Hide and Seek
with
Hidden Markov Models

SKA AFRICA
SQUARE KILOMETRE ARRAY

Arun Aniyan

SKA South Africa

Hide and Seek
with
~~Hidden Markov Models~~

Probabilistic Graphic Models

SKA AFRICA
SQUARE KILOMETRE ARRAY

Arun Aniyan

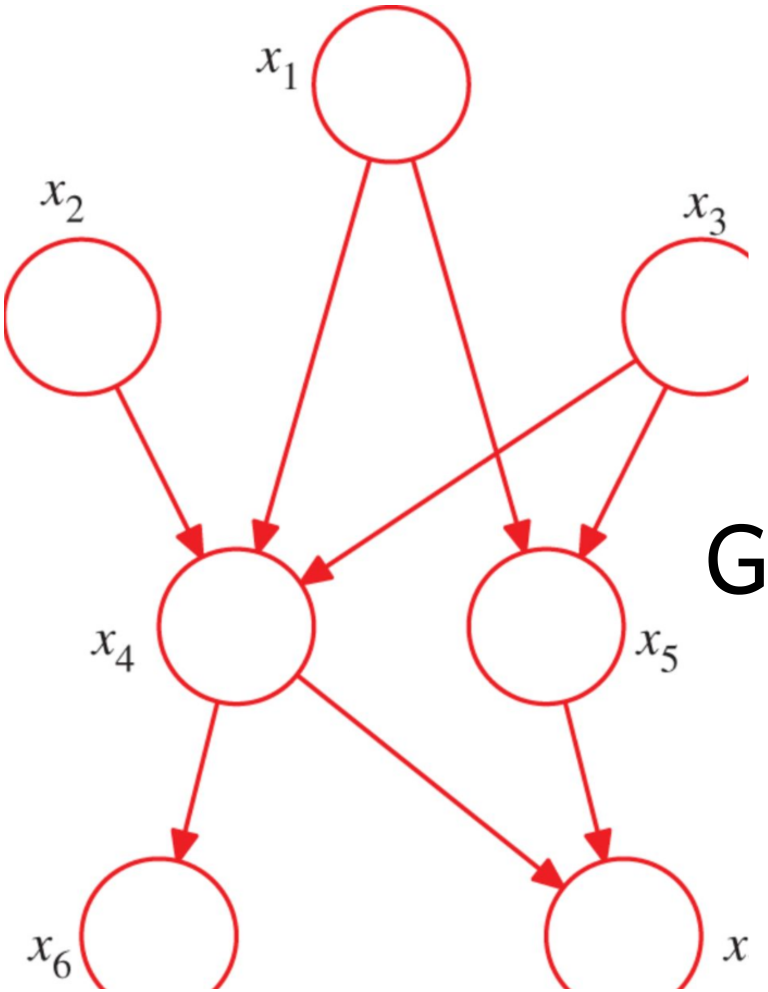
SKA South Africa

No neural nets or deep funda ?

.... so having a notebook or lively debate on the best way to do neural network magic would be cool too. - Thuso

Will talk about neural nets

...... eventually



Graphical Models

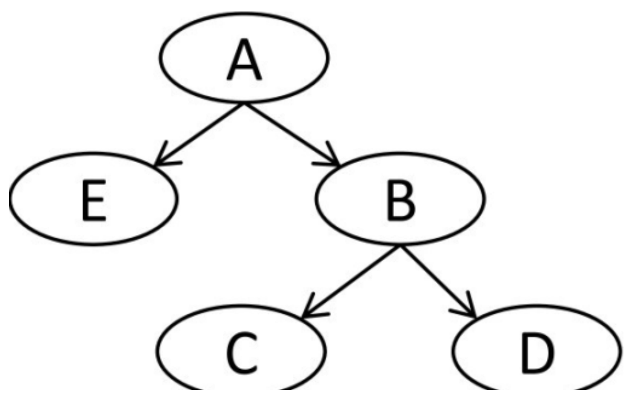
Uses Graph theory for underlying computational machinery

Probabilistic Graphical Models are a marriage of Graph Theory with Probabilistic Methods

Bayesian Networks

Bayesian Networks

Directed Acyclic Graphs (DAG)



Bayesian Networks

Directed Acyclic Graphs (DAG)

