Introduction to Probabilistic Machine Learning with PyMC3

Bayesian Data Science DC Meetup

Data Science & Cybersecurity Meetup

Daniel Emaasit

Data Scientist

Haystax Technology

April 26, 2018

Materials

Download slides & code: bit.ly/intro-pml-dc

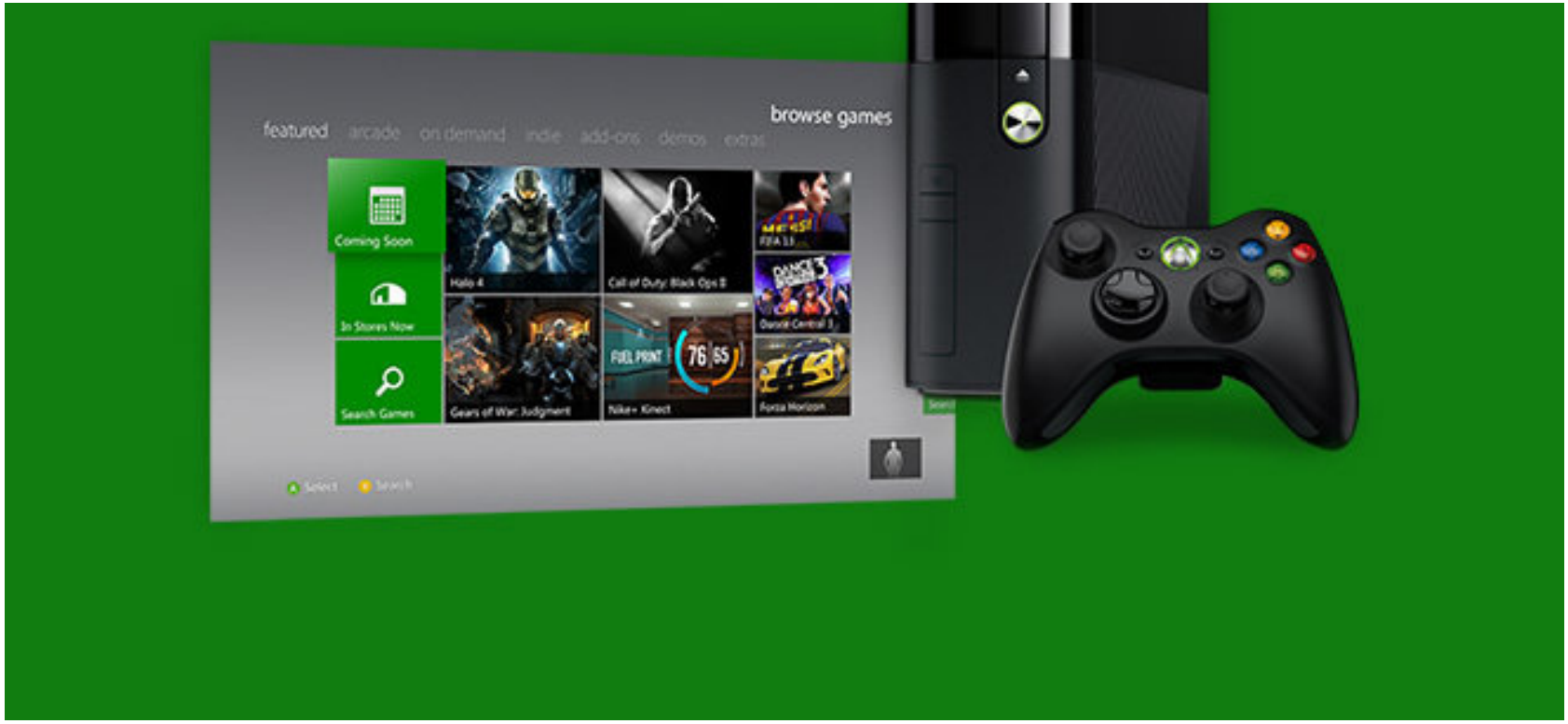
Application (1/3)

- Optimizing expensive functions in autonomous vehicles (using Bayesian optimizatio

Schneider et al., 2016. (Uber ATG).

Application (2/3)

- Probabilistic approach to ranking & matching gamers.



Herbrich et al., 2009. (Microsoft Research).

Application (3/3)

- Supplying internet to remote areas (using Gaussian processes).

