Lecture 1: Intro to Machine Learning

Statistics 208

At BlackRock, Machines Are Rising Over Managers to Pick Stocks

By **LANDON THOMAS Jr.** MARCH 28, 2017

RELATED COVERAGE

At BlackRock, a W
Trillion Comeback

BlackRock Hires C
Professor to Guide
Strategy JUNE 9, 2015

Investing With a C
Robot APRIL 6, 2016

What does Steve think?

“The development of full artificial intelligence could spell the end of the human race.”

-Steven Hawkings (Physicist)

"If we build these devices to take care of everything for us, eventually they'll think faster than us and they'll get rid of the slow humans to run companies more efficiently."

-Steve Wosniak (Apple co-founder)

"It's not even on our radar screen,” he said at a media event, adding that significant workforce disruption due to AI is “50 to 100" years away. "I'm not worried at all" about robots displacing humans in the near future, Mnuchin said. "In fact I'm optimistic."

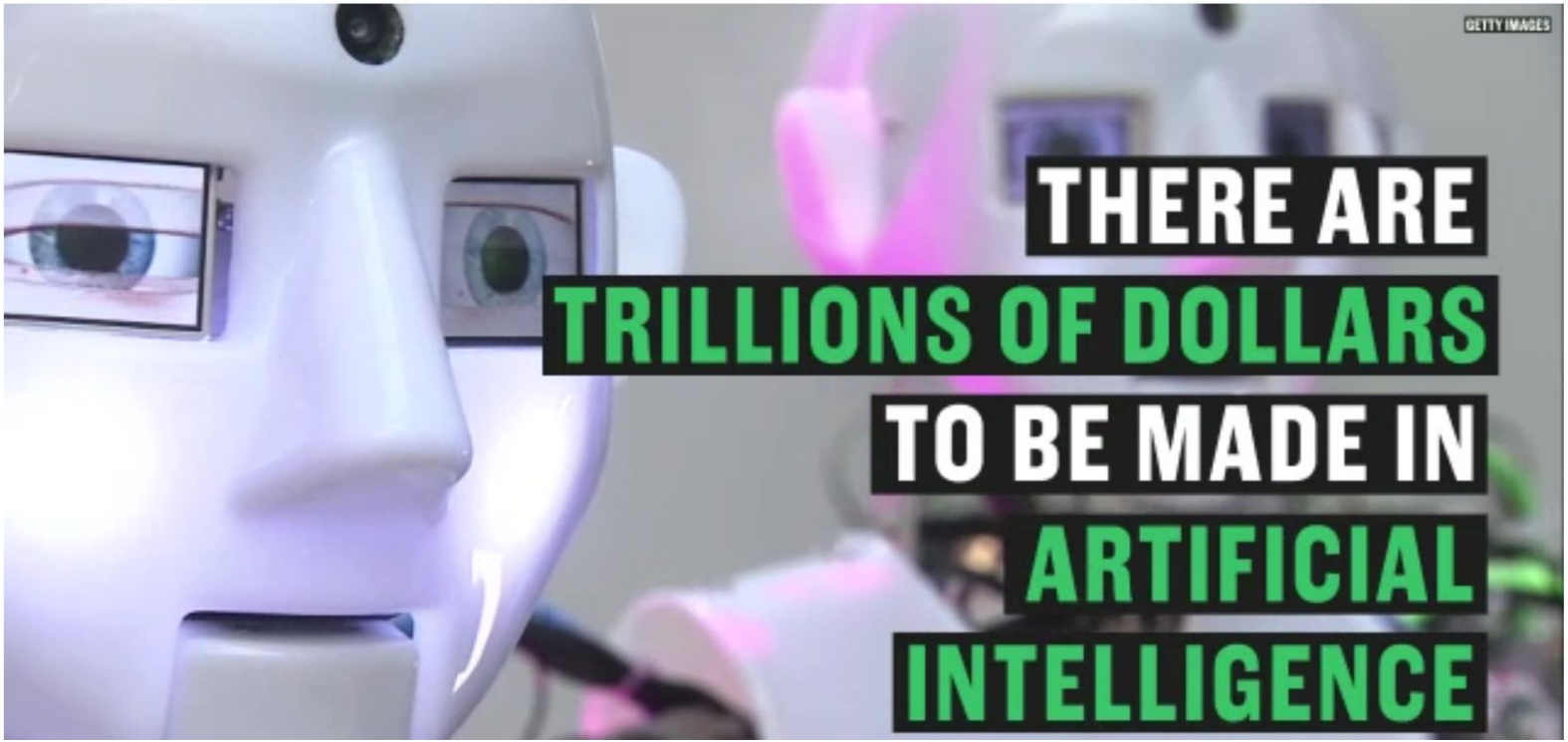
-Steve Mnuchin (US Treasury Secretary)

