From HFT to Laplace Demon

When timed data technology curves the market

@abifet

TELECOM ParisTech

@erichoresnyi

QCon LONDON

HFT in the hey days

High Frequency Trading

5ms=20m$

Source: Tabb Group

$100trn

HFT context

Rank	Manager	Market	Assets (US$million)
1	BlackRock	US	$4,651,896
2	Vanguard Group	US	$3,148,496
3	State Street Global	US	$2,448,112
4	Allianz Group	Germany	$2,189,296
5	Fidelity Investments	US	$1,974,077
6	J.P. Morgan Chase	US	$1,748,849
7	Bank of New York Mellon	US	$1,710,282
8	AXA Group	France	$1,491,394
9	Capital Group	US	$1,396,777
10	Deutsche Bank	Germany	$1,262,884
11	Goldman Sachs Group	US	$1,178,000
12	Prudential Financial	US	$1,175,947
13	UBS	Switzerland	$1,158,763
14	BNP Paribas	France	$1,114,595
15	Legal & General Group	UK	$1,077,425
16	Amundi	France	$1,052,587

Blackrock*
Pimco

CapGroup

*Blackrock is actually headquartered in NY, main AUM coming from ETF/ passive originally BGI in SF

Fidelity
StateStreet
GS **BoNY**
JPM
Prudential
Vanguard

AUM>$1trn, source: Towers Watson

Approx 3xGDP in USA ie 155k$/hab

QCon LONDON

TELECOM ParisTech

Liquidity Flow

HFT context

PCX

NASDAQ

Buy Side 1

NSX

NYSE

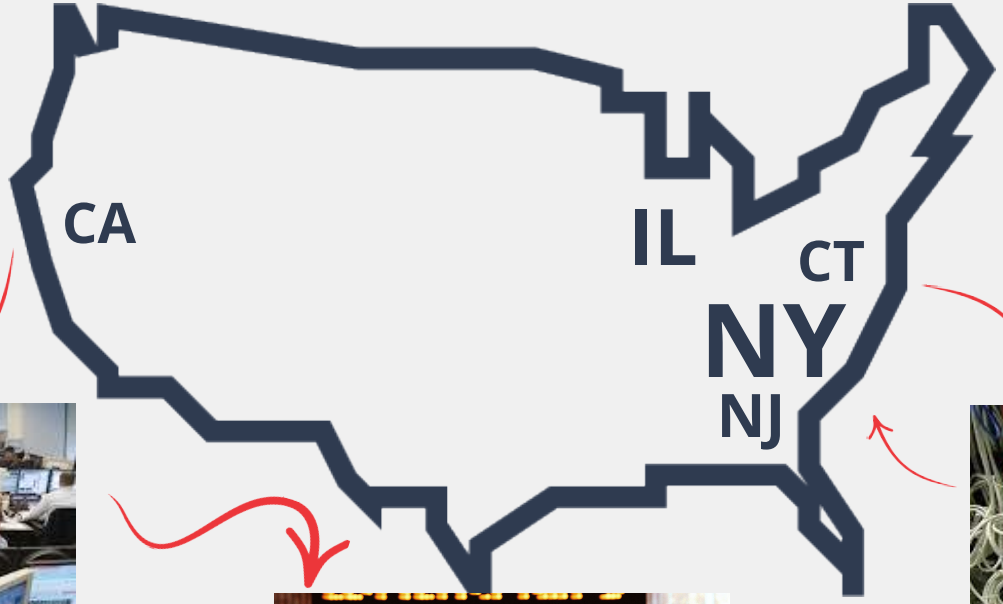
Sell Side 2

Sell Side 1

Buy Side 2

Order Flow

HFT context



Sell Side 1

Market Maker

Sell Side 2

Cambrian Explosion

HFT context

ARCA

CME

CBOT

CBOE

NASDAQ

IEX

BRUT

NYSE

INET

NYMEX

BATS

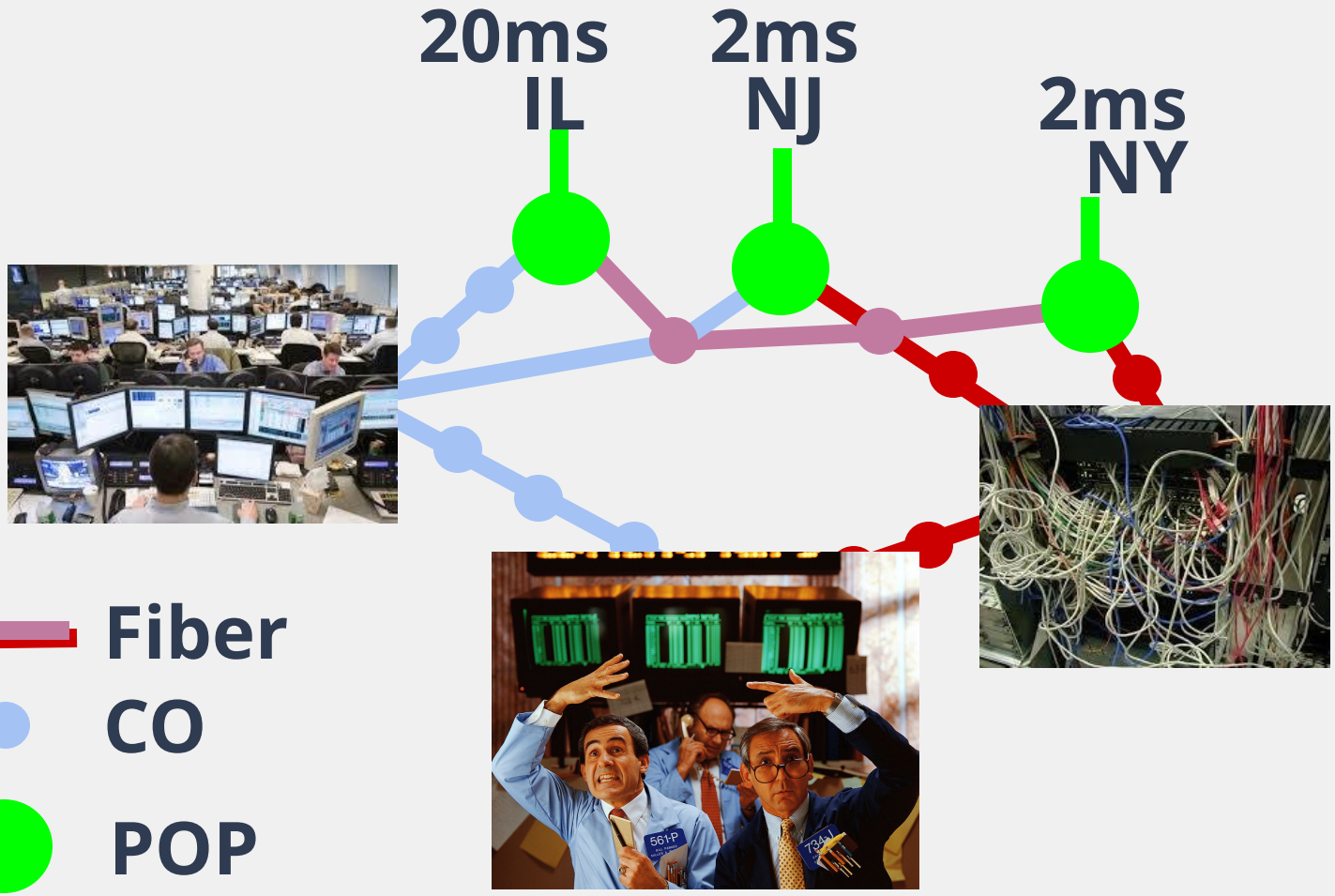
ICE

Reg.ATS'98-Reg.NMS'05

PCX

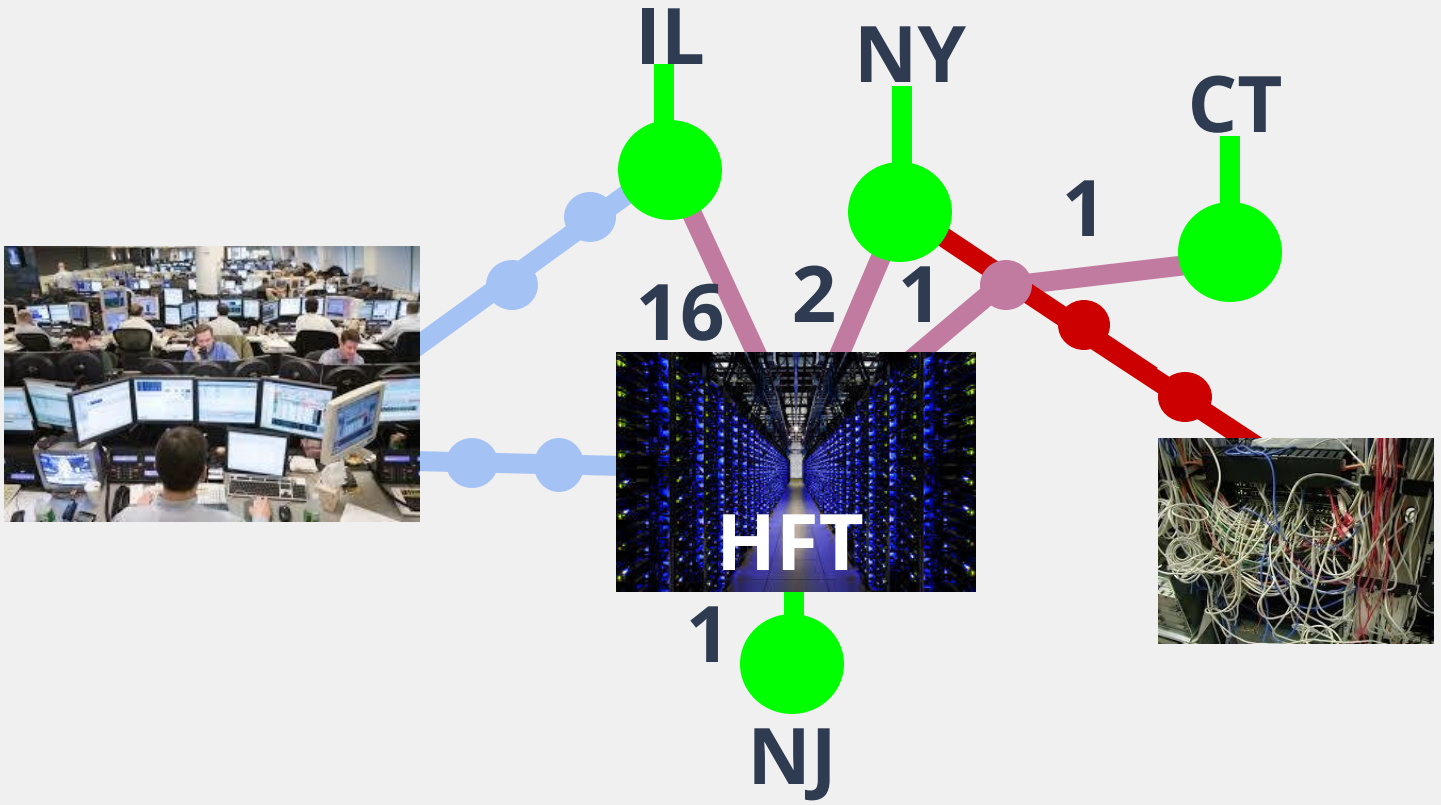
Infra view

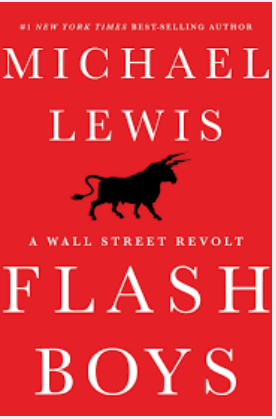
HFT context



HFT: Proximity

Host in Network Nodes, then Exchanges



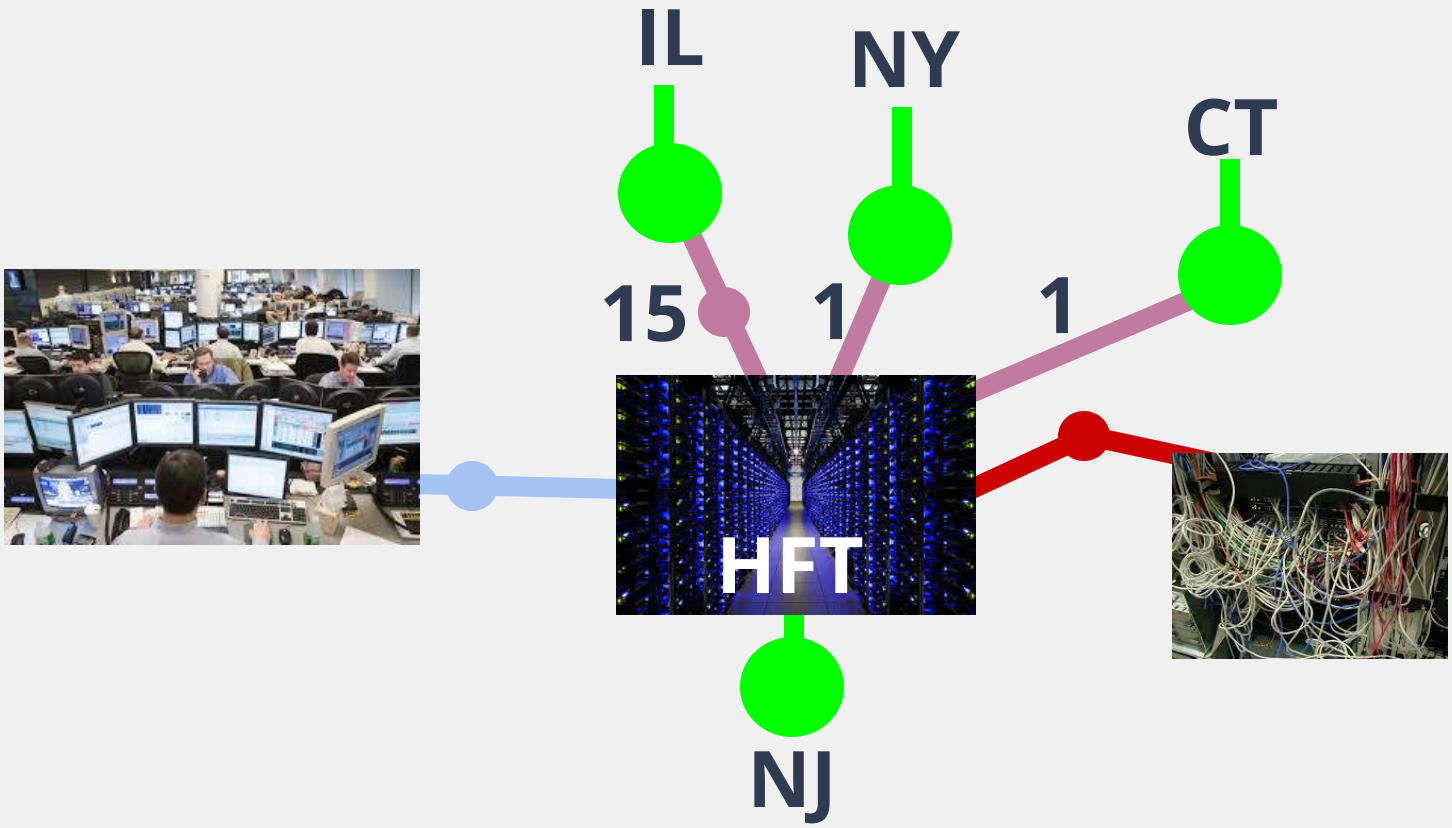


Latency
=
Propagation
+
Serialization
+
Processing


#1 NEW YORK TIMES BEST-SELLING AUTHOR
MICHAEL
LEWIS
A WALL STREET REVOLT
FLASH
BOYS
QCon
LONDON
TELECOM
ParisTech

HFT: Ultra

Dark Fiber



Buy-Side view of HFT

It's not a ghost...