$+$

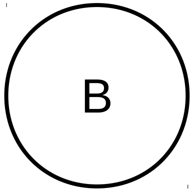
Conditional Probability

Bayesian Networks

Exam Fear

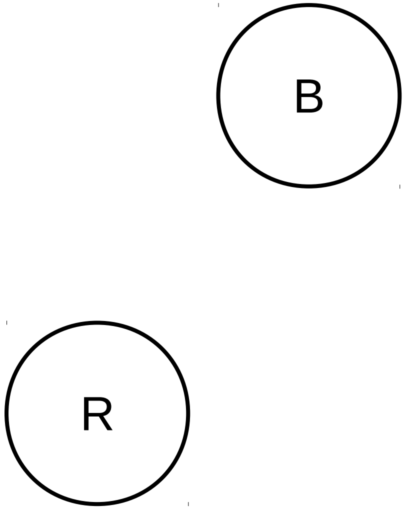
Bayesian Networks

Exam Fear



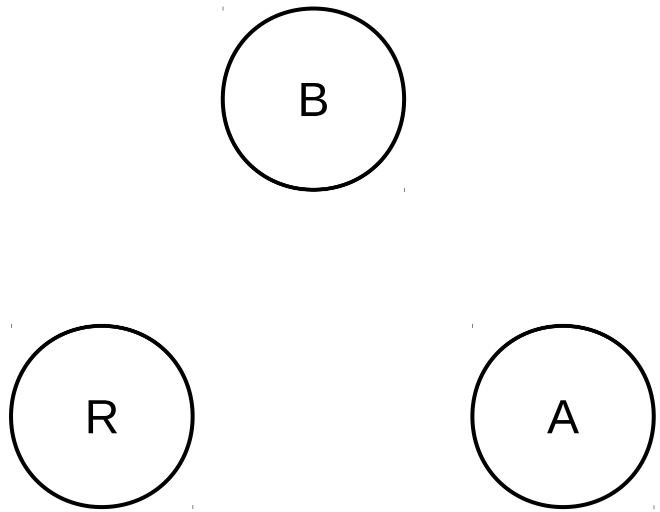
Bayesian Networks

Exam Fear



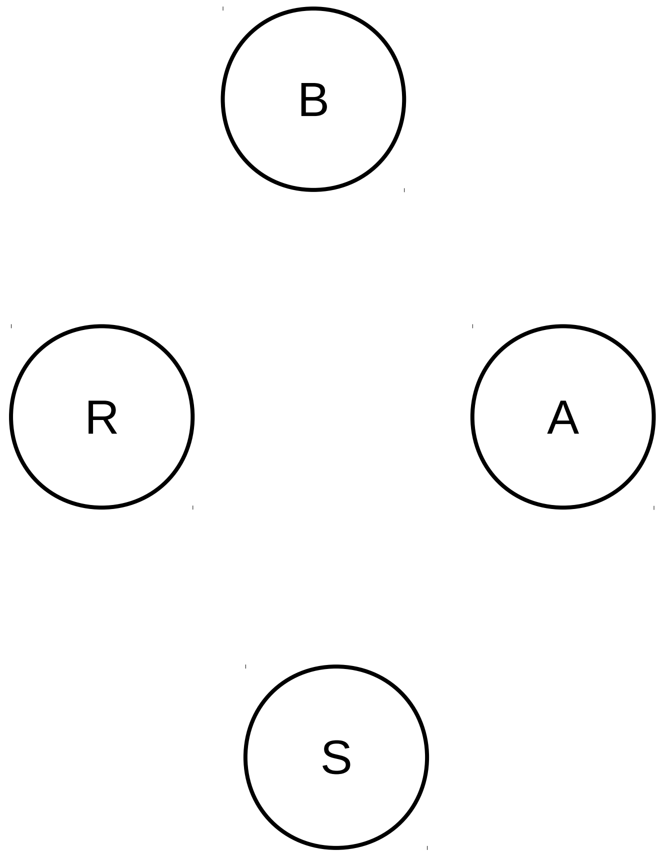
Bayesian Networks

Exam Fear



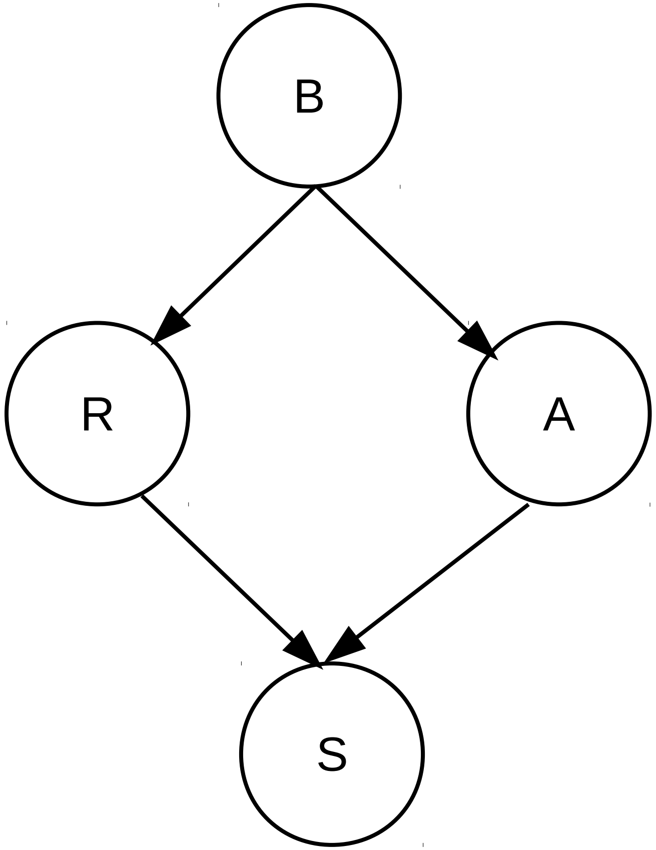
Bayesian Networks

Exam Fear



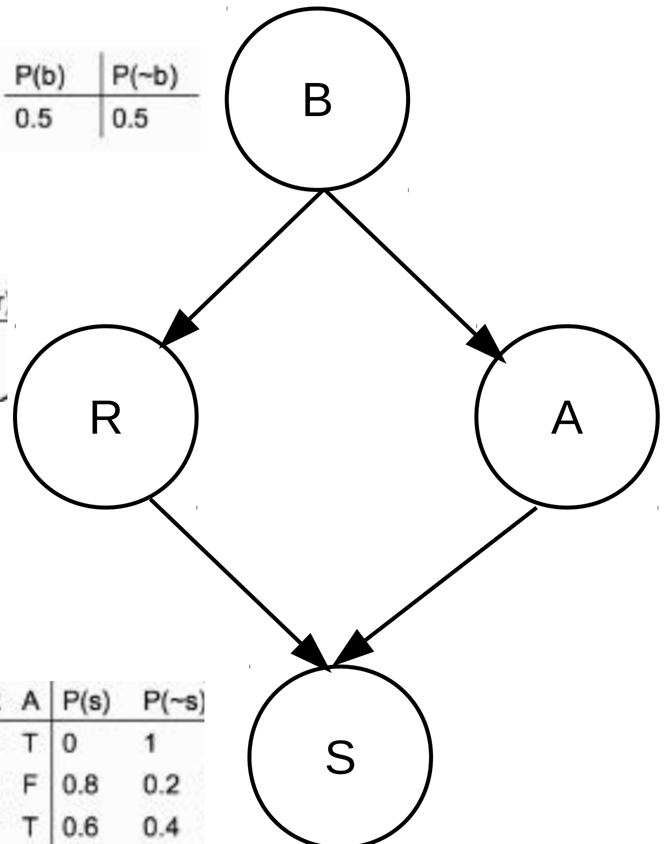
Bayesian Networks