Op-Ed Trump's Treasury secretary is an Artificial Intelligence denier

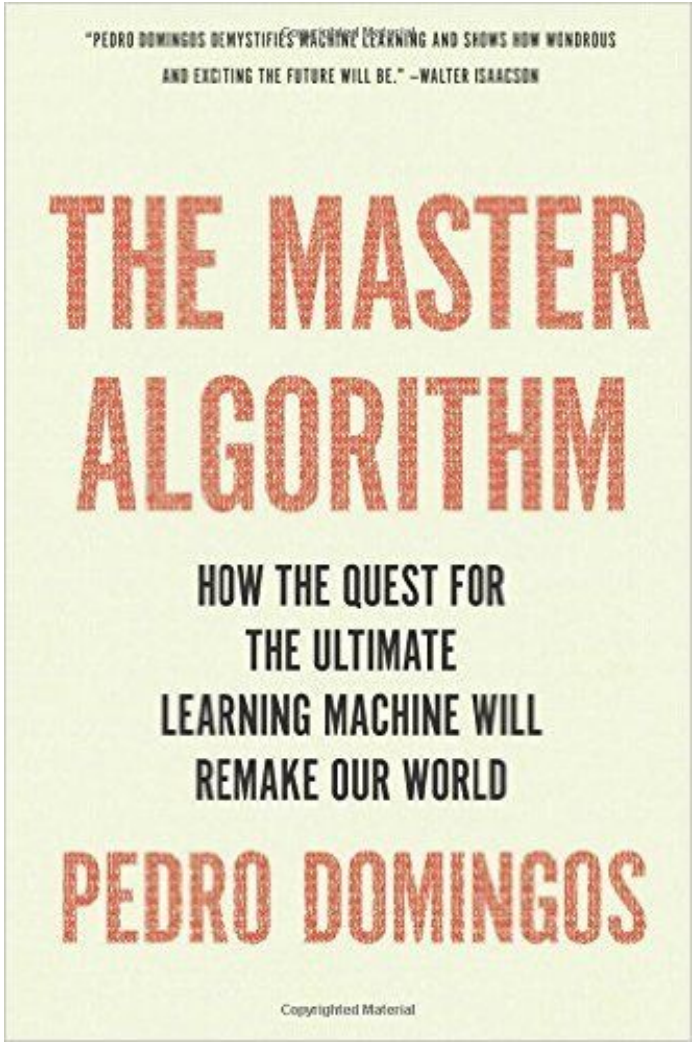
In Case You Misse

Mark Cuban: The world's first trillionaire will be an artificial intelligence entrepreneur

Catherine Clifford | Monday, 13 Mar 2017 | 9:54 AM ET



cnbc.com


"PEDRO DOMINGOS DEMYSTIFIES MACHINE LEARNING AND SHOWS HOW WONDROUS AND EXCITING THE FUTURE WILL BE." –WALTER ISAACSON
THE MASTER ALGORITHM
HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD
PEDRO DOMINGOS
Copyrighted Material

What is Machine Learning?

“How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?”

Task: What do you want the system to do?

Performance Metric: How do we evaluate how well it is doing the task?

Experience: How does the system collect data and interact with its environment?

Speech Recognition

Data: audio snippets and word labels

Task: given audio input (predictor), label the snippet with a word (response)

Performance: % of correct words (accuracy)

“'Cause I knew you were trouble when you walked in
So shame on me now
Flew me to places I'd never been
'Til you put me down, oh”

Supervised learning

Data: $\{x_i, y_i\}_{i=1}^{n}$ w/ $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$

Task: Predict $y^* | x^*$ for some new x^*

Regression

Input : y_i is continuous

Task : learn $\mathbb{E}\, y^* | x^* \rightarrow \hat{y}(x^*)$

Loss : $\ell(\hat{y}(x^*), y^*) = (\hat{y}(x^*) - y^*)^2$

Binary Classification

Input : $y_i \in \{0,1\}$

Task : learn $\max_{y \in \{0,1\}} \mathbb{P}\{y^* = y \mid x^*\}$

Loss : $\ell(\hat{y}(x^*), y^*) = 1\{\hat{y}(x^*) \neq y^*\}$

(Hamming Error)

