochs>
for $e = 1, \ldots, N_e$ **do**
Generate $\mathcal{D}_e$
<Iterate over mini-batches>
for $m = 1, \ldots, M$ **do**
Generate $\mathcal{D}_m$
<Feed-Forward network construction>
for $l = 2, \ldots, L$ **do**
Compute all $x_j^{(l)}$
end for
for $l = L, \ldots, 2$ **do**
<Backpropagation>
Compute all $\delta_j^{(l)} := \nabla_{s_j^{(l)}} E$
<Update the weights>
$\mathbf{w}^{(l)} \leftarrow \mathbf{w}^{(l)} - \gamma X^{(l-1)} \left(\delta^{(l)}\right)^T$
end for
end for
end for
If crossentropy(e) $\leq$ crossentropy(e-1) then $\gamma \leftarrow \gamma/2$
end for
Return final weights $w_{i,j}^{(l)}$

Walk forward optimization



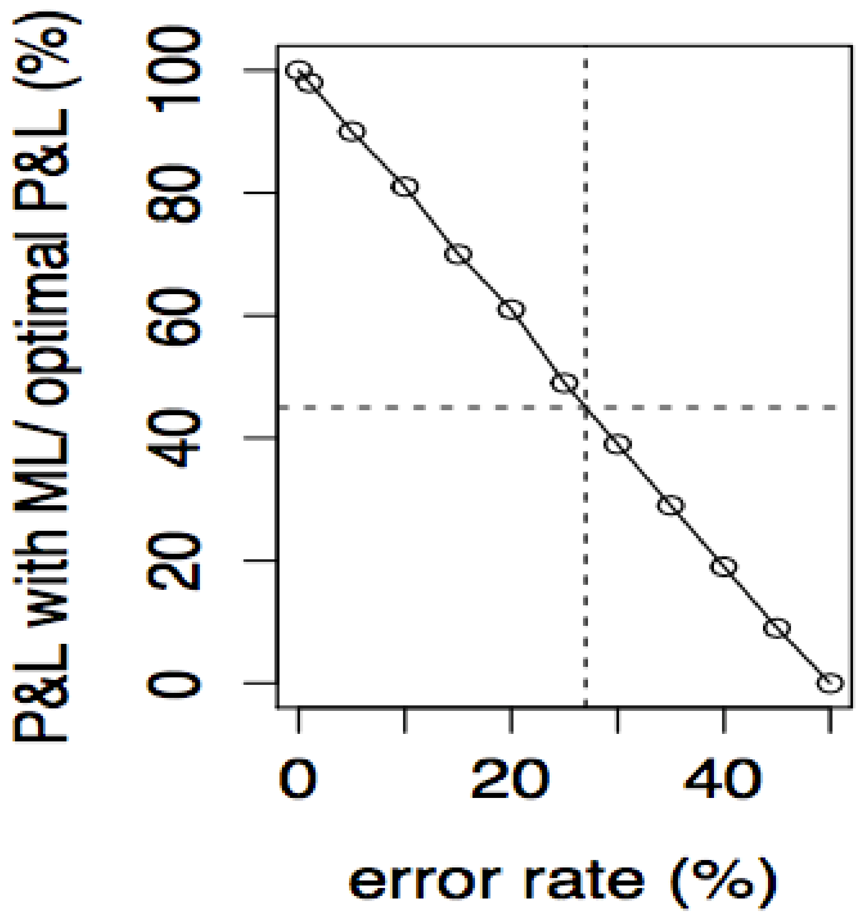
DNN Performance

method	in sample	out-of-sample	change (%)
ANN	0.75	0.66	-12.0
DNN	0.78	0.73	-6.4

Comparison of the classification accuracy of a classical ANN (with one hidden layer) and a DNN with four hidden layers. The in-sample and out-of-sample error rates are also compared to check for over-fitting.

