Machine Learning for Finance

Neal Parikh

Cornell University

Spring 2018

Introduction

Finding spy planes

US Federal Agents Flew A Secret Spy Plane To Hunt Drug Cartel Leaders In Mexico

Neither the US Marshals Service nor the Mexican government wants to talk about their joint efforts to hunt drug kingpins. But BuzzFeed News spotted a Marshals spy plane circling around the time of a prominent capture in Sinaloa.

Posted on August 3, 2017, at 8:00 a.m.

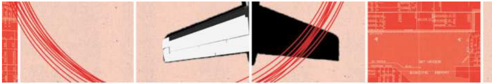
Peter Aldhous
BuzzFeed News Reporter

Karla Zabludovsky
BuzzFeed News Reporter

in August 2017, Buzzfeed News publishes articles finding

- military contractors flying over SF Bay Area
- secret US Marshals plane hunting drug cartel kingpins in Mexico
- Air Force special operations planes flying over US
- ...

Finding spy planes

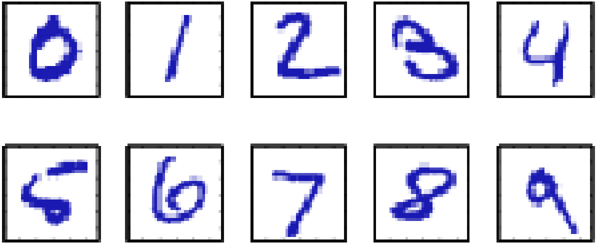


BuzzFeed News Trained A Computer To Search For Hidden Spy Planes. This Is What We Found.

From planes tracking drug traffickers to those testing new spying technology, US airspace is buzzing with surveillance aircraft operated for law enforcement and the military.

1. pull 4 months of flight-tracking data from website Flightradar24
2. extract ‘features’: turning rates, speeds, altitudes, manufacturers
3. train a binary classifier to distinguish between previously identified FBI/DHS planes and not
4. validate

Handwritten digit recognition



Examples

- Adobe (font recognition)
- Amazon (speculative shipping, Kindle browser prefetching)
- American Express (fraud detection, individual credit limits)
- Cheesecake Factory (predict food ingredient demand)
- Microsoft (traffic prediction for Bing maps, Xbox player matching)
- NASA (anomaly detection for aircraft)
- Nest Thermostat (embedded control of smart thermostat)
- Target (market research, individualized product catalogues)
- USPS (handwriting recognition)
- Walmart (inventory, product placement)

What is machine learning?

- no precise technical definition
- usage evolved over time
- ‘classical’ usage is as a sub-discipline of AI research
- but classical thinking had no special commitment to modeling uncertainty

What is machine learning?

- intersection of computer science and statistics
- computationally tractable algorithms that learn from data
- the mathematical foundation of modern AI
- but now also used in a huge variety of other domains

What is machine learning?

- intersection of computer science and statistics
- computationally tractable algorithms that learn from data
- the mathematical foundation of modern AI
- but now also used in a huge variety of other domains

- important for (finance) practitioners to now consider what machine learning is and isn't, and how it relates to existing methods

What is machine learning?

- modern usage: how to build *learning procedures*, *i.e.*, how to use historical data to build a *prediction rule* with complexity automatically adapted to ther problem at hand
- prediction rule: algorithm mapping observable inputs to prediction of unknown quantity (the *response*)
  - in finance, the response is often a security return
- focus is on making predictions

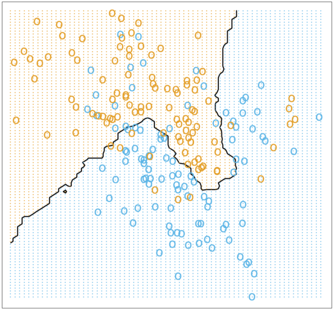
What is machine learning?

In current usage, 'machine learning' means
'modern statistical prediction':
focus on choosing model complexity.