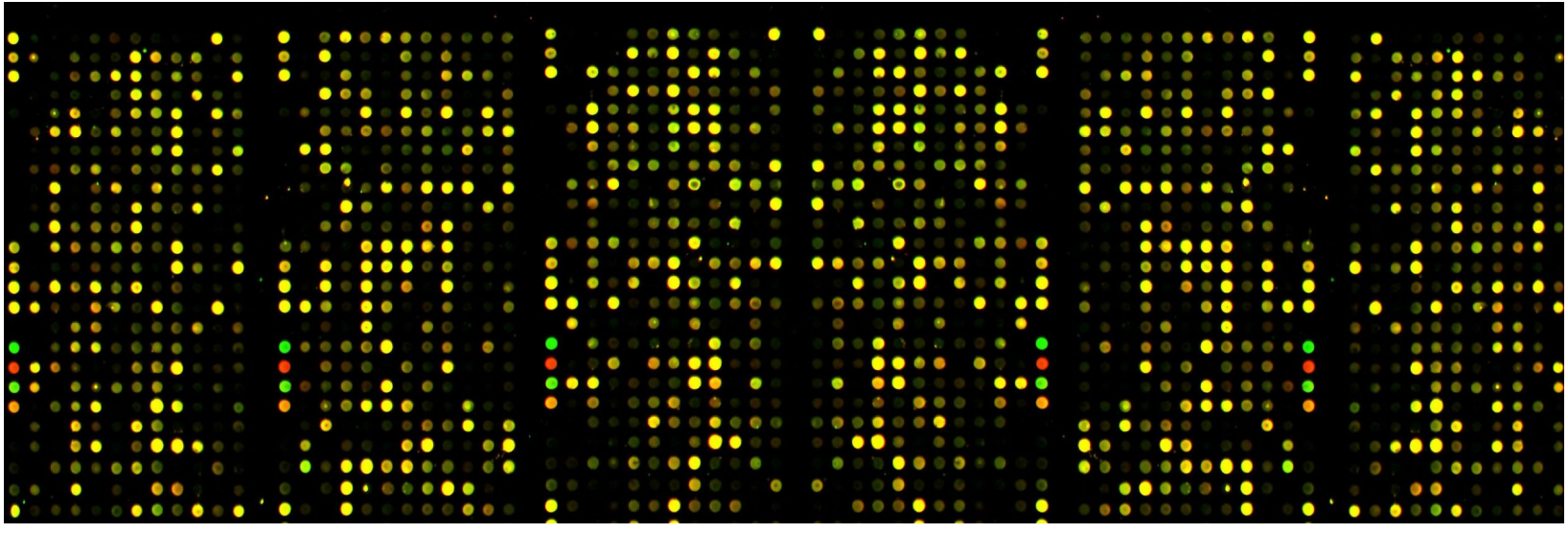
Mortgage Insurance

Mortgage insurer pays the mortgage company if the homeowner defaults on loan. The predictor variables are extracted from quotes, credit inquiry, browser history, etc.