Wired magazine, accessed 2016. (Google Project Loon)

Media Attention

CADE METZ BUSINESS 02.03.17 07:00 AM

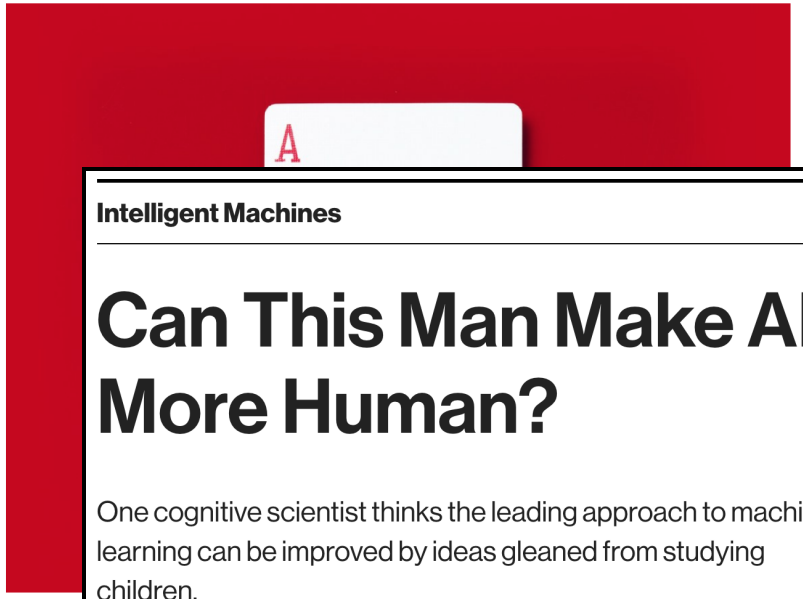
AI IS ABOUT TO LEARN MORE LIKE HUMANS—WITH A LITTLE UNCERTAINTY

CADE METZ BUSINESS 12.05.16 08:00 AM

UBER BUYS A MYSTERIOUS STARTUP TO MAKE ITSELF AN AI COMPANY

CADE METZ BUSINESS 02.01.17 07:00 AM

INSIDE LIBRATUS, THE POKER AI THAT OUT-BLUFFED THE BEST HUMANS



Intelligent Machines

Can This Man Make AI More Human?

One cognitive scientist thinks the leading approach to machine learning can be improved by ideas gleaned from studying children.

by Will Knight December 17, 2015

Like any proud father, Gary Marcus is only too happy to talk about the latest achievements of his two-year-old son. More unusually, he believes that the way his toddler learns and reasons may hold the key to making machines much more intelligent.

Academics flocking to Industry

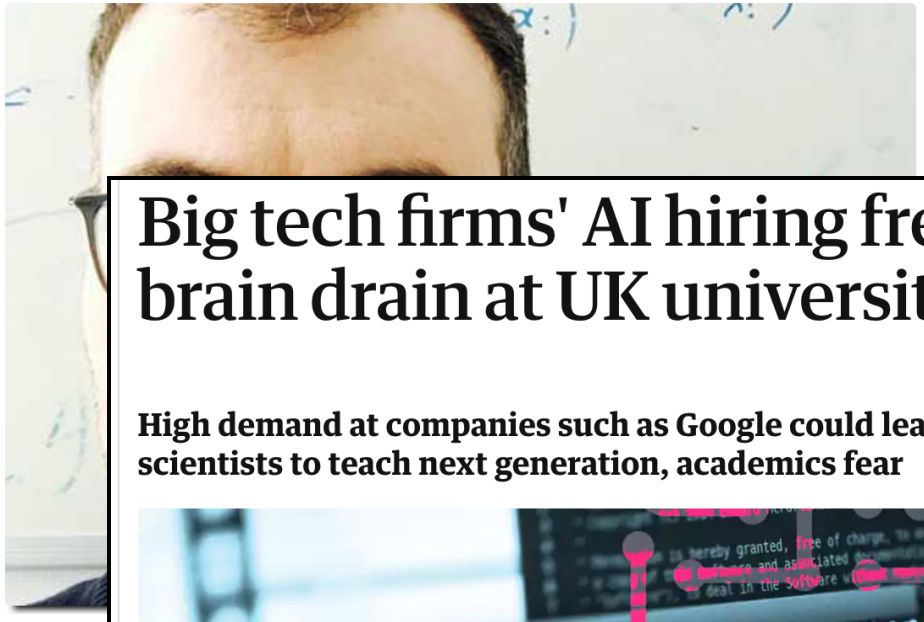
Uber appoints machine learning professor Zoubin Ghahramani as chief scientist 3 months after acquiring his startup