DNN Performance

IIT Stuart School of Business
ILLINOIS INSTITUTE OF TECHNOLOGY



Implementation

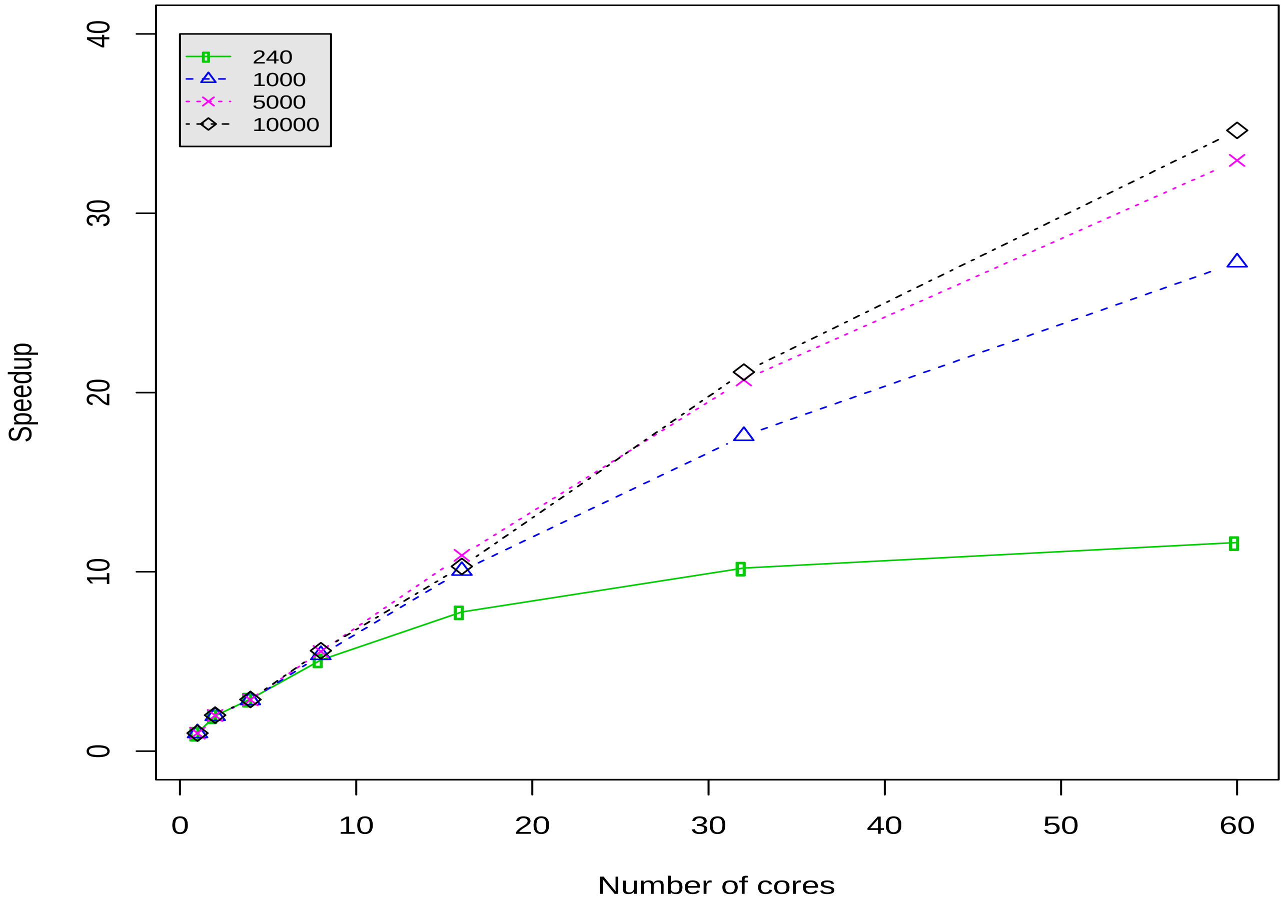
In designing an algorithm for parallel efficiency on a shared memory architecture, three design goals have been implemented:

1. The algorithm has to be designed with good data locality properties.
2. The dimension of the matrix or for loop being parallelized is at least equal to the number of threads.
3. BLAS routines from the MKL should be used in preference to openmp parallel for loop primitives.