HFT: Straight Fiber

1,000 miles > 825 miles

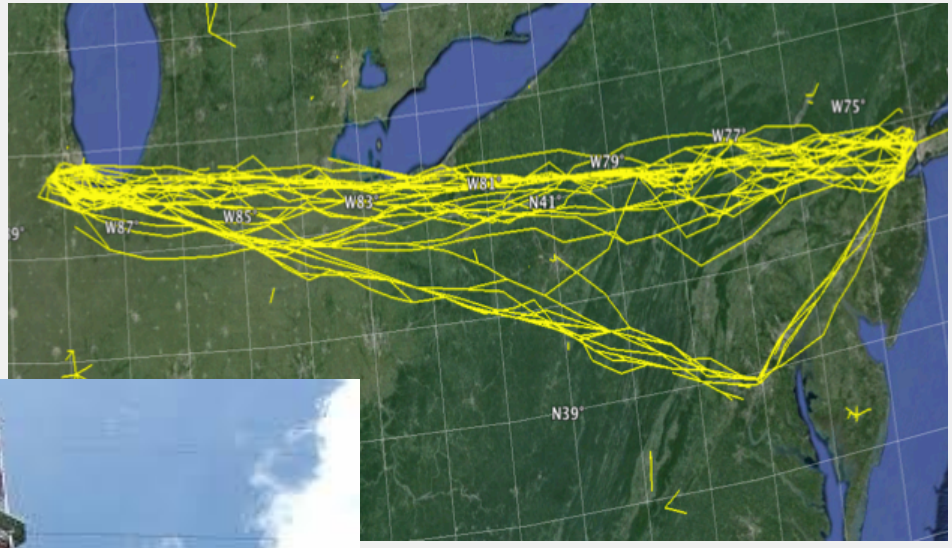
14.5 ms > 11.5 ms

HFT: Microwaves

11.5 > 8.5ms

N:1.33 > 1.0003

v = c/n



HFT: FPGA

Nanosecs

Choose your lane

HFT <> Algo Trading

"Once you get into milliseconds it's almost not HFT any more"

QCon LONDON

TELECOM ParisTech

Spacetime is relative

Market Events: [ct,x,y,z]

Speed curves spacetime

HFT built a wormhole to win on [ct',x,y,z] events

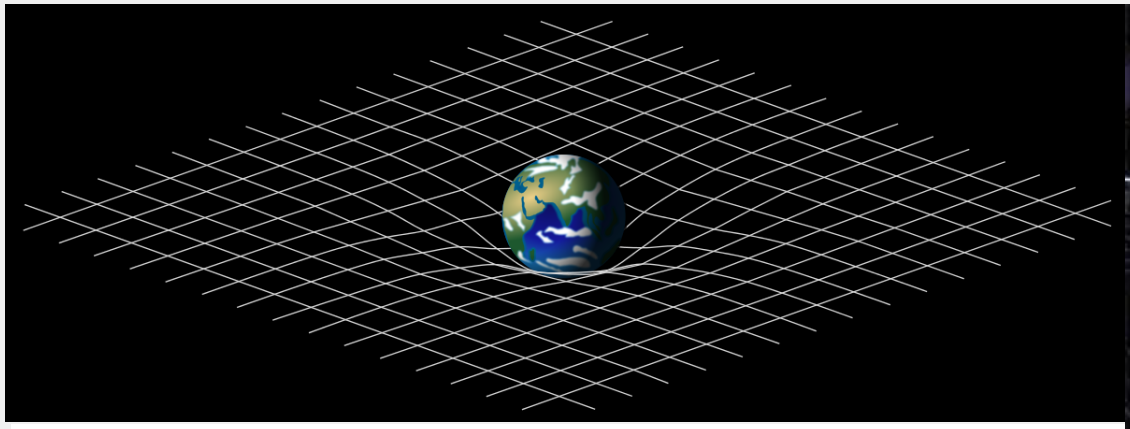
$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{v^2}{c^2}}}$$

QCon LONDON

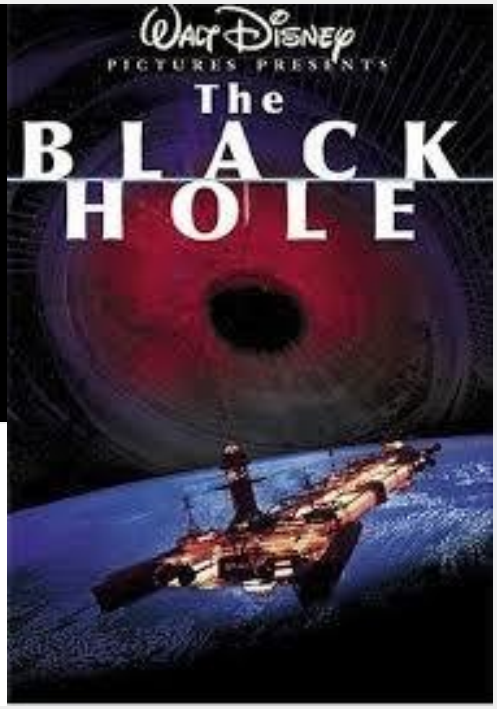
TELECOM ParisTech

Mass curves spacetime

AI builds a blackhole by massively processing [ct,x,y,z] events



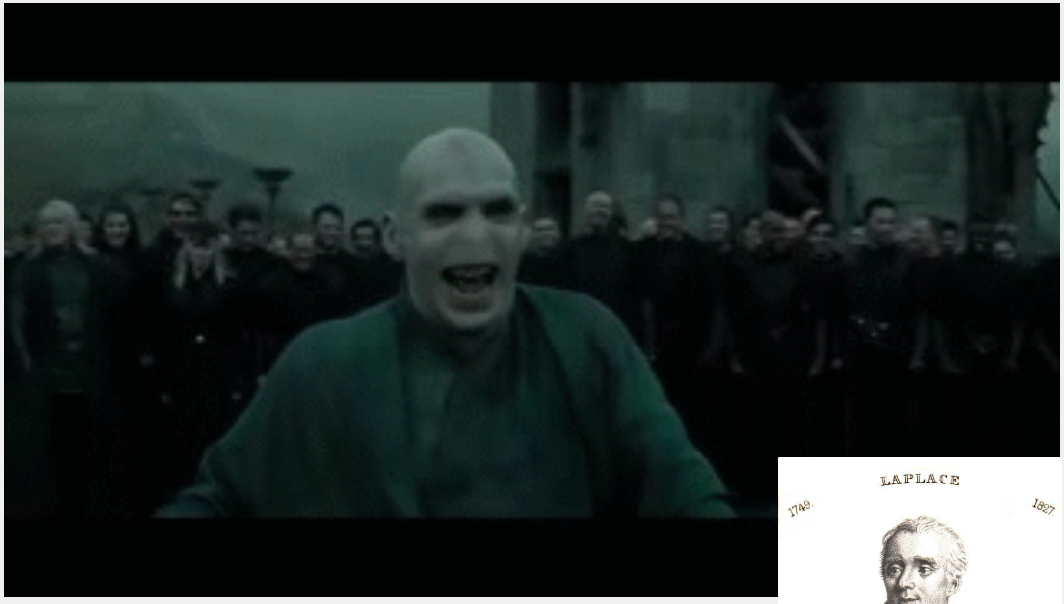
$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$