PAUL SAWERS @PSAWERS MARCH 15, 2017 4:49 AM

Above: Professor Zoubin Ghahramani

6 March, 2017 - 12:51 By **Kate Sweeney**

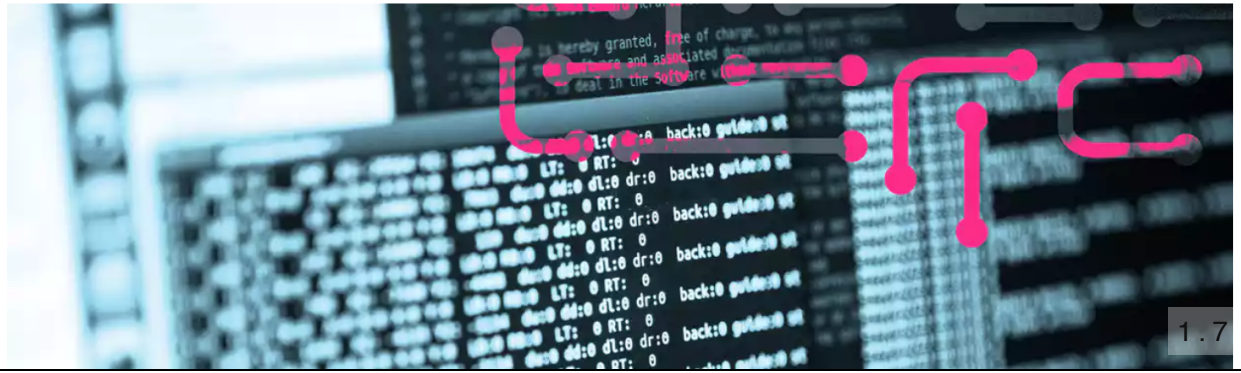
In-demand scientist opts for Cambridge AI role



Researc
Cambri
scientist

Big tech firms' AI hiring frenzy leads to brain drain at UK universities

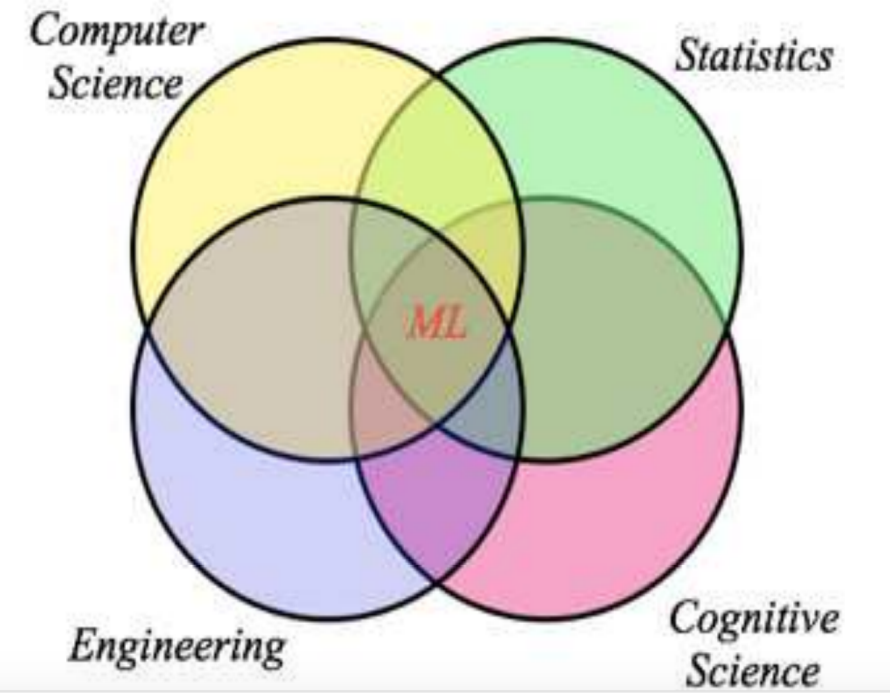
High demand at companies such as Google could leave fewer talented scientists to teach next generation, academics fear