1. Will the homeowner default on the loan?
2. How much will the homeowner pay before they default?

Transcriptome Analysis

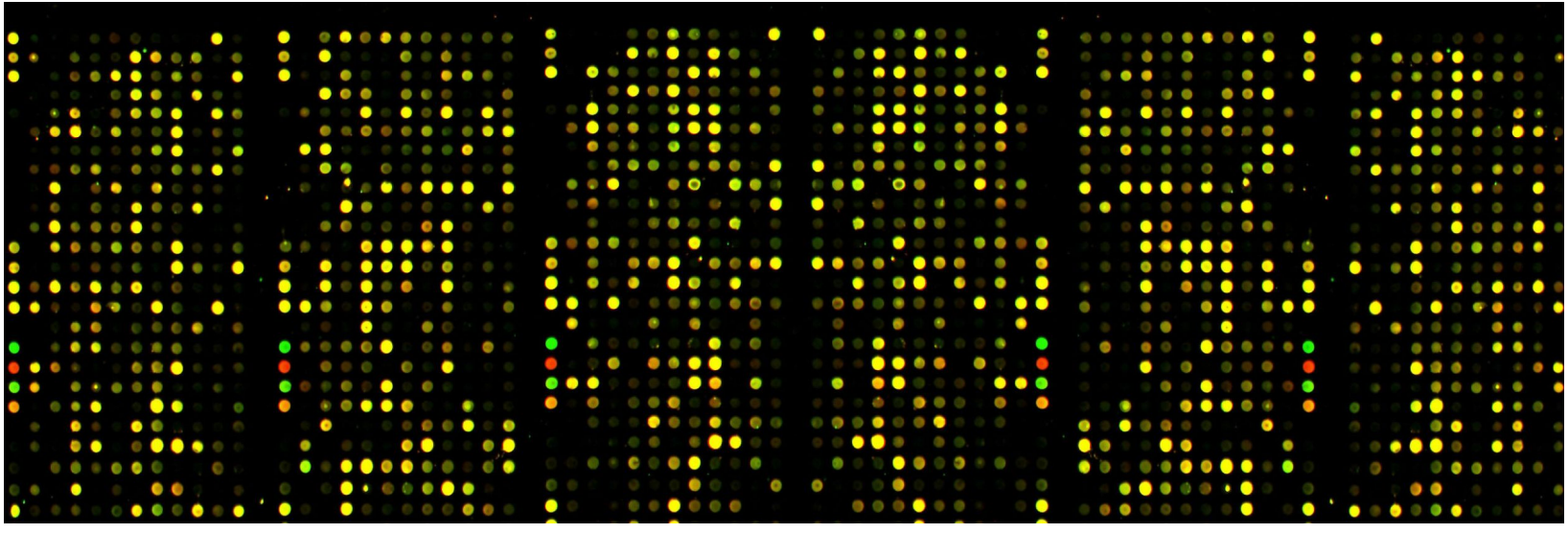
Researcher 1 has RNA expression levels for 1,000 genes in 100 normal and 100 cancerous tissue samples.



Transcriptome Analysis

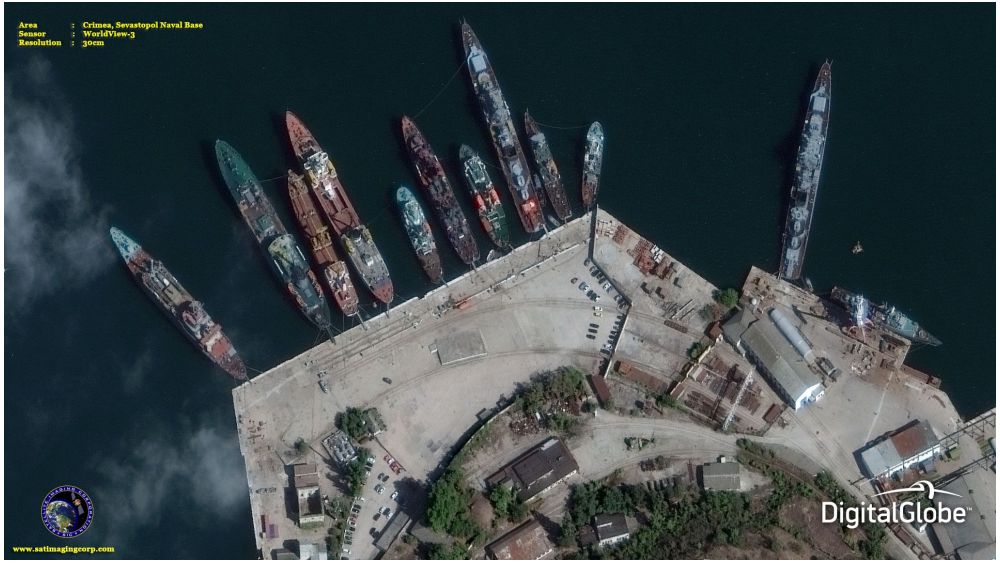
Researcher 1 has RNA expression levels for 1,000 genes in 100 normal and 100 cancerous tissue samples.

They want to know which genes are causing cancer?