Exam Fear



Bayesian Networks

Exam Fear



P(b)	P(~b)
0.5	0.5

B	P(r)	P(~r)
T	0.3	0.7
F	0.8	0.2

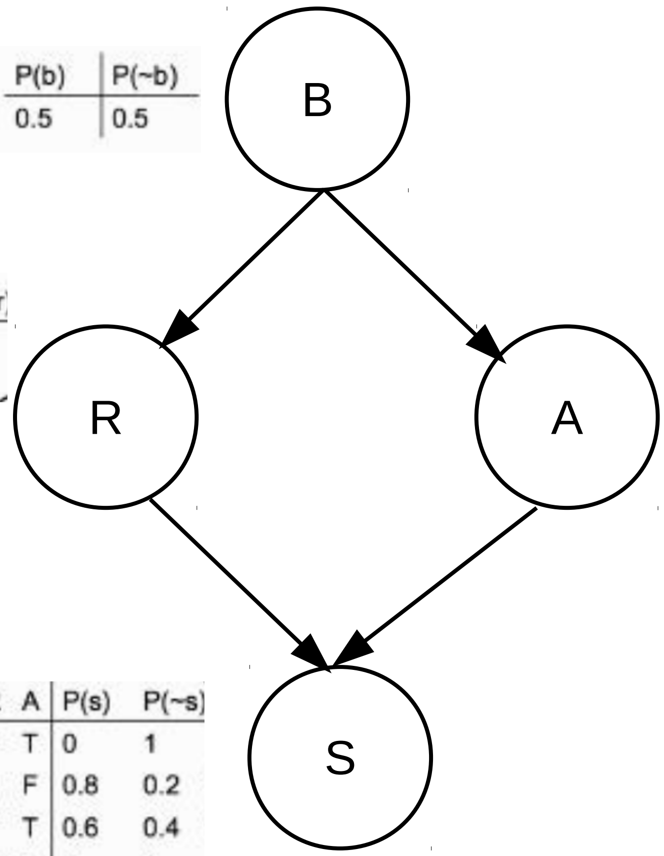
B	P(a)	P(~a)
T	0.1	0.9
F	0.5	0.5

R	A	P(s)	P(~s)
T	T	0	1
T	F	0.8	0.2
F	T	0.6	0.4
F	F	1	0

Bayesian Networks

Exam Fear

P(b)	P(~b)
0.5	0.5



B	P(r)	P(~r)
T	0.3	0.7
F	0.8	0.2

B	P(a)	P(~a)
T	0.1	0.9
F	0.5	0.5

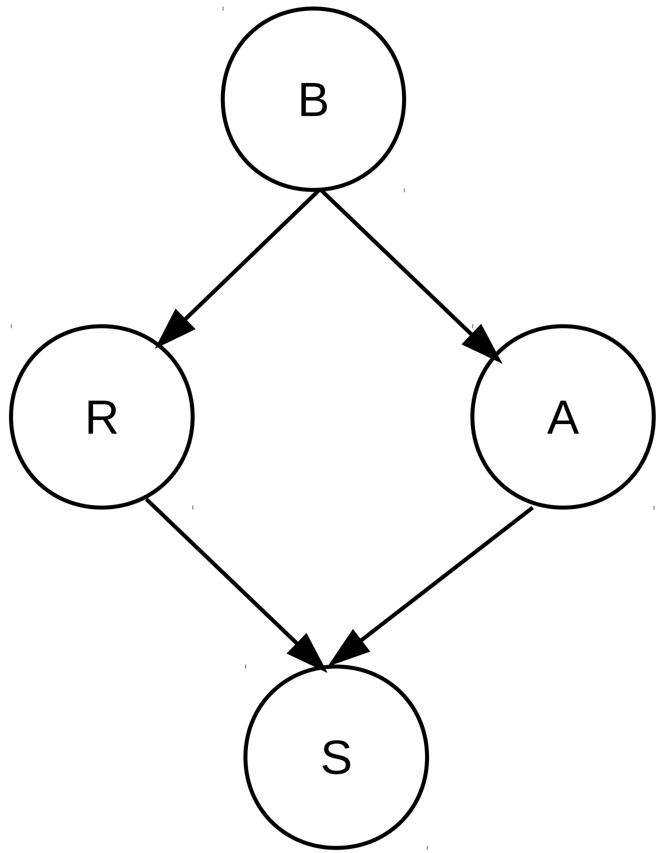
R	A	P(s)	P(~s)
T	T	0	1
T	F	0.8	0.2
F	T	0.6	0.4
F	F	1	0