Intro to Probabilistic Machine Learning

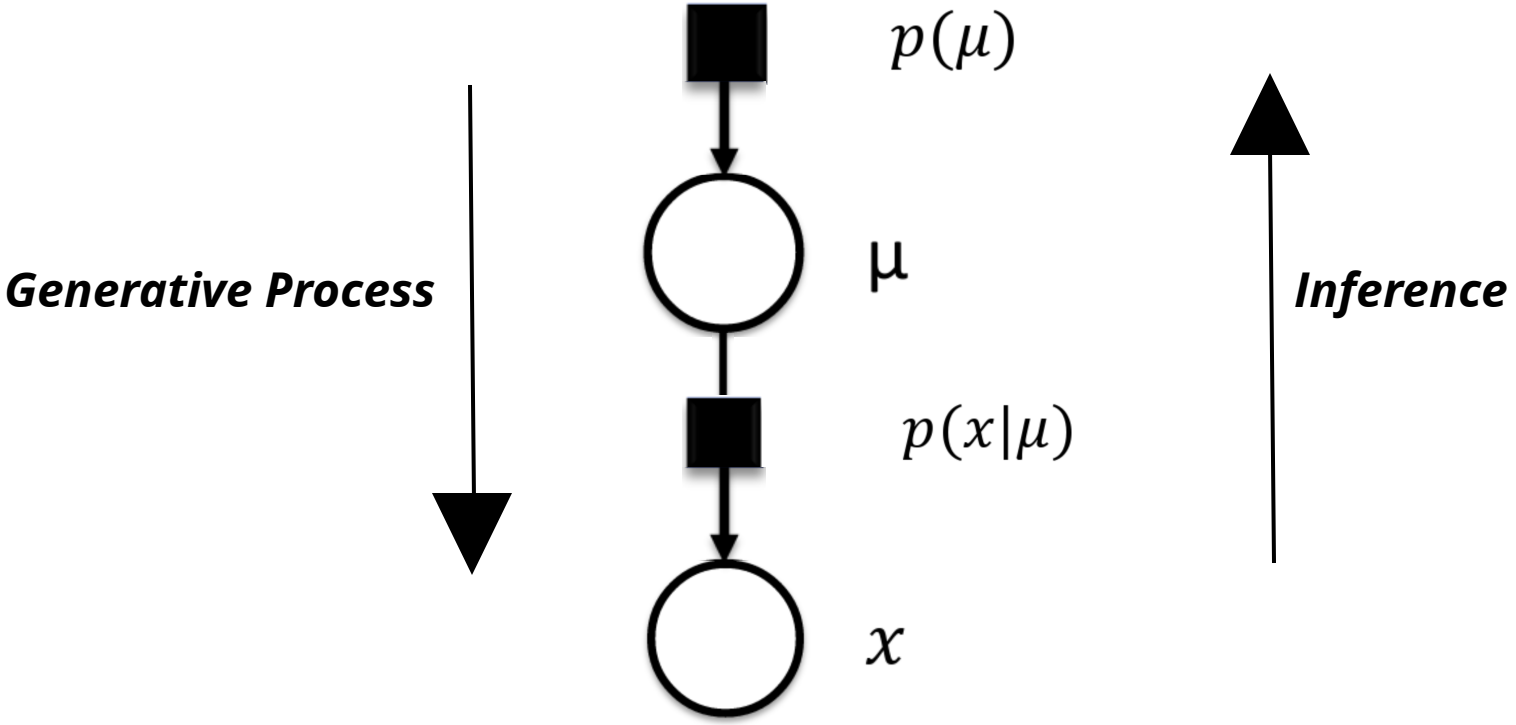
ML: A Probabilistic Perspective (1/3)

- Probabilistic ML:
  - An Interdisciplinary field that develops both the mathematical foundations and practical applications of systems that learn **models** of data. (*Ghahramani, 2018*)



ML: A Probabilistic Perspective (2/3)

- A Model:
  - A model describes data that one could observe from a system (*Ghahramani, 2014*)



  - Use the mathematics of **probability theory** to express all forms of uncertainty

ML: A Probabilistic Perspective (3/3)

Everything follows from two simple rules:

Sum rule: $P(x) = \sum_y P(x, y)$

Product rule: $P(x, y) = P(x)P(y|x)$

$$P(\theta \mid y, x) = \frac{P(y|\theta, x)\, P(\theta)}{P(y|x)}$$

where:

- θ = parameters e.g. coefficients

$p(\theta \mid \mathbf{y}, \mathbf{x})$ = posterior over the parameters, given observed data

$p(\theta)$ = prior over the parameters

$p(\mathbf{y} \mid \theta, \mathbf{x})$ = likelihood given the covariates & parameters