Laplace Demon

The endgame of Determinism

$$\forall\ [ct,x,y,z] \in \mathbf{R}n \vdash \forall\ [ct',x',y',z']$$

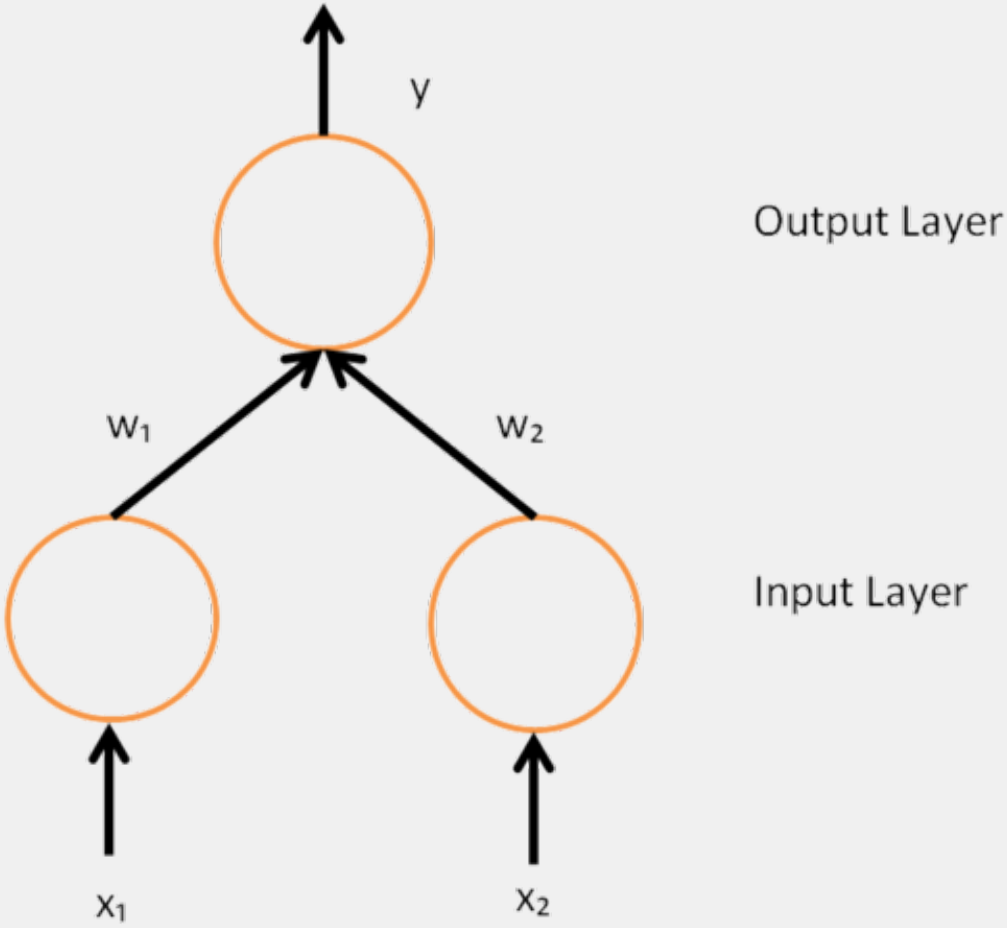


QCon LONDON

TELECOM ParisTech

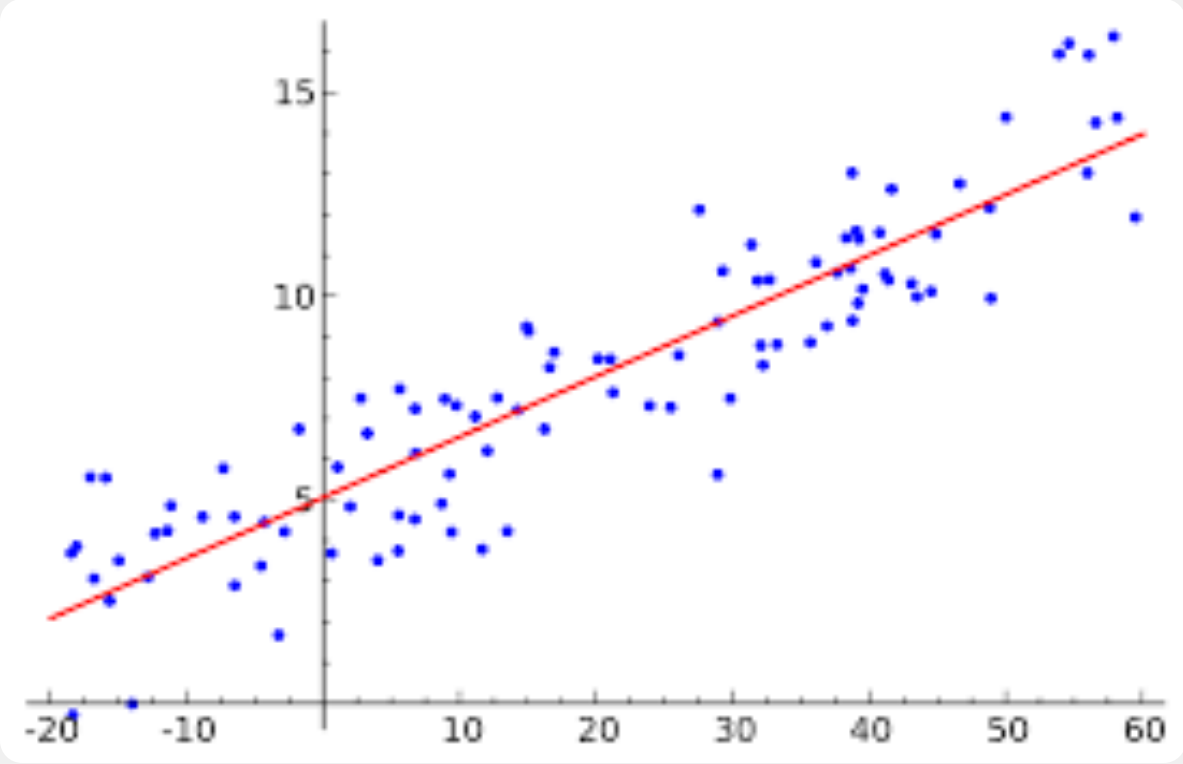
The Endgame 1/3

Event Machine View



The Endgame 2/3

Graph View : Regression



Loss aka Cost Function = J(θ) : distance points to line

The Endgame 3/3

Matrix view

Features

$$[x_0, x_1, \ldots, x_n]$$

$$\begin{bmatrix} w_{1,1} w_{1,2} \ldots w_{1,n} \\ w_{2,1} w_{2,2} \ldots w_{2,n} \\ w_{\ldots,1} w_{\ldots,2} \ldots w_{\ldots,n} \\ w_{n,1} w_{n,2} \ldots w_{n,n} \end{bmatrix} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ . \\ y_n \end{bmatrix}$$

Matrices of Weights

Labels

$AAPL

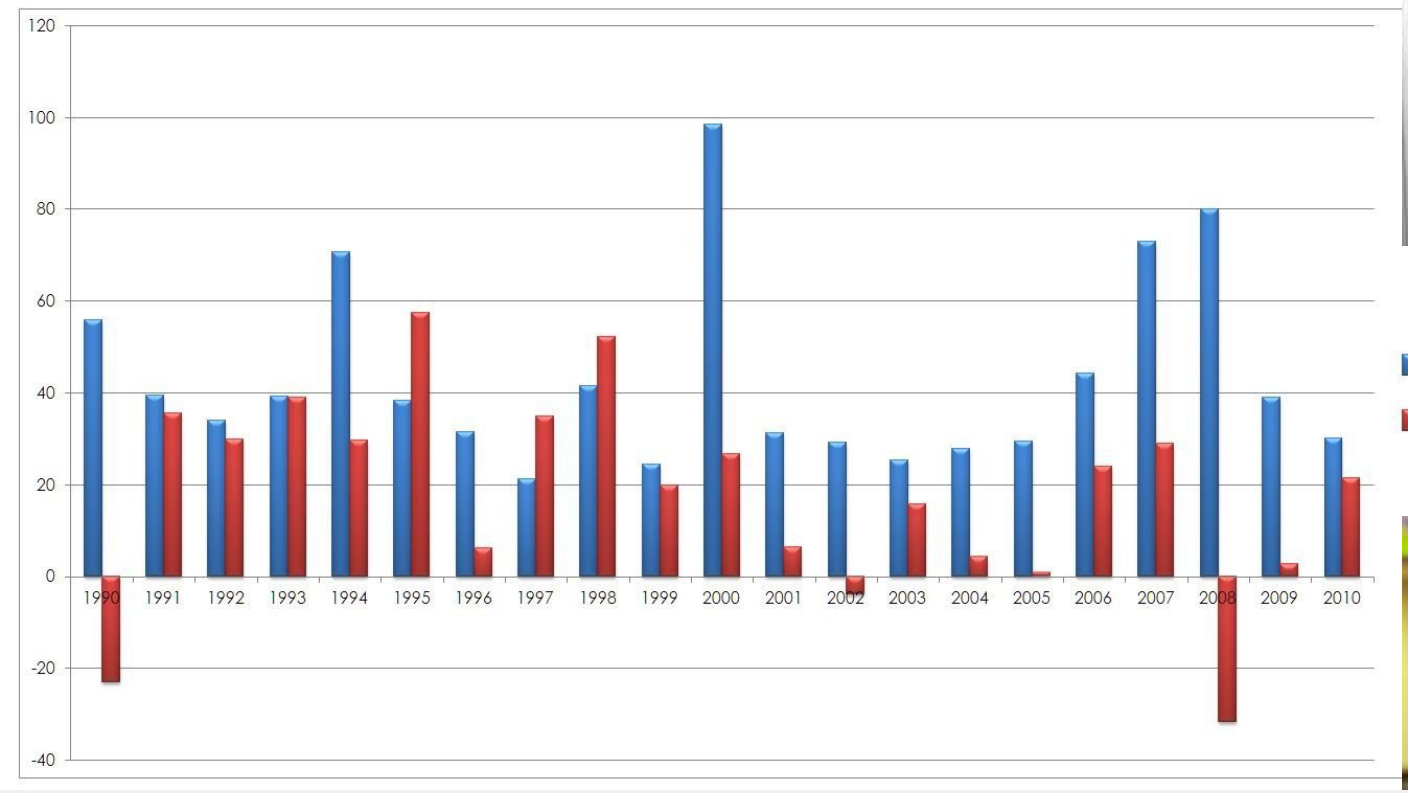
$GOOG

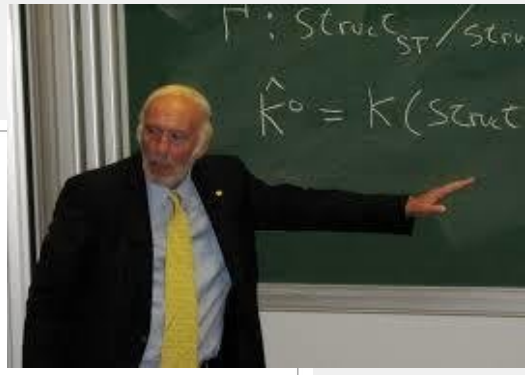
.