Bayes Rule

Bayesian Networks

Exam Fear

Prediction

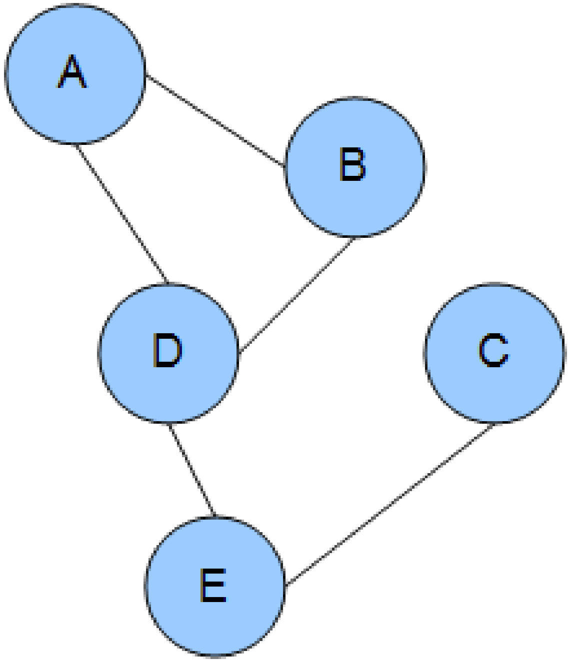


Diagnosis

Markov Random Fields (MRF)

Markov Random Fields (MRF)

Nodes are Conditionally Independent



Hidden Markov Models

Hidden Markov Models

Glecoph

Hidden Markov Models

Glecoph

Hidden Markov Models

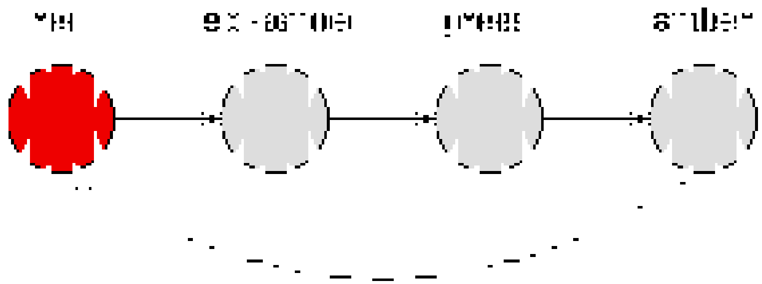
A Markov process is a random process in which the future is independent of the past, given the present. Thus, Markov processes are the natural stochastic analogs of the deterministic processes described by differential and difference equations. They form one of the most important classes of random processes.

Hidden Markov Models

A Markov process is a random process in which the future is independent of the past, given the present. Thus, Markov processes are the natural stochastic analogs of the deterministic processes described by differential and difference equations. They form one of the most important classes of random processes.

- Number of outcomes / states is finite
- Output at any stage depends only on the previous stage
- Probabilities are constant over time