$p(\mathbf{y} \mid \mathbf{x})$ = data distribution to ensure normalization

Probabilistic ML Vs Traditional ML

	Algorithmic ML	**Probabilistic ML**
Examples	K-Means, Random Forest	GMM, Gaussian Process
Specification	Model + Algorithm combined	Model & Inference separate
Unknowns	Parameters	Random variables
Inference	Optimization (MLE)	Bayes (MCMC, VI)
Regularization	Penalty terms	Priors
Solution	Best fitting parameter	Full posterior distribution

Limitations of deep learning

- Deep learning systems give amazing performance on many benchmark tasks but they are generally:
  - very **data hungry** (e.g. often millions of examples)
  - very **compute-intensive** to train and deploy (cloud GPU resources)
  - poor at representing **uncertainty**
  - **easily fooled** by adversarial examples
  - **finicky to optimize**: choice of architecture, learning procedure, etc, require expert knowledge and experimentation
  - uninterpretable **black-boxes**, lacking in transparency, difficult to trust

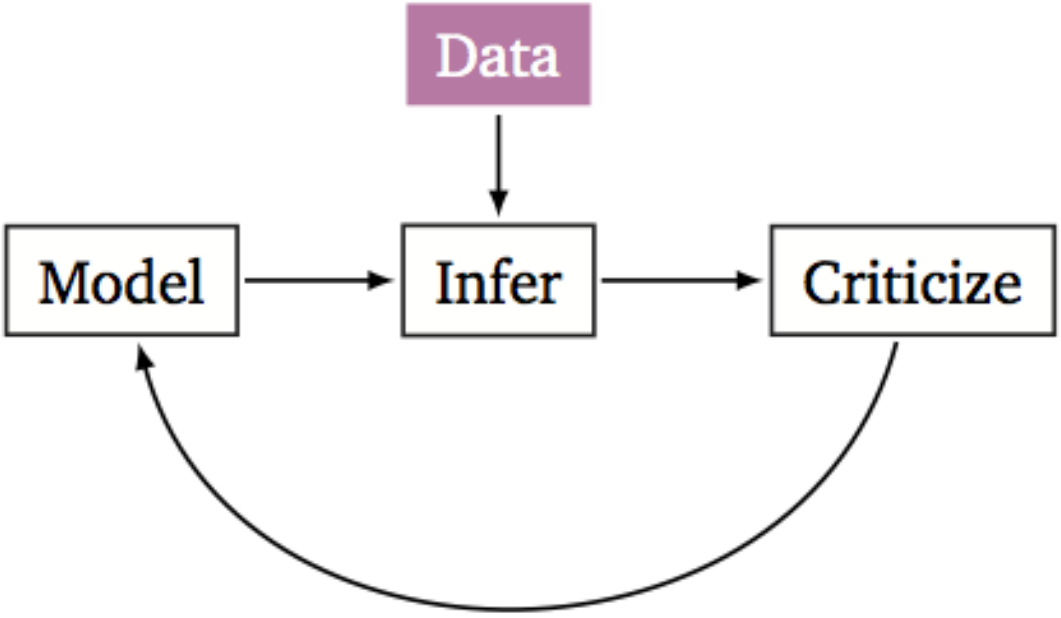
Probabilistic Programming

Probabilistic Programming (1/2)

- Probabilisic Programming (PP) Languages:
  - Software packages that take a model and then automatically generate inference routines (even source code!) e.g Pyro, Stan, Infer.Net, PyMC3, TensorFlow Probability, etc.

Probabilistic Programming (2/2)

- Steps in Probabilisic ML:
  - Build the model (Joint probability distribution of all the relevant variables)
  - Incorporate the observed data
  - Perform inference (to learn distributions of the latent variables)



Box's Loop (Image credit: http://dustintran.com/)

Demo

bit.ly/intro-pml-dc

Resources to get started

- PyMC3 documents

- Winn, J., Bishop, C. M., Diethe, T. (2015). Model-Based Machine Learning. Microsoft Research Cambridge.

- R. McElreath (2012) Statistical Rethinking: A Bayesian Course with Examples in R and Stan (& PyMC3 & brms too)

- Probabilistic Programming and Bayesian Methods for Hackers: Fantastic book with many applied code examples.

Thank You!

Appendix

Limitations of deep learning

- Deep learning systems give amazing performance on many benchmark tasks but they are generally:
  - very **data hungry** (e.g. often millions of examples)
  - very **compute-intensive** to train and deploy (cloud GPU resources)
  - poor at representing **uncertainty**
  - **easily fooled** by adversarial examples
  - **finicky to optimize**: choice of architecture, learning procedure, etc, require expert knowledge and experimentation
  - uninterpretable **black-boxes**, lacking in transparency, difficult to trust