US Navy

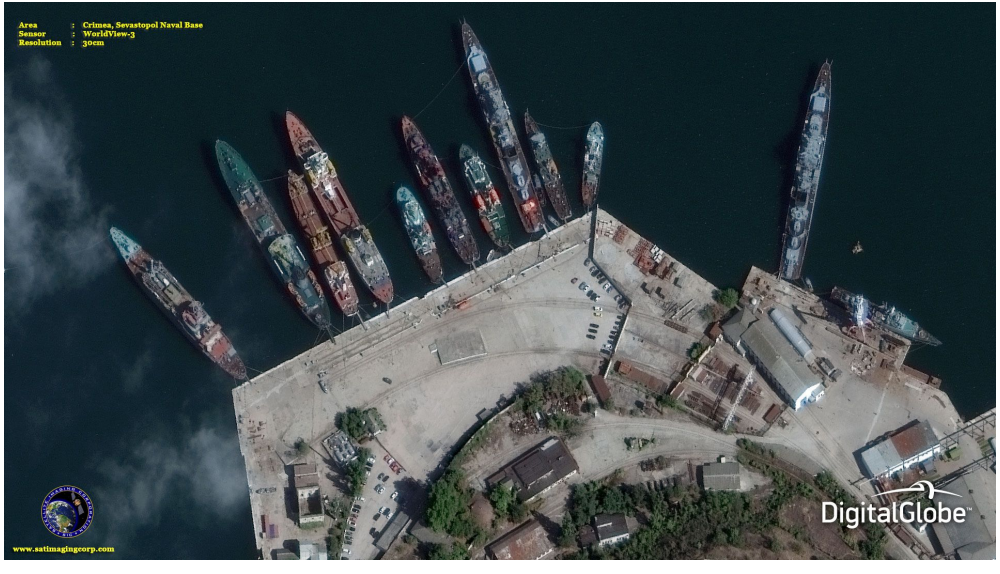
Naval research center wants to use satellite images to detect naval traffic and determine the class of ship.



US Navy

Naval research center wants to use satellite images to detect naval traffic and determine the class of ship.

None of the images are labelled, and the intern will have to label each by hand!



Principles of Statistical Machine Learning

Why learn principles and not just as many methods as possible (in a quarter)?

1. **Machine learning is a field**, not just a course.
2. **Methods change** and adapt to new tasks, computational architectures, data types, but **principles persist**
3. In the meantime, we can learn the most well developed subfield of ML, supervised learning.

Things we won’t cover...

1. Graphical model inference
2. Graphical model learning
3. Dynamic models (hidden Markov models)
4. Matrix completion and recommender systems
5. Reinforcement learning
6. Update learning
7. Distributed optimization (most optimization - take Cho Jui’s class)
8. Multi-armed bandits
9. Topic models
10. Anything Bayesian

Things we won’t cover...

1. Graphical model inference
2. Graphical model learning
3. Dynamic models (hidden Markov models)
4. Matrix completion and recommender systems
5. Reinforcement learning
6. ~~Update learning~~ (not a real thing)
7. Distributed optimization (most optimization - take Cho Jui’s class)
8. Multi-armed bandits
9. Topic models
10. Anything Bayesian

Stages of supervised learning

Consider the supervised learning problem

fit : data $(X, y) \longrightarrow \hat{f}$

predict: $x^* \longrightarrow \hat{y} = \hat{f}(x^*)$

Ordinary least squares (OLS)

$$\hat{f}(x) = x^T \hat{\beta} = \sum_j \hat{\beta}_j x_j$$

fit: $X^T X \hat{\beta} = X^T y$ (normal eqn)

predict: $\hat{f}(x) = x^T \hat{\beta}$

loss: $l(\hat{y}, y^*) = (\hat{y} - y^*)^2$

Empirical (training) risk

For predictor $\hat{f}$, training risk is

$$R_n(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} l(\hat{f}(x_i), y_i)$$

for regression

$$R_n(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{f}(x_i) - y_i)^2$$

Empirical risk minimizer

Empirical risk : $\frac{1}{n}\sum_{i=1}^{n} l(\hat{f}(x_i), y_i) = \frac{1}{n}\sum_{i=1}^{n} (\hat{f}(x_i) - y_i)^2 = R_n(\hat{f})$

Hope is that $\approx \mathbb{E}\,(\hat{f}(x^*) - y^*)^2$ = True Risk

def $\hat{f}$ is the empirical risk minimizer (ERM) within some class, ~~of~~ $\mathcal{F}$, of response functions, if

$$\hat{f} \equiv \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ R_n(\hat{f})$$

OLS is the ERM

ex (linear) Regression $\mathcal{F} = \{ f(x) = x^T\beta : \beta \in \mathbb{R}^p \}$

$$R_n(f) = \frac{1}{n}\sum_{i=1}^{n} (f(x_i) - y_i)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} (x_i^T\beta - y_i)^2$$

OLS is the ERM for linear functions, with square error loss.