$FB

AI not news to trading

+35% yoy for 20 years : $2,500 > $1,000,000



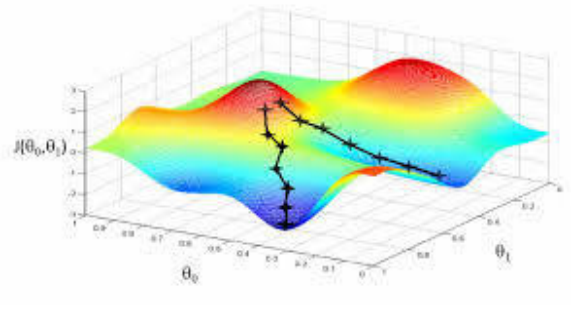


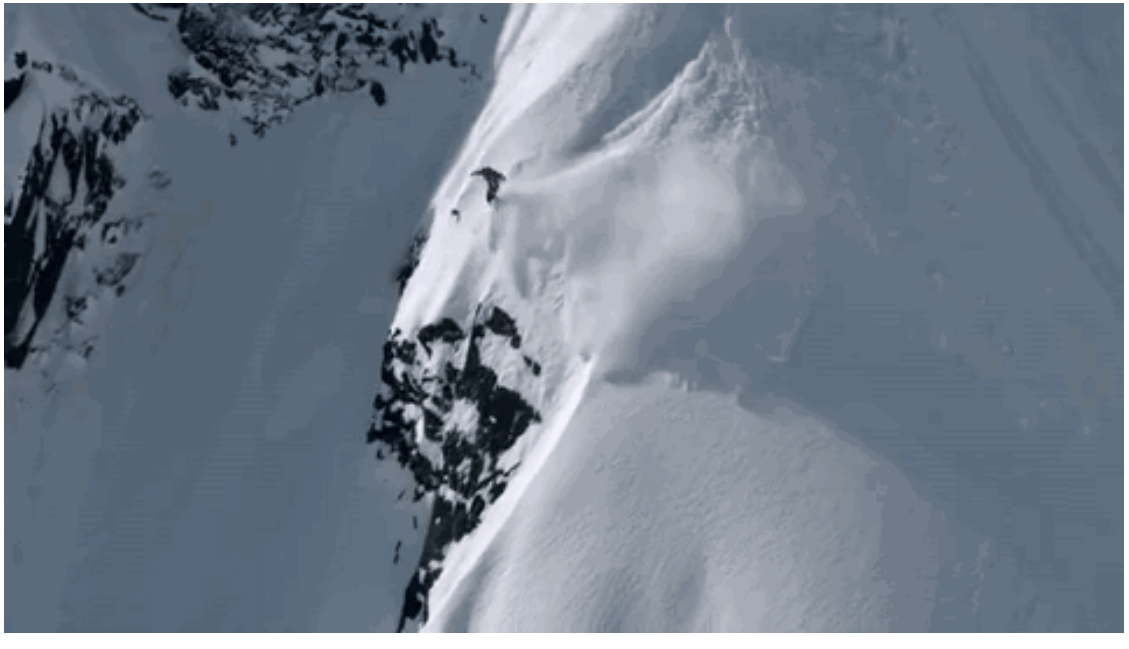
PhD Mathematics, Berkeley - String Theory Chern-Simons Form

AI age:Gradient Descent

Follow the steepest slope, 100m+ features

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$





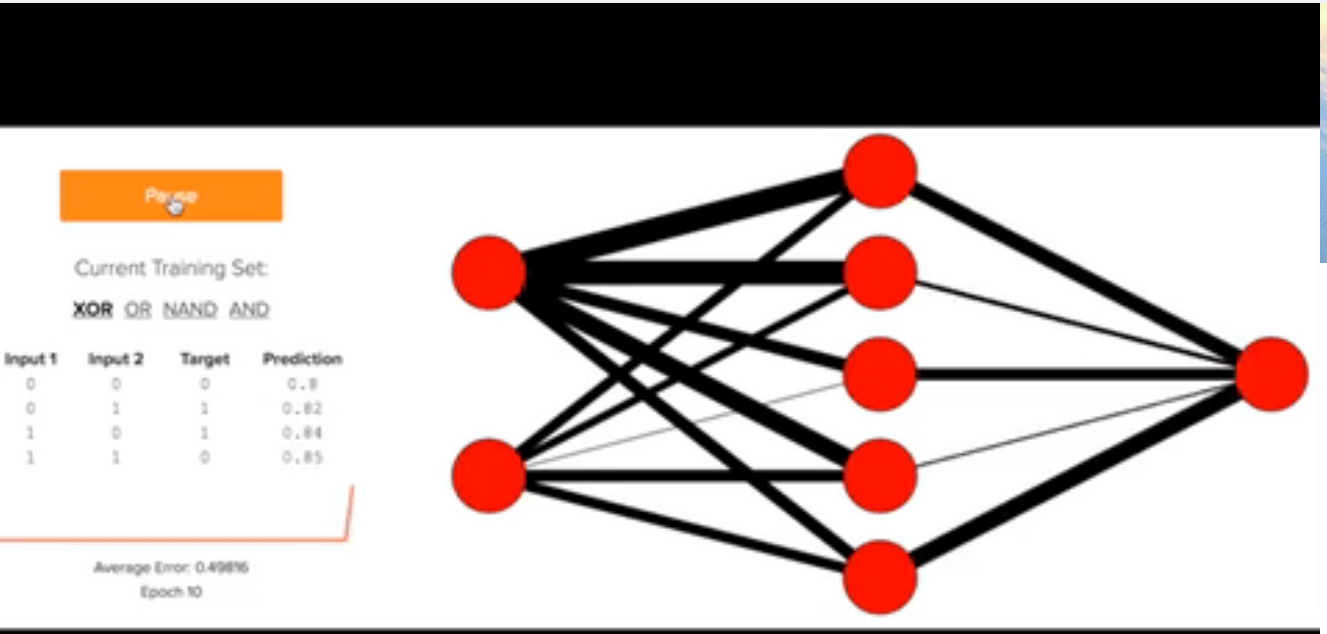
α : Learning Rate, ∇J : Gradient

QCon LONDON

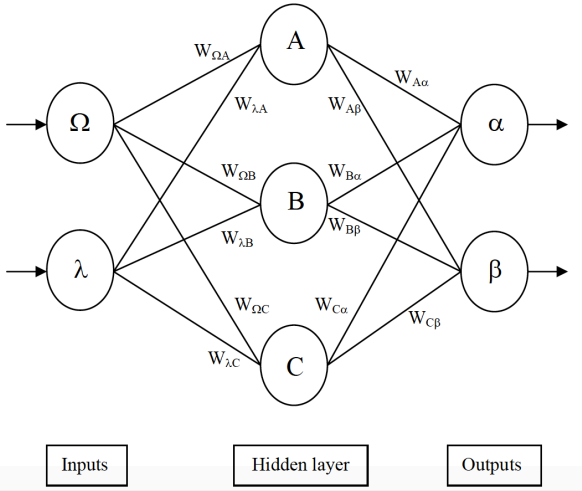
TELECOM ParisTech

AI age:Back Propagation

Adapt weight to control error from previous layer's input, 150+ layers



Source: Neural Networks simulation by Matt Mazur at Emergent Mind



AI age: GPU

From Final Fantasy to Autonomous Car

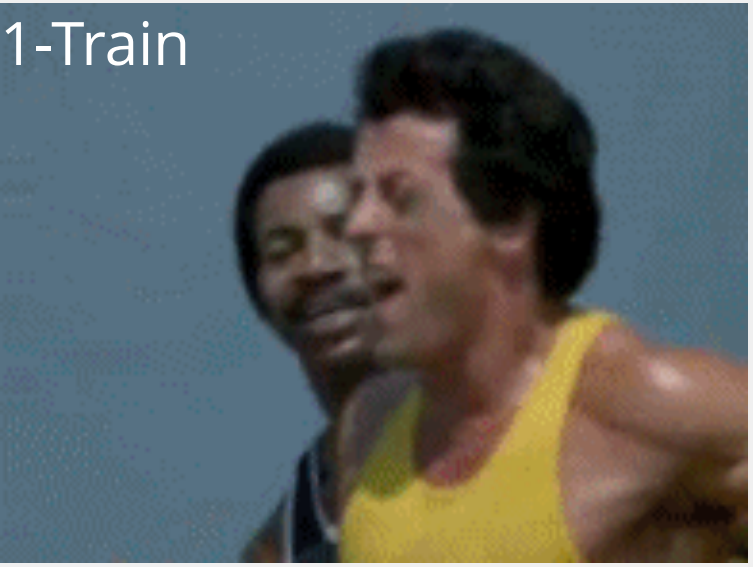
$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