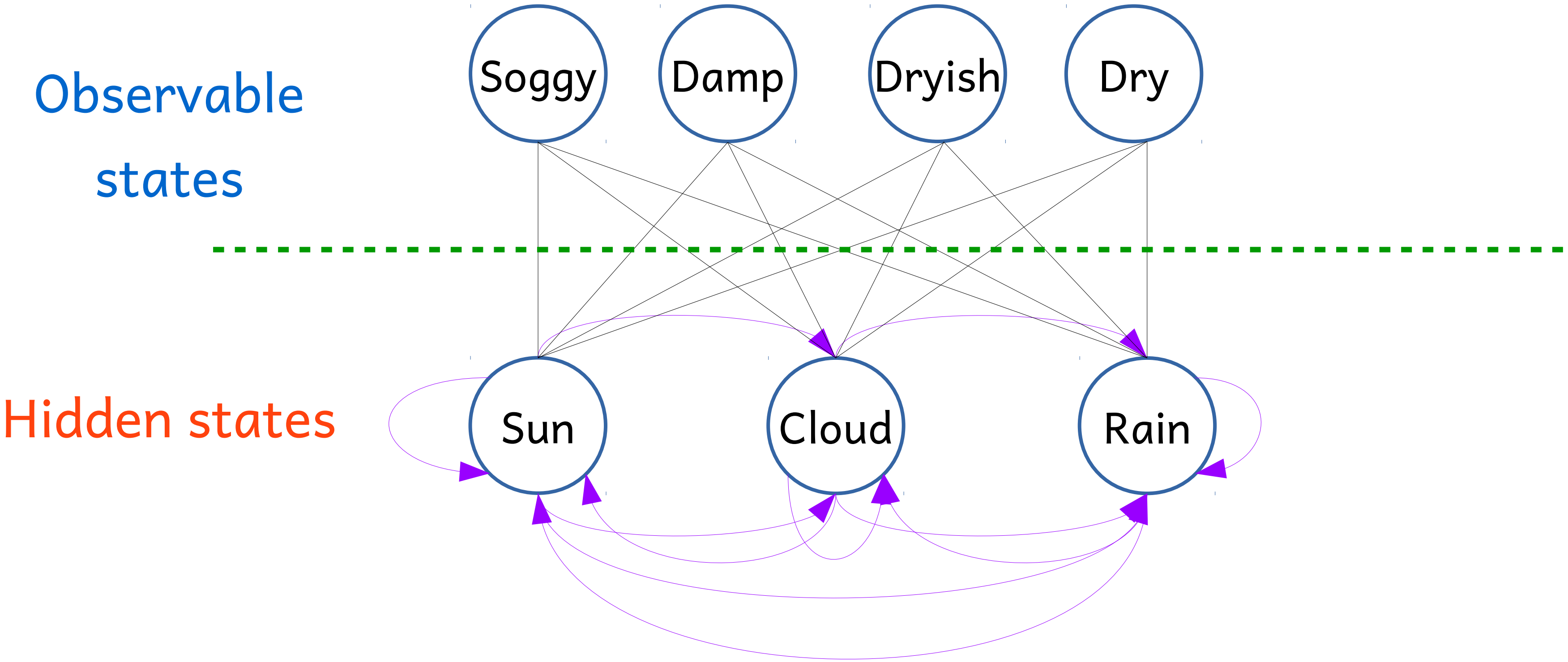
Hidden Markov Models



- Number of outcomes / states is finite
- Output at any stage depends only on the previous stage
- Probabilities are constant over time

Hidden Markov Models

Weather



Hidden Markov Models

Weather

$$
\begin{array}{c c}
 & \begin{array}{cccc} \text{Dry} & \text{Dryish} & \text{Damp} & \text{Soggy} \end{array} \\
\begin{array}{c} \text{Sun} \\ \text{Cloud} \\ \text{Rain} \end{array} &
\begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.5 \end{pmatrix}
\end{array}
$$

Hidden Markov Models

Forward Algorithm : Calculates the probability of (hidden) state at a particular time instant.

Viterbi Algorithm : Produces maximum likelihood estimates of successive states

Forward-Backward Algorithm (Baum-Welch Algo): Calculate MLE parameters from data.

Hidden Markov Models

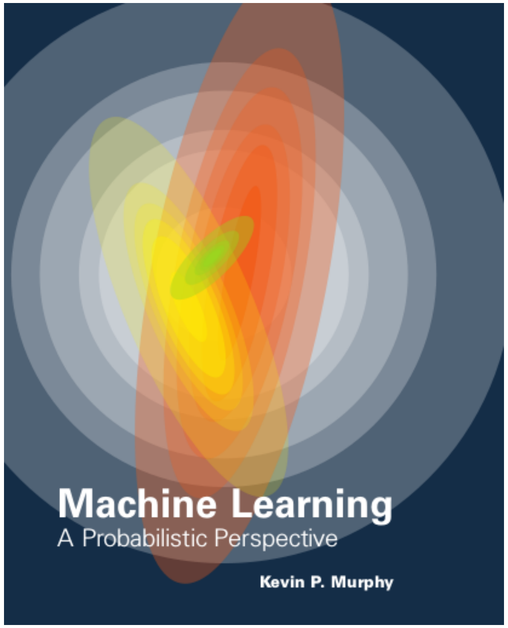
Kalman Filter and Particle Filters

Kalman Filter and Particle Filters

- Recursive Bayesian Filters
- Kalman is for linear systems with gaussian noise
- Particle filters work for systems whic is not linear and non gaussian

Probabilistic Graphic Models (PGM)

- Can represent conditional dependence or independence
- Powerful way of understanding /explaining underlying mechanics
- Most of the ML algorithms can be expressed with PGM !