Implementation

	CPU System	Co-processor System
Processor	Xeon E5-2690 v2 - 16 cores - 20 threads (HT on) - 2.30GHz	Xeon Xeon Phi 7120 - 61 cores - 244 threads (HT on) - 1.24GHz
ECC	on	on
RAM	128GB	16GB
OS	GNU 2.6.32	GNU 2.6.38
XE Composer	2015	
Compiler	ICC 15.0.2	
Flags	-O3	-O3 -mmic

Performance on the Intel Xeon Phi

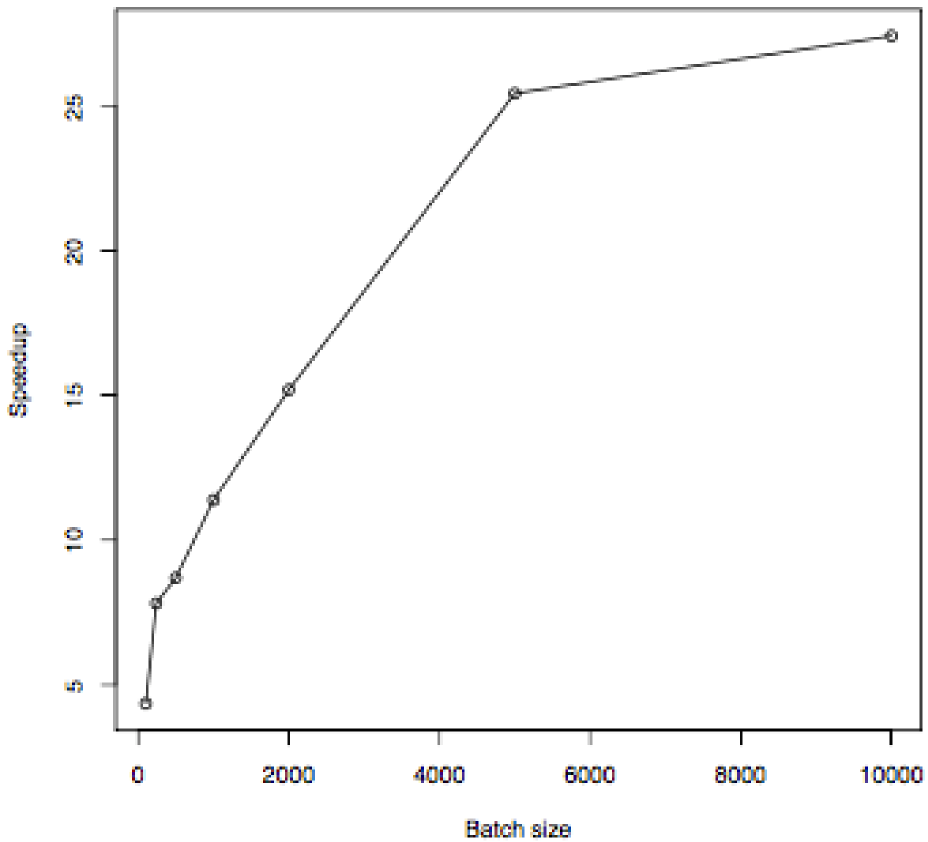


Speedup of the batched back-propagation algorithm on the Intel Xeon Phi as the number of cores is increased.

Performance on the Intel Xeon Phi

b	platform	ffwd	delta	weights	total
1000	baseline	2.03	2.22	1.34	5.59
	Xeon Phi	0.295	0.105	0.0915	0.491
	speedup	6.88 x	21.1 x	14.6x	11.4x
5000	baseline	8.45	21.0	6.71	36.1
	Xeon Phi	0.626	0.383	0.412	1.42
	speedup	13.9 x	54.8 x	16.3x	25.4x
10000	baseline	17.9	42.2	13.9	74.0
	Xeon Phi	1.15	0.704	0.846	2.70
	speedup	15.6 x	59.9 x	16.4x	27.4x