"The implementation of streaming algorithms, typied by highly parallel computations with little reuse of input data, has been widely explored on GPUs." (Stanford, 2004)

QCon LONDON

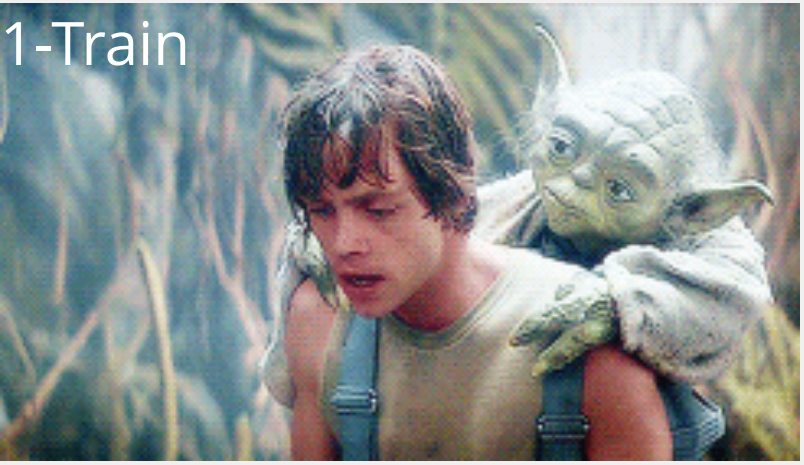
TELECOM ParisTech

Bullish Fitness Drill



Overfitting?

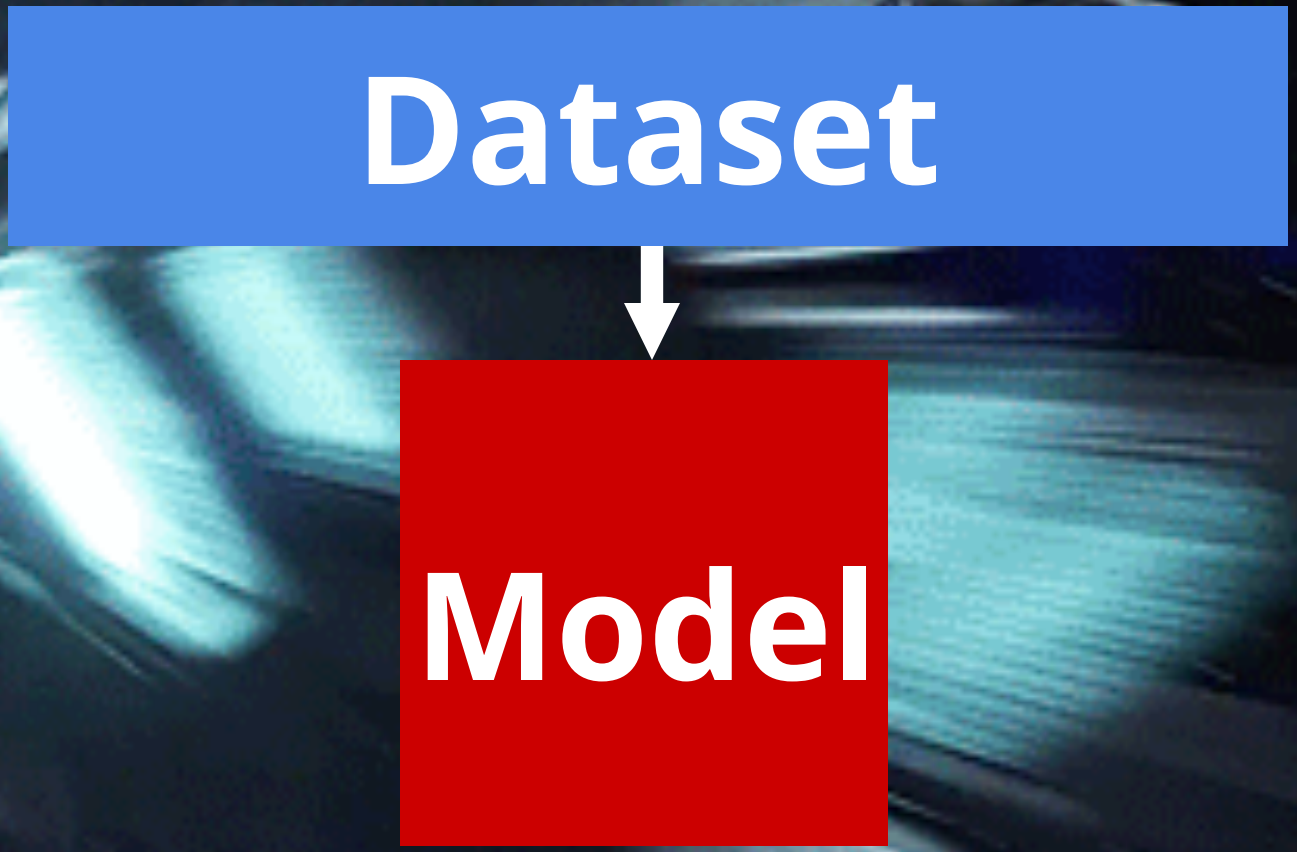
Bearish Fitness Drill



Overfitting?

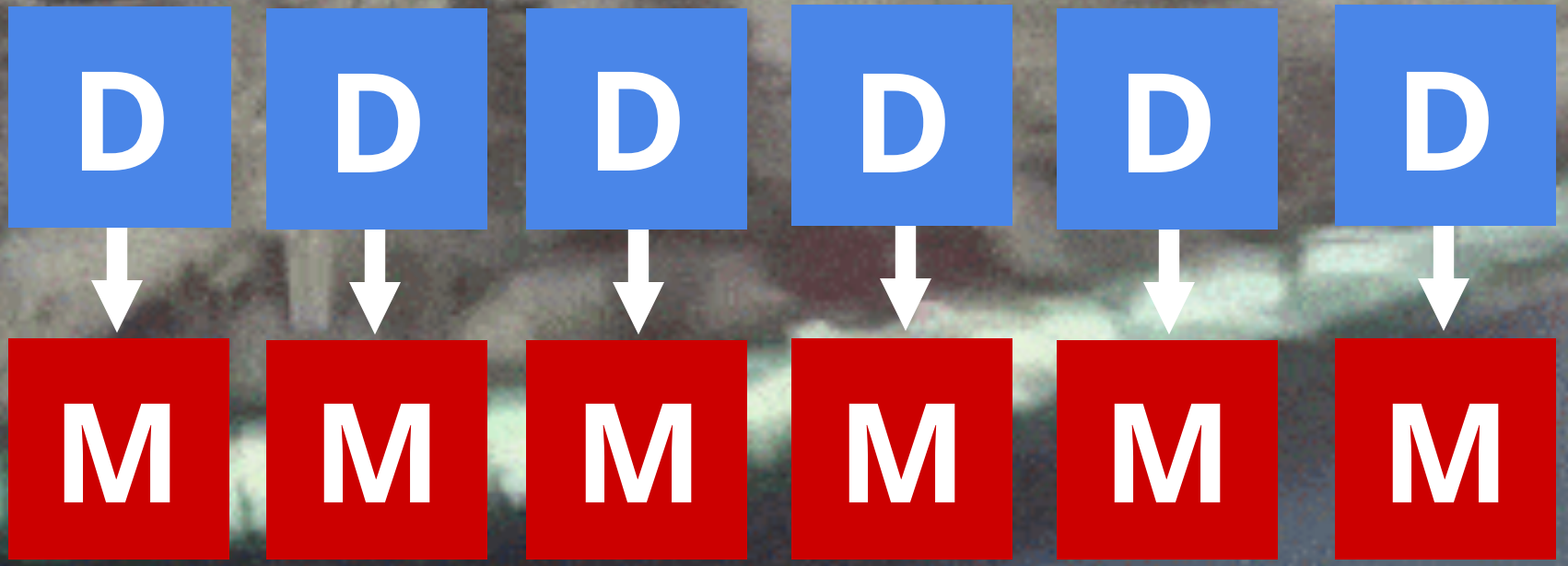
Standard Approach

Batch-based, finite training sets, static models



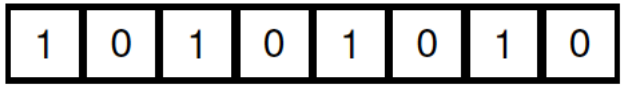
Data Stream Approach

Infinite training sets, dynamic models



Approximation Algo

What is the largest number that we can store in 8 bits?



Programming Techniques

S.L. Graham, R.L. Rivest Editors

Counting Large Numbers of Events in Small Registers

Robert Morris
Bell Laboratories, Murray Hill, N.J.