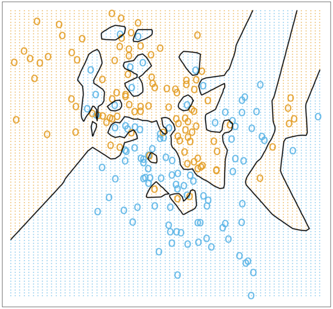
Model complexity

- 'auto-adapted model complexity' is a central idea in machine learning
- model complexity:
  - informally: expressiveness of a prediction rule
  - high-complexity rule can well-approximate rich input-response relationships, while a low-complexity one cannot
- key issue is to perform well on (unobserved) out of sample data
- example: k-nearest neighbors
  - prediction for label of x^{new} is average or majority vote of labels of its k nearest neighbors (in some metric, often ℓ_2)

Model complexity of k-nearest neighbors



$k = 15$



$k = 1$

Source: Hastie, Tibshirani, Friedman, *The Elements of Statistical Learning*

Model complexity

- how to measure model complexity?
- generally done by working with a family of related prediction rules indexed by a *complexity parameter*
  - k-NN: neighbor count k (better, $1/k$)
  - polynomial regression: degree of polynomial
- choosing the prediction rule complexity for a given problem is called *model selection*

Approaches to machine learning

1. (regularized) loss minimization
2. Bayesian methods

these will turn out to be connected in a variety of ways

Regularized loss minimization

many model fitting problems have the form

$$\text{minimize} \quad l(w) + \lambda r(w)$$

- $w \in \mathbf{R}^n$ are the **model parameters** or **weights**
- $l : \mathbf{R}^n \to \mathbf{R}$ is a **loss function** measuring misprediction or lack of fit on training data
- $r : \mathbf{R}^n \to \mathbf{R}$ is a **regularizer** that attempts to improve generalization ability (*e.g.*, by penalizing more complex models)
- $\lambda > 0$ is a **regularization parameter**

A (very) crude history of AI & machine learning

1950s	Dartmouth conferences; chess & checkers; LISP; perceptron
1960s	early foundational & philosophical work; formal logic
1970s	neural networks; AI winter
1980s	expert systems; AI winter

1990s	probabilistic revolution; graphical models; kernel methods
2000s	continuing development; convex optimization
2010s	large-scale & widespread applications; deep learning

Tasks and techniques

supervised learning: predict output value based on inputs

- regression
- classification

unsupervised learning: no outputs, find association among inputs

- clustering
- dimensionality reduction

Machine learning and statistics

(Wasserman; Tibshirani)

statistics	computer science
estimation/fitting	learning
regression/classification	supervised learning
clustering/density estimation	unsupervised learning
data	training sample
covariates	features, inputs
response	outputs
test set performance	generalization ability

Notes on notation

- $\mathbf{Z}$: integers
- $\mathbf{R}$: real numbers
- $\mathbf{R}^{m\times n}$: real $m \times n$ matrices
- $\mathbf{S}^n$: symmetric $n \times n$ matrices

- $\mathbf{R}_+$ is nonnegative orthant, $\mathbf{R}_{++}$ is positive orthant
- if $x \in \mathbf{R}^n$, then $x \in \mathbf{R}^n_+$ ($\mathbf{R}^n_{++}$) also written $x \succeq 0$ ($x \succ 0$)
- $\mathbf{S}^n_+$ is positive semidefinite, $\mathbf{S}^n_{++}$ is positive definite matrices
- if $X \in \mathbf{S}^n$, then $X \in \mathbf{S}^n_+$ ($\mathbf{S}^n_{++}$) also written $X \succeq 0$ ($X \succ 0$)

Notes on notation

- $[n]$, where $n \in \mathbf{Z}_+$, is $\{1, 2, \ldots, n\}$
- $[\text{predicate}]$ is indicator function (Iverson bracket), *e.g.*,

$$[z \leq 4] = \begin{cases} 1 & z \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

- $\mathbf{1}$ is vector of ones, so $\mathbf{1}^T x = \sum_{i=1}^n x_i$
- $e_1, \ldots, e_n$ is standard basis for $\mathbf{R}^n$
- will sometimes use overline for averages; given $x_1, \ldots, x_N$, define

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Acknowledgements

- Andrew Ng (Stanford)
- Stephen Boyd (Stanford) & Lieven Vandenberghe (UCLA)
- Michael Jordan (Berkeley)
- Jon McAuliffe (Voleon, Berkeley)
- Daphne Koller (Stanford)