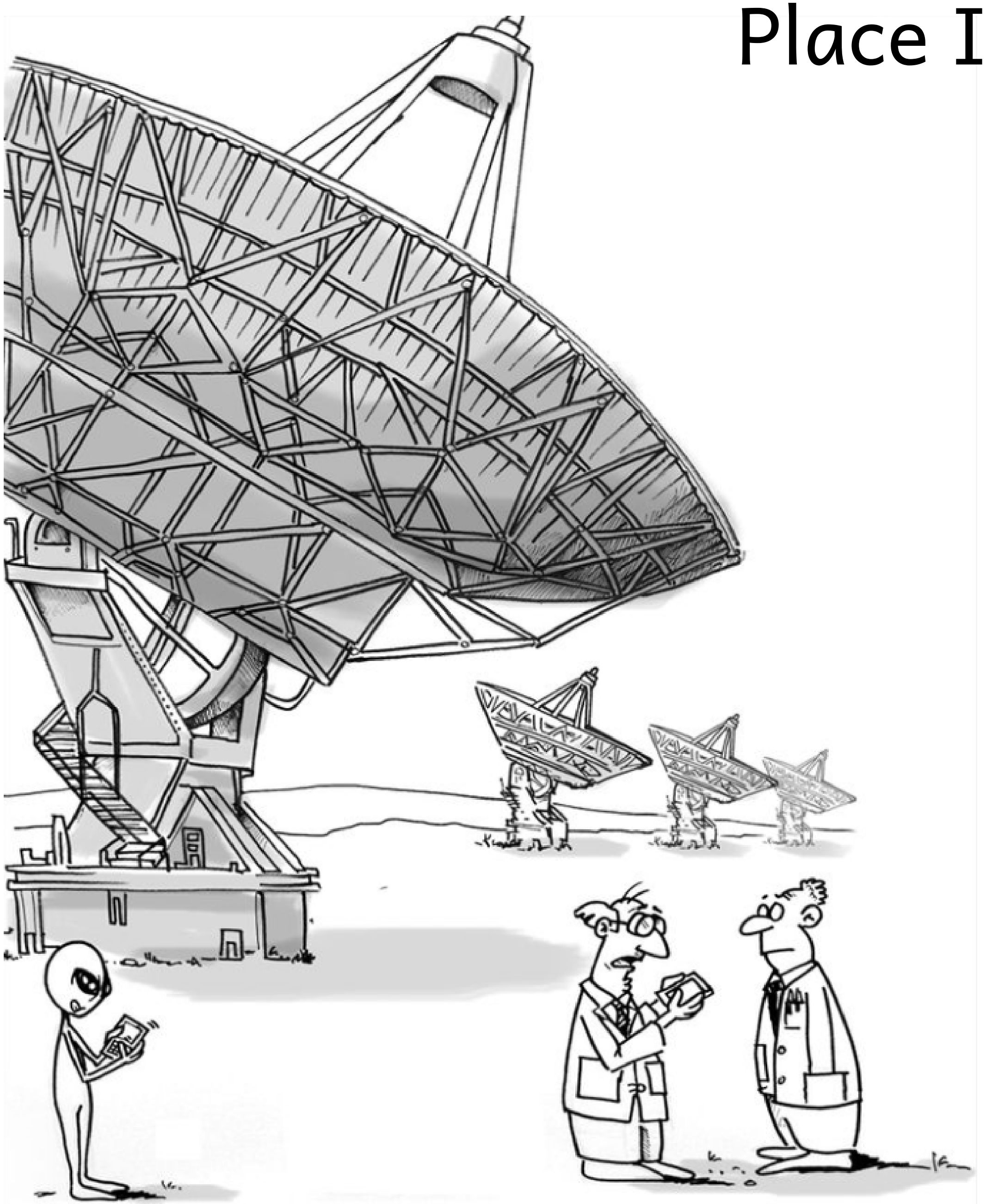
Probabilistic Graphic Models (PGM)

Fully Bayesian

Probabilistic Graphic Models

Neural Nets & Probabilistic Graphic Models

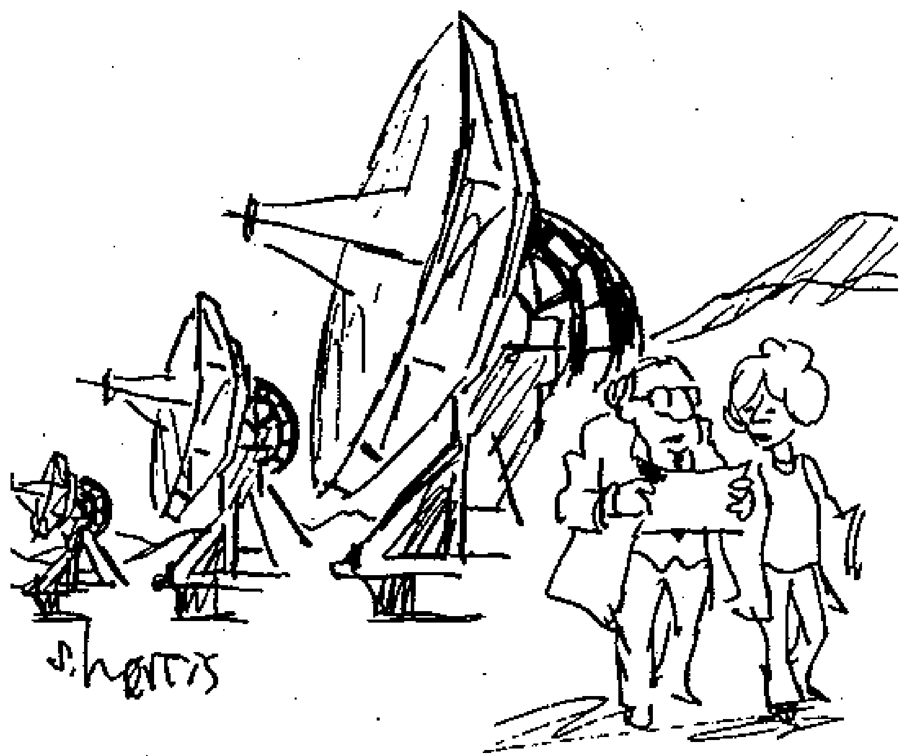
Place I work



“The aliens want us to stop spamming them.”

We are trying to find the known knowns, the known unknowns and the unknown unknowns.

Place I work



“ We’ve discovered a massive dust and gas cloud which is either the beginning of a new star or just a hell of a lot of dust and gas.”

Why does deep and cheap learning work so well?

Henry W. Lin and Max Tegmark
Dept. of Physics, Harvard University, Cambridge, MA 02138 and
Dept. of Physics & MIT Kavli Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

(Dated: August 31, 2016)

We show how the success of deep learning depends not only on mathematics but also on physics: although well-known mathematical theorems guarantee that neural networks can approximate arbitrary functions well, the class of functions of practical interest can be approximated through "cheap learning" with exponentially fewer parameters than generic ones, because they have simplifying properties tracing back to the laws of physics. The exceptional simplicity of physics-based functions hinges on properties such as symmetry, locality, compositionality and polynomial log-probability, and we explore how these properties translate into exceptionally simple neural networks approximating both natural phenomena such as images and abstract representations thereof such as drawings. We further argue that when the statistical process generating the data is of a certain hierarchical form prevalent in physics and machine learning, a deep neural network can be more efficient

"Why Should I Trust You?" Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

ABSTRACT

Despite widespread adoption, machine learning models remain mostly black boxes. Understanding the reasons behind predictions is, however, quite important in assessing *trust*, which is fundamental if one plans to take action based on a prediction, or when choosing whether to deploy a new model. Such understanding also provides insights into the model, which can be used to transform an untrustworthy model or prediction into a trustworthy one.