It is possible to use a small counter to keep approximate counts of large numbers. The resulting expected error can be rather precisely controlled. An example is given in which 8-bit counters (bytes) are used to keep track of as many as 130,000 events with a relative error which is substantially independent of the number n of events. This relative error can be expected to be 24 percent or less 95 percent of the time (i.e. $\sigma = n/8$). The techniques could be used to advantage in multichannel counting hardware or software used for the monitoring of experiments or processes.

QCon LONDON

TELECOM ParisTech

Approximation Algo

What is the largest number that we can store in 8 bits?

MORRIS APPROXIMATE COUNTING ALGORITHM

```
1   Init counter c ← 0
2   for every event in the stream
3       do rand = random number between 0 and 1
4           if rand < p
5               then c ← c + 1
```

Approximation Algorithm

101100011110101 0111010

Sliding Window

We can maintain simple statistics over sliding windows, using $O(\frac{1}{\epsilon}\log^2 N)$ space, where

- N is the length of the sliding window
- ϵ is the accuracy parameter

M. Datar, A. Gionis, P. Indyk, and R. Motwani.
Maintaining stream statistics over sliding windows. 2002

Massive Online Analysis

moa

QCon LONDON

TELECOM ParisTech

Stream Setting

Process an example at a time

Inspect it only once (at most)

Use a limited amount of memory

Work in a limited amount of time

Be ready to predict at any point

Prequential Evaluation

Sequence of examples > Error of a model

$$S = \sum_{i=1}^{n} L(y_i, \hat{y}_i).$$

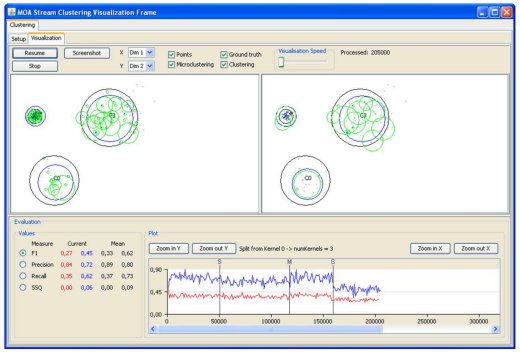
Command Line


```
java -cp .:moa.jar:weka.jar -javaagent:sizeofag.jar moa.DoTask
EvaluatePrequential

-l DecisionStump //training DecisionStump classifier ...

-s generators.WaveformGenerator //...on WaveformGenerator data

-n 100000 //using the first 100 thousand examples for testing

-i 100000000 //training on a total of 100 million examples

-f 1000000 //testing every one million examples

> dsresult.csv
```

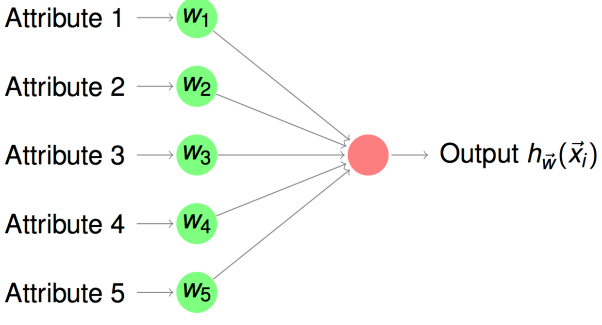
Resourceful

moa

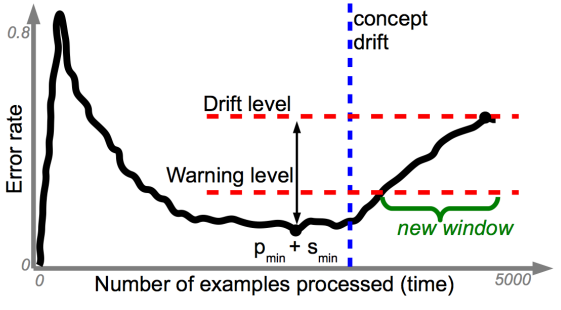
Classification



Regression



Concept Drift



Sentiment Analysis

Stock Price

Alerting

QCon LONDON

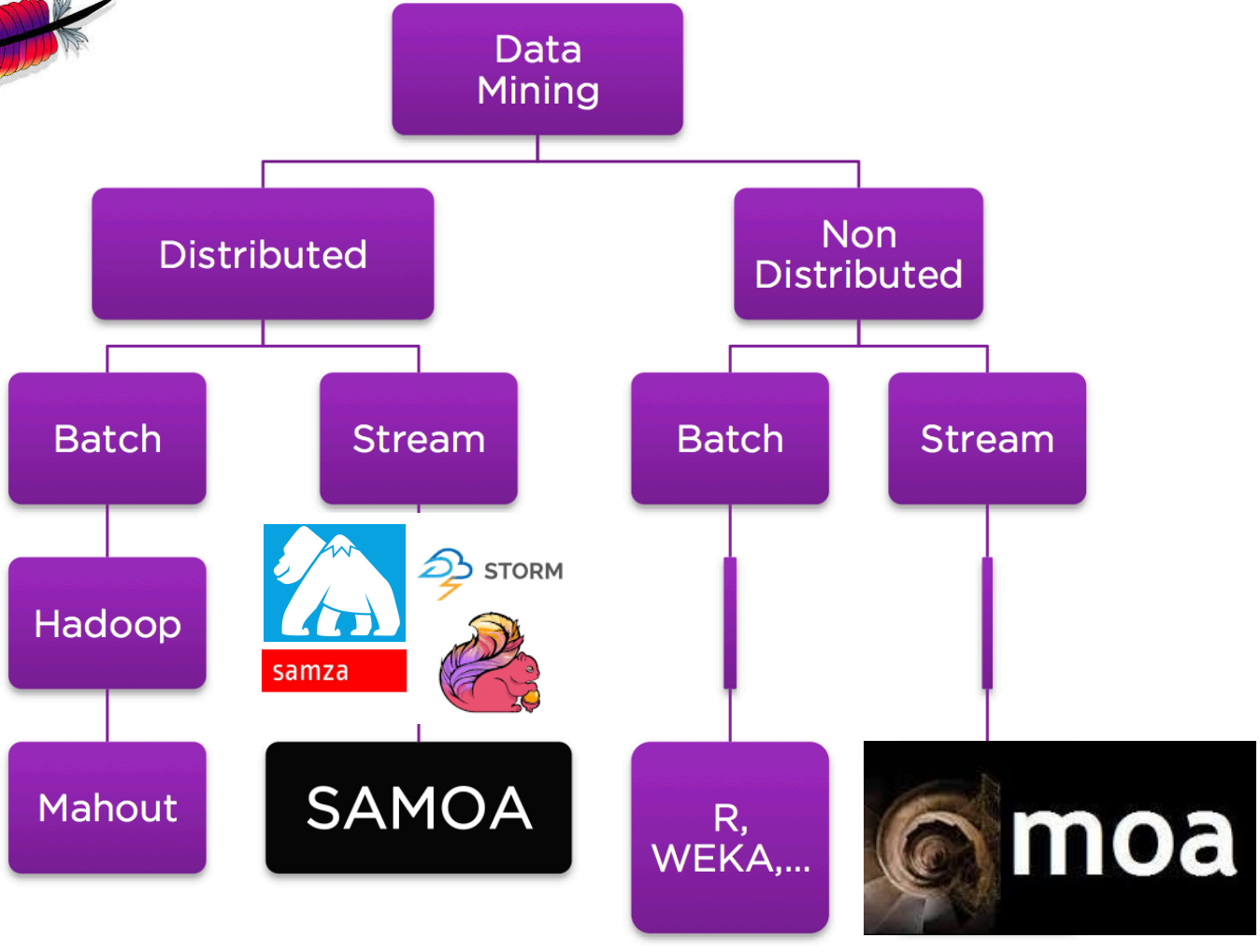
TELECOM ParisTech

Simple


```
learner.getVotesForInstance(instance)
learner.trainOnInstance(instance)
```