Performance on the Intel Xeon Phi

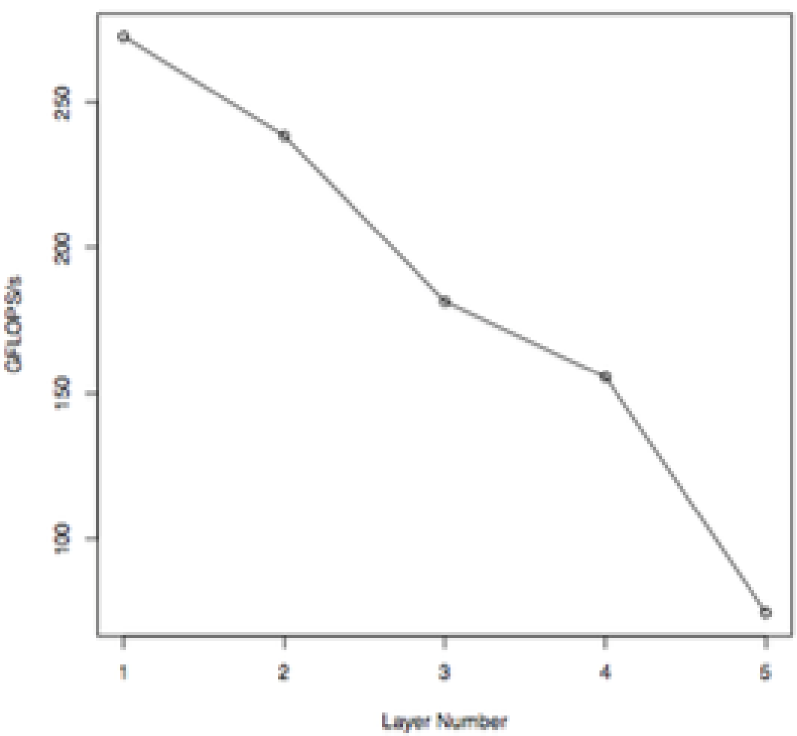


Speedup of the batched back-propagation algorithm on the Intel Xeon Phi relative to the baseline for various batch sizes.

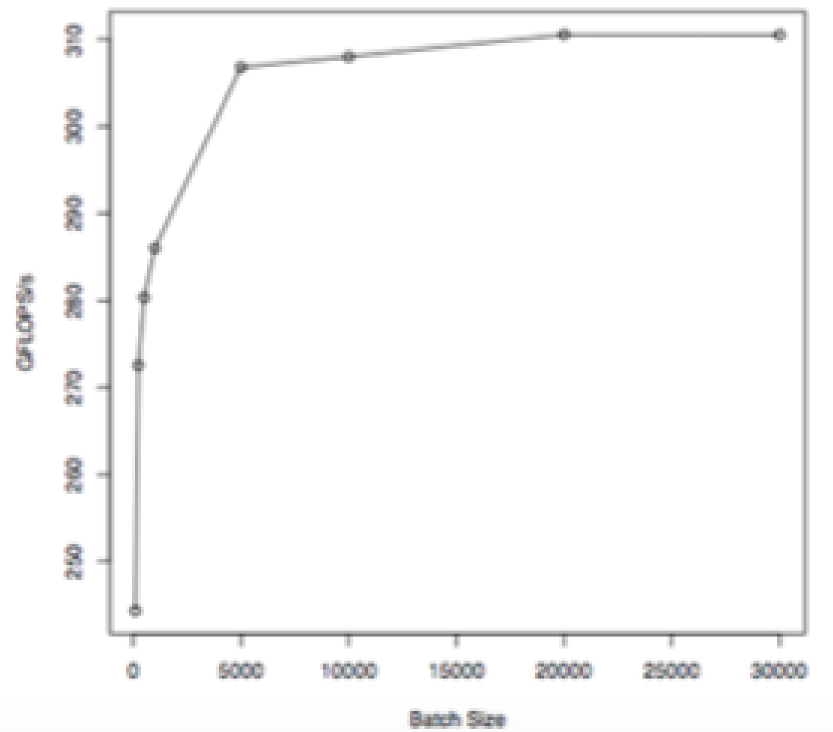
AACSB ACCREDITED

Performance on the Intel Xeon Phi

IIT Stuart School of Business
ILLINOIS INSTITUTE OF TECHNOLOGY



The GFLOPS of the weight matrix update (using dgemm) for each layer.



Variation of the GFLOPS of the first layer weight matrix update with the batch size.

Summary

- DNNs exhibit less overfitting than ANNs.
- Previous studies have applied ANNs to single instruments. We combine multiple instruments across multiple markets and engineer features based on lagging/moving averages and correlations of prices and returns respectively.
- The training of DNNs requires many parameters and is very computationally intensive –we solved this problem by using the Intel Xeon Phi.