In this work, we propose LIME, a novel explanation technique that explains the predictions of *any* classifier in an interpretable and faithful manner, by learning an interpretable model locally around the prediction. We also propose a method to explain models by presenting representative indi-

how much the human understands a model's behaviour, as opposed to seeing it as a black box.

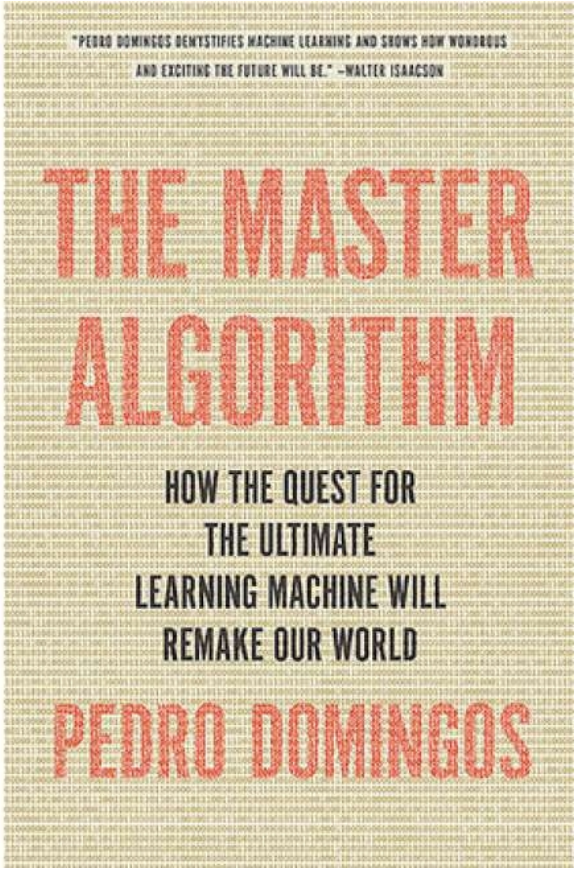
Determining trust in individual predictions is an important problem when the model is used for decision making. When using machine learning for medical diagnosis [6] or terrorism detection, for example, predictions cannot be acted upon on blind faith, as the consequences may be catastrophic.

Apart from trusting individual predictions, there is also a need to evaluate the model as a whole before deploying it "in the wild". To make this decision, users need to be confident that the model will perform well on real-world data, according to the metrics of interest. Currently, models are evaluated using accuracy metrics on an available validation dataset. However, real-world data is often significantly different, and

Neural Network

- The model (architecture)
- Optimizer algorithm

Neural Network



Combining probabilistic models with neural nets will be the best way to understand them.

Neural Network

UNIVERSITY *of* WASHINGTON

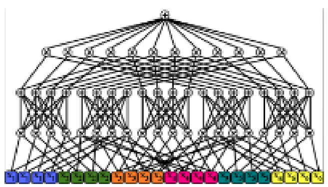
ABOUT US | CONTACT US | MY CSE | INTERNAL

Custom Search

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

News & Events | People | Education | Research | Current Students | Prospective Students | Faculty Candidates | Alumni | Industry Affiliates | Support CSE

Tractable Deep Learning



In machine learning, as throughout computer science, there is a tradeoff between expressiveness and tractability. On the one hand, we need powerful model classes to capture the richness and complexity of the real world. On the other, we need inference in those models to remain tractable, otherwise their potential for widespread practical use is limited. Deep learning can induce powerful representations, with multiple layers of latent variables, but these models are generally intractable. We are developing new classes of similarly expressive but still tractable models, including sum-product networks and tractable Markov logic. These models capture both class-subclass and part-subpart structure in the domain, and are in some aspects more expressive than traditional graphical models like Bayesian networks and Markov random fields. Research includes designing representations, studying their properties, developing efficient algorithms for learning them, and applications to challenging problems in natural language understanding, vision, and other areas.

Awards

- **NIPS 2012 Outstanding Student Paper:** Discriminative Learning of Sum-Product Networks
- **UAI 2011 Best Paper:** Sum-Product Networks: A New Deep Architecture
- **EMNLP 2009 Best Paper:** Unsupervised Semantic Parsing

Neural Network

Marrying Graphical Models with Deep Learning

Special Theme: Machine Learning   Hits: 4457

by Max Welling (University of Amsterdam)

In our research at the University of Amsterdam we have married two types of models into a single comprehensive framework which we have called “Variational Auto Encoders”. The two types of models are: 1) generative models where the data generation process is modelled, and 2) discriminative models, such as deep learning, where measurements are directly mapped to class labels.

Deep learning is particularly successful in learning powerful (e.g., predictive/ discriminative) features from raw, unstructured sensor data. Deep neural networks can effectively turn raw data streams into new representations that represent abstract, disentangled and semantically meaningful concepts. Based on these, a simple linear classifier can achieve the state of the art. But to learn them one needs very large quantities of annotated data. They are flexible input-output mappings but do not incorporate a very sophisticated inductive bias about the world. An important question is how far will this take us?