Scalable

http://samoa-project.net

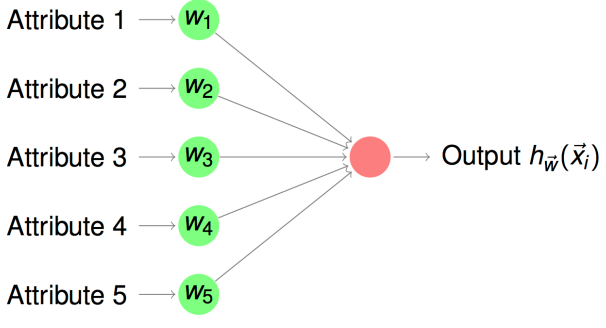


An experiment

Public Stock Dataset

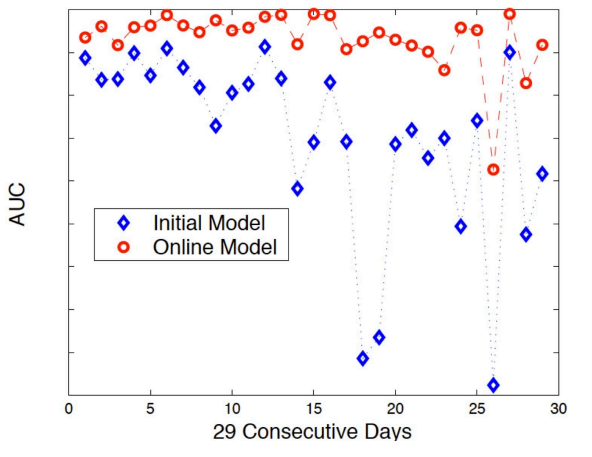
Name	Price	Time	+/-	Change
HOME DEPOT	71.18 $	9:47:05	▲	2.95 %
AT&T	36.71 $	9:47:00	▲	4.88 %
ALCOA	8.61 $	9:47:05	▲	1.30 %
BOEING CO	78.91 $	9:47:01	▼	-7.17 %
HEWLETT PACKARD	22.42 $	9:46:54	▼	-4.20 %
CISCO SYSTEMS	21.52 $	9:47:05	▲	5.50 %
TRAVELERS COMPANIES (THE)	88.50 $	9:46:59	▲	6.47 %
UNITEDHEALTH GROUP	56.06 $	9:46:53	▼	-0.56 %
BANK OF AMERICA	13.13 $	9:47:05	▲	9.43 %

MOA Regression

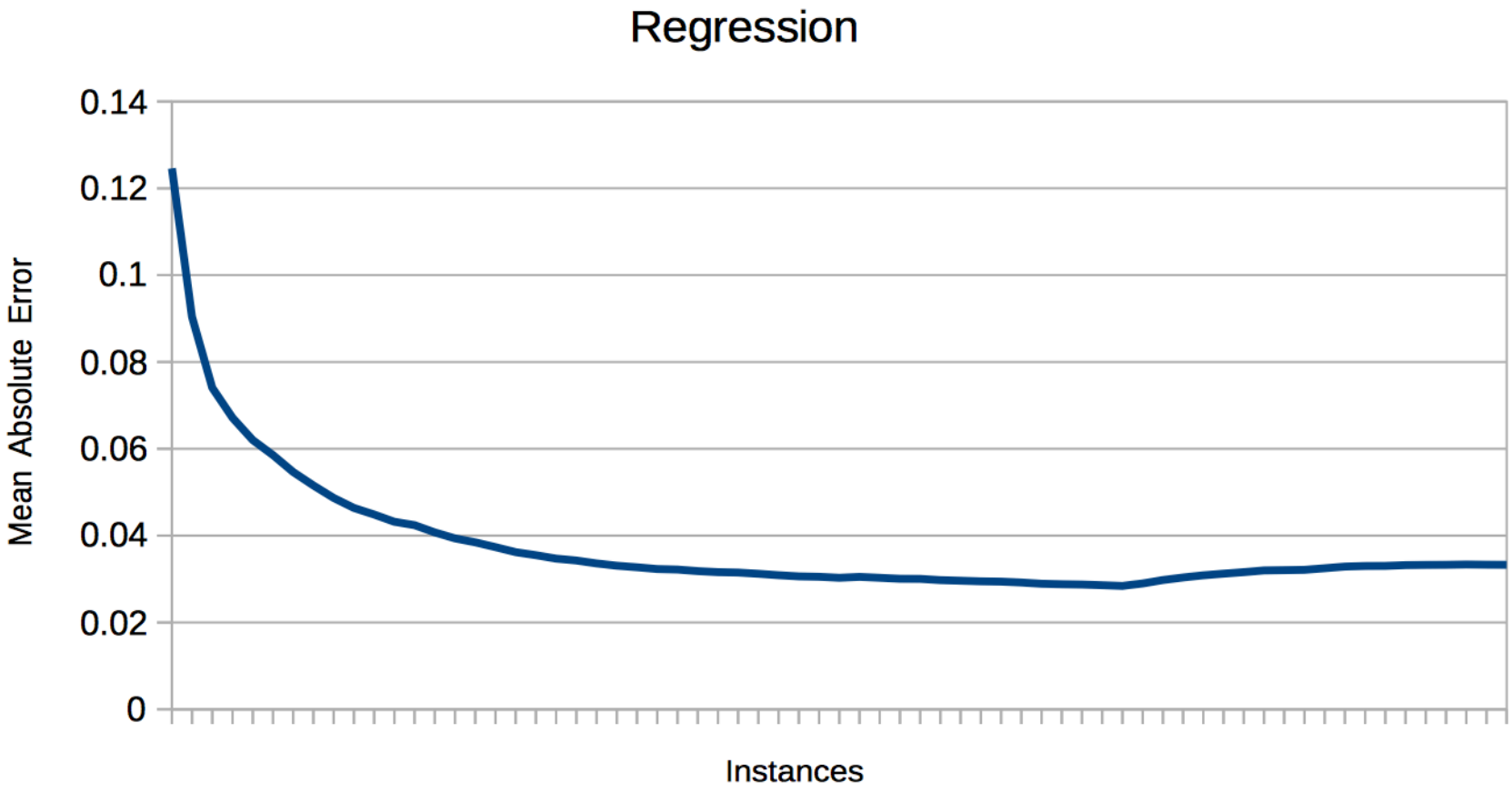


Stock Price

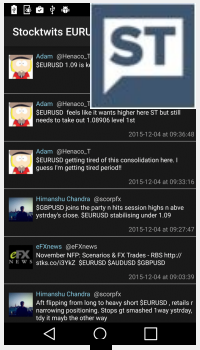
Error



An Experiment



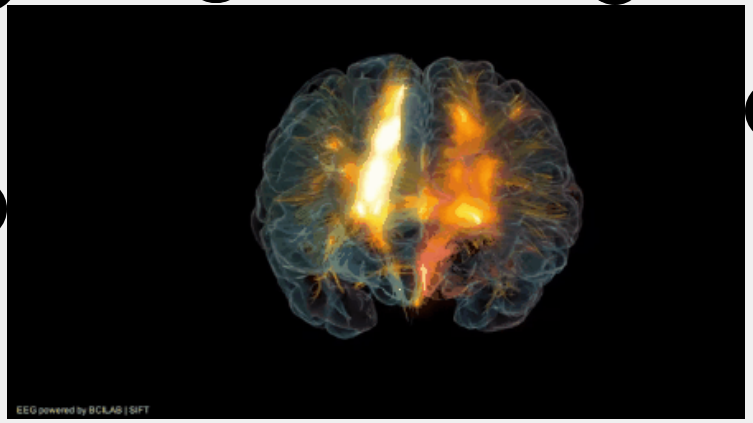
The New HF Frontier: AI



estimize

PSYCHSIGNAL
TRADER MOOD, QUANTIFIED

Sentiment Analysis



Regression/Perceptron

Alerts

{API}

NYSE

xignite®
The Market Data Cloud

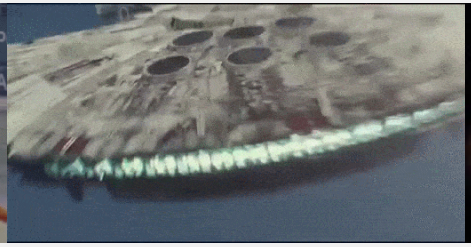
QCon LONDON

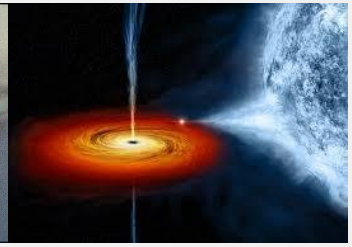
TELECOM ParisTech

Fast vs Smart

Data Stream a compromise

ct x,y,z





HFT Data Stream AI

Thanks!

Apache & Wikipedia Foundation : please donate!

MOA, Kaggle & Giphy : please contribute!

Books & Lectures

Data Stream Mining, MOA team

Yann LeCun Deep Learning Class, NYU

Matt Mazure, Emergent Mind & Andrew Ng, Coursera on AI

My Life as a Quant:Reflections on Physics&Finance, E.Derman

The Value of a Millisecond: Finding the Optimal Speed of a Trading Infra., TabbGroup

Flashboys, M.Lewis

Movies & Games

The Big Short, Back to the Future, Interstellar, The Black Hole,

Harry Potter, Rocky, Into the Mind, Star Wars, Matrix; Final Fantasy

TELECOM ParisTech