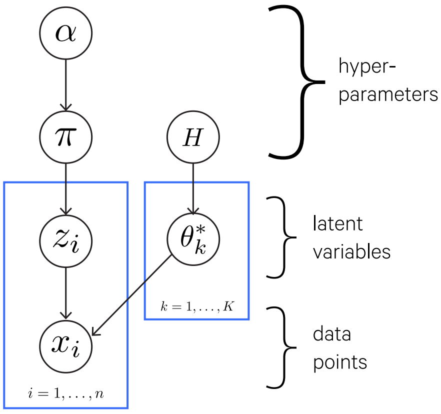
Bayesian Learning and Probabilistic Programming

- https://dsteinberg.github.io/pages/research-projects.html
- https://tminka.github.io/papers/index.html
- http://am207.info/resources.html
- http://mlg.eng.cam.ac.uk/tutorials/07/
- https://www.cs.toronto.edu/~radford/res-bayes-ex.html
- understanding computational Bayesian statistics

Bayesian learning can be regarded as the extension of Bayesian statistics. The core topic of Bayesian learning is thought as prior information to explain the uncertainty of parameters.
It is related with `Bayesian statistics, computational statistics, probabilistic programming and machine learning`.

- [Bayesian] “我是bayesian我怕谁”系列 - Gaussian Process
- Bayesian Learning
- Lecture 9: Bayesian Learning
- Bayesian Learning
- Bayesian machine learning
- Infinite Mixture Models with Nonparametric Bayes and the Dirichlet Process
- Bayesian Learning for Machine Learning: Part I - Introduction to Bayesian Learning
- Bayesian Learning for Machine Learning: Part II - Linear Regression
- Bayesian Learning by Burr H. Settles, CS-540, UW-Madison, www.cs.wisc.edu/~cs540-1
- Bayesian machine learning @fastML
- Understanding emprical Bayesian hierarchical learning
- While My MCMC Gently Samples - Bayesian modeling, Data Science, and Python
- Probabilistic Models in the Study of Language
- Statistical Rethinking A Bayesian Course with Examples in R and Stan (& PyMC3 & brms too)

- COS597C: Advanced Methods in Probabilistic Modeling
- COS513: Foundations of Probabilistic Modeling
- CSSS 564: Bayesian Statistics for the Social Sciences (University of Washington, Spring 2018)
- Radford Neal's Research: Bayesian Inference
- Learn Bayesian statistics
- http://ryanrossi.com/search.php
- https://bayesianwatch.wordpress.com/



Bayes Formulae	**Inverse Bayes Formulae**
$f_X(x) = \frac{f_{X,Y}(X,Y)}{f_{Y\mid X}(y\mid x)} = \frac{f_{X\mid Y}(x\mid y)f_Y(y)}{f_{Y\mid X}(y\mid x)}$	$f_X(x) = (\int_{S_y} \frac{f_{Y\mid X}(y\mid x)}{f_{X\mid Y}(x\mid y)}\mathrm{d}y)^{-1}$
$f_X(x) \propto f_{X\mid Y}(x \mid y) f_Y(y) (= f_{X,Y}(X,Y))$	$f_X(x) \propto \frac{f_{X\mid Y}(x\mid y_0)}{f_{Y\mid X}(y_0\mid x)}$

Naive Bayes

`Naive Bayes` is to reconstruct the joint distribution of features and labels $Pr(\vec{x}, y)$ given some training dataset/samples $T = \{(\vec{X}_i, y_i)\}_{i=1}^{n}$.
However, the features are usually in high dimensional space in practice so the dimension curse occurs which makes it impossible to compute the joint distribution $Pr(\vec{X}, y)$ via the (emprical) conditional probability $Pr(\vec{X} \mid y)$ and the prior $Pr(y)$.
A `naive` idea is to simplify the computation process by assumption that the features are conditional independence so that

$$Pr(\vec{X} \mid y) = \prod_{i=1}^{p} Pr(\vec{X}^{(i)} \mid y). \tag{1}$$

And the predicted labels will be computed via

$$Pr(y \mid \vec{X}) = \frac{Pr(y)Pr(\vec{x} \mid y)}{\sum Pr(y)Pr(\vec{x} \mid y)}. \tag{2}$$

where the conditional probability $Pr(\vec{X} \mid y)$ is simplified by the conditional independence assumption in formula (1).
Thus the naive Bayes classifier is represented as maximum a posteriori (MAP)

$$y = f(x) = \arg_y Pr(y \mid \vec{X}).$$

The prior probability $Pr(y)$ can be emprical or estimated.

- Naive Bayesian
- Ritchie Ng on Machine Learning
- MLE/MAP + Naïve Bayes

Gaussian Naive Bayes

- Gaussian Naive Bayes Classifier
- Gaussian Naïve Bayes
- Naive Bayes and Gaussian Bayes Classifier

Average One-Dependence Estimator (AODE)

- Not so naive Bayes: Aggregating one-dependence estimators
- https://link.springer.com/article/10.1007%2Fs10994-005-4258-6

Variational Bayes Methods

Variational inference is an umbrella term for algorithms which cast posterior inference as optimization.

- Variational-Bayes .org
- A tutorial on variational Bayesian inference
- The Variational Bayesian EM Algorithm for Incomplete Data: with Application to Scoring Graphical Model Structures
- Variational Inference: A Review for Statisticians
- Another Walkthrough of Variational Bayes
- The Variational Approximation for Bayesian Inference
- High-Level Explanation of Variational Inference by Jason Eisner (2011)

Hierarchical Bayesian Regression

`Hierarchical Bayesian Regression` extends the Bayesian models by setting the uncertainty of the uncertainty such as

$$
\begin{gathered}
y \sim Pr(\phi(x) \mid \theta) \\
Pr(\phi(x) \mid \theta) = \frac{Pr(\theta \mid \phi(x)) Pr(\phi(x))}{P(\theta)} \\
Pr(\phi(x)) = Pr(\phi(x) \mid \eta) Pr(\eta) \\
\vdots
\end{gathered}
$$

<img src = https://images2015.cnblogs.com/blog/941880/201607/941880-20160713184714686-1547363080.png width = 50%>
<img src = http://www.indiana.edu/~kruschke/BMLR/HierarchicalDiagram.jpg width = 50%>

We can take any factor into consideration in this hierarchical Bayesian model. And it is a graphical probability model, which consists of the connections and probability.

- [ ] https://twiecki.io/
- [ ] https://twiecki.io/blog/2014/03/17/bayesian-glms-3/
- [ ] https://twiecki.io/blog/2018/08/13/hierarchical_bayesian_neural_network/
- [ ] BAYESIAN HIERARCHICAL MODELS

- [ ] https://www.cnblogs.com/huangxiao2015/p/5667941.html
- [ ] https://www.cnblogs.com/huangxiao2015/p/5668140.html
- [ ] Chapter 4: Regression and Hierarchical Models
- [ ] https://docs.pymc.io/notebooks/GLM-hierarchical.html
- [ ] http://varianceexplained.org/r/hierarchical_bayes_baseball/
- [ ] http://idiom.ucsd.edu/~rlevy/pmsl_textbook/chapters/pmsl_8.pdf
- [ ] https://www.wikiwand.com/en/Bayesian_hierarchical_modeling

---

- [ ] SAS/STAT Examples Bayesian Hierarchical Poisson Regression Model for Overdispersed Count Data
- [ ] Hierarchical Regression and Spatial models
- [ ] Short Course for ENAR 2009 - Sunday, March 15, 2009: Hierarchical Modeling and Analysis of Spatial-Temporal Data: Emphasis in Forestry, Ecology, and Environmental Sciences
- [ ] https://sites.google.com/site/doingbayesiandataanalysis/
- [ ] Bayesian methods for combining multiple Individual and Aggregate data Sources in observational studies.
- [ ] https://www.stat.berkeley.edu/~census/goldbug.pdf
- [ ] http://www.biostat.umn.edu/~ph7440/pubh7440/Lecture5.pdf
- [ ] CS&SS/STAT 564: Bayesian Statistics for the Social Sciences, University of Washington, Spring 2018
- [ ] https://jrnold.github.io/bayesian_notes/

http://doingbayesiandataanalysis.blogspot.com/

Optimal Learning

The Bayesian perspective casts a different interpretation
on the statistics we compute, which is particularly useful in the context of optimal learning.
In the frequentist perspective, we do not start with any knowledge about the system before we have collected any data.
By contrast, in the Bayesian perspective we assume that we begin with a prior distribution of belief about the unknown parameters.

Everyday decisions are made without the benefit of accurate information. Optimal Learning develops the needed principles for gathering information to make decisions, especially when collecting information is time-consuming and expensive.
Optimal learning addresses the problem of efficiently collecting information with which to make decisions.
Optimal learning is an issue primarily in applications where observations
or measurements are expensive.

It is possible to approach the learning problem using classical and familiar ideas from optimization. The operations research community is very familiar with the use of gradients to minimize or maximize functions. Dual variables in linear programs are a form of gradient, and these are what guide the simplex algorithm. Gradients capture the value of an incremental change in some input such as a price, fleet size or the size of buffers in a manufacturing system.
We can apply this same idea to learning.

There is a list of optimal learning problems.

- http://yelp.github.io/MOE/
- Peter I. Frazier@cornell

- Optimal Learning book
- Optimal Learning Course
- https://onlinelibrary.wiley.com/doi/book/10.1002/9781118309858
- 有没有依靠『小数据』学习的机器学习分支? - 覃含章的回答 - 知乎