If we are asked to analyse a scene depicted in an image we seek a story that can explain the things we see in the image. Yes, there is a fast feedforward pipeline that quickly segments out the objects and classifies them into object classes. But when you need to truly understand a scene you will try to infer a story about which events caused other events, which in turn led to the image you are looking at. This causal story is also a powerful tool to predict how the events may unfold into the future.

So, to understand and reason about the world we need to find its causal atoms and their relationships. Now this is precisely what Bayesian networks [1] were intended to do. Each random variable connects to other random variables and their directed relations model their causal relationships. (Bayesian networks do not necessarily represent the causal relationships, but an extension called “structural equation models” does.) Another key advantage of interpretable models like Bayesian networks is that they can express our expert knowledge. If we know X causes Y then we can simply hard-code that relation into the model. Relations that we do not know will need to be learned from the data. Incorporating expert knowledge (e.g., the laws of physics) into models is the everyday business of scientists. They build sophisticated simulators with relatively few unidentified parameters, for instance implemented as a collection of partial differential equations (PDEs).

Neural Network

Composing graphical models with neural networks for structured representations and fast inference

Matthew James Johnson
Harvard University
mattjj@seas.harvard.edu

David Duvenaud
Harvard University
dduvenaud@seas.harvard.edu

Alexander B. Wiltschko
Harvard University, Twitter
awiltsch@fas.harvard.edu

Sandeep R. Datta
Harvard Medical School
srdatta@hms.harvard.edu

Ryan P. Adams
Harvard University, Twitter
rpa@seas.harvard.edu

Abstract

We propose a general modeling and inference framework that combines the complementary strengths of probabilistic graphical models and deep learning methods. Our model family composes latent graphical models with neural network observation likelihoods. For inference, we use recognition networks to produce local evidence potentials, then combine them with the model distribution using efficient message-passing algorithms. All components are trained simultaneously with a single stochastic variational inference objective. We illustrate this framework by automatically segmenting and categorizing mouse behavior from raw depth video, and demonstrate several other example models.

Neural Network

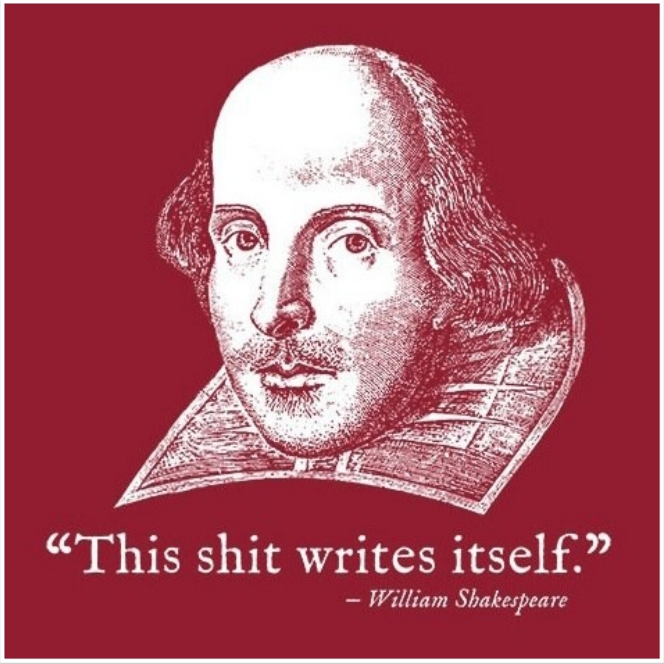
23
Tuesday
Feb 2016

Is deep learning a Markov chain in disguise?

Posted by Brian Lee Yung Rowe in artificial intelligence, data science, R

≈ 7 Comments

Andrej Karpathy's post "The Unreasonable Effectiveness of Recurrent Neural Networks" made splashes last year. The basic premise is that you can create a recurrent neural network to learn language features character-by-character. But is the resultant model any different from a Markov chain built for the same purpose? I implemented a character-by-character Markov chain in R to find out.



— Source: @shakespeare

Neural Network

Ferenc Huszár, Twitter

We only need a small, vague claim that SGD does something Bayesian, and then we're winning.

Keeping Neural Networks Simple by Minimizing the Description Length of the Weights

Geoffrey E. Hinton and Drew van Camp
Department of Computer Science
University of Toronto
10 King's College Road
Toronto M5S 1A4, Canada

FLAT MINIMA

NEURAL COMPUTATION 9(1):1–42 (1997)

Sepp Hochreiter
Fakultät für Informatik
Technische Universität München
80290 München, Germany
hochreit@informatik.tu-muenchen.de
http://www7.informatik.tu-muenchen.de/~hochreit

Jürgen Schmidhuber
IDSIA
Corso Elvezia 36
6900 Lugano, Switzerland
juergen@idsia.ch
http://www.idsia.ch/~juergen

March 1996

Neural Network

Stochastic Gradient Descent as Approximate Bayesian Inference

Stephan Mandt
Data Science Institute, Department of Computer Science
Columbia University

Matthew D. Hoffman
Adobe Research
Adobe Systems Incorporated

David M. Blei
Data Science Institute, Departments of Computer Science and Statistics
Columbia University

April 17, 2017

arun@ska.ac.za