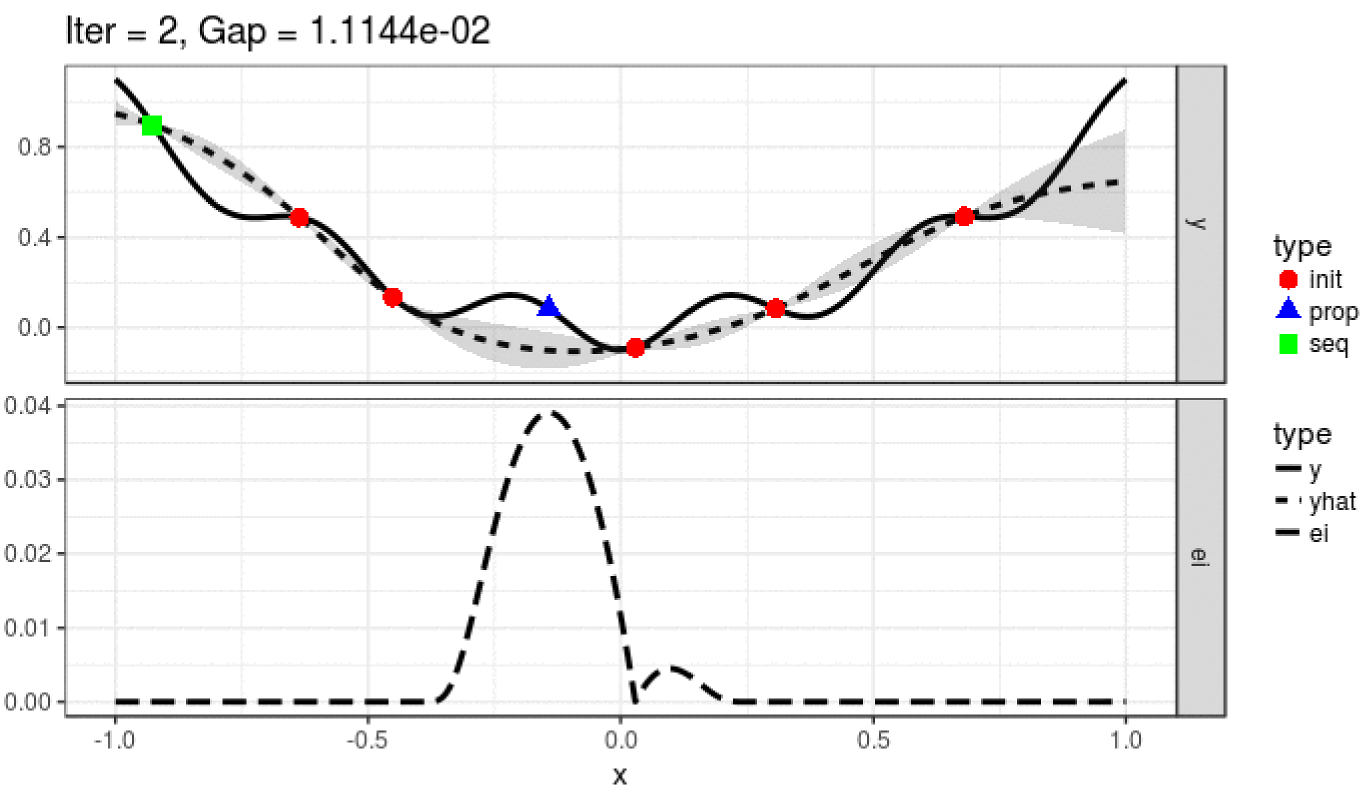
Bayesian Optimization

Bayesian optimization has been successful at global optimization of expensive-to-evaluate multimodal objective functions. However, unlike most optimization methods, Bayesian optimization typically does not use derivative information.

As `response surface methods`, they date back to Box and Wilson in 1951.
Bayesian optimization usually uses `Gaussian process` regression.

Bayesian Optimization

Bayesian Optimization



BayOpt

- http://www.sigopt.com/
- https://jmhessel.github.io/Bayesian-Optimization/
- https://mlrmbo.mlr-org.com/index.html
- Let’s Talk Bayesian Optimization
- Bayesian optimization tutorial slides and article (from INFORMS 2018)
- Practical Bayesian optimization in the presence of outliers

- Bayesian Optimization with Gradients
- https://www.iro.umontreal.ca/~bengioy/cifar/NCAP2014-summerschool/slides/Ryan_adams_140814_bayesopt_ncap.pdf
- https://haikufactory.com/
- Bayesian Optimization at Imperial London College
- PROBABILISTIC-NUMERICS.ORG
- Bayesian optimization@http://krasserm.github.io/
- Introduction to Bayesian Optimization by Javier Gonz´alez
- A Python implementation of global optimization with gaussian processes
- Bayesian Optimization using Pyro
- Taking the Human Out of the Loop:A Review of Bayesian Optimization
- Bayesian Optimization @modAL
- RoBO – a Robust Bayesian Optimization framework written in python.
- Bayesian Optimization@Ployaxon
- BOAT: Building auto-tuners with structured Bayesian optimization
- The Intuitions Behind Bayesian Optimization

Probabilistic Graphical Model

A graphical model or probabilistic graphical model (PGM) or structured probabilistic model is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.
They are commonly used in probability theory, statistics — particularly Bayesian statistics — and machine learning. It is a marriage of graph theory and probability theory.
It is aimed to solve the causal inferences, which is based on principles rather than models.

- Probabilistic inference in graphical models by Jordan
- Foundations of Graphical Models, Fall 2016, Columbia University
- CS 228: Probabilistic Graphical Models, Stanford / Computer Science / Winter 2017-2018
- Probabilistic Graphical Models 10-708, • Spring 2019 • Carnegie Mellon University
- Probabilistic Graphical Models

- https://www.wikiwand.com/en/Graphical_model
- https://blog.applied.ai/probabilistic-graphical-models-for-fraud-detection-part-1/
- https://blog.applied.ai/probabilistic-graphical-models-for-fraud-detection-part-2/
- https://blog.applied.ai/probabilistic-graphical-models-for-fraud-detection-part-3/
- https://frnsys.com/ai_notes/foundations/probabilistic_graphical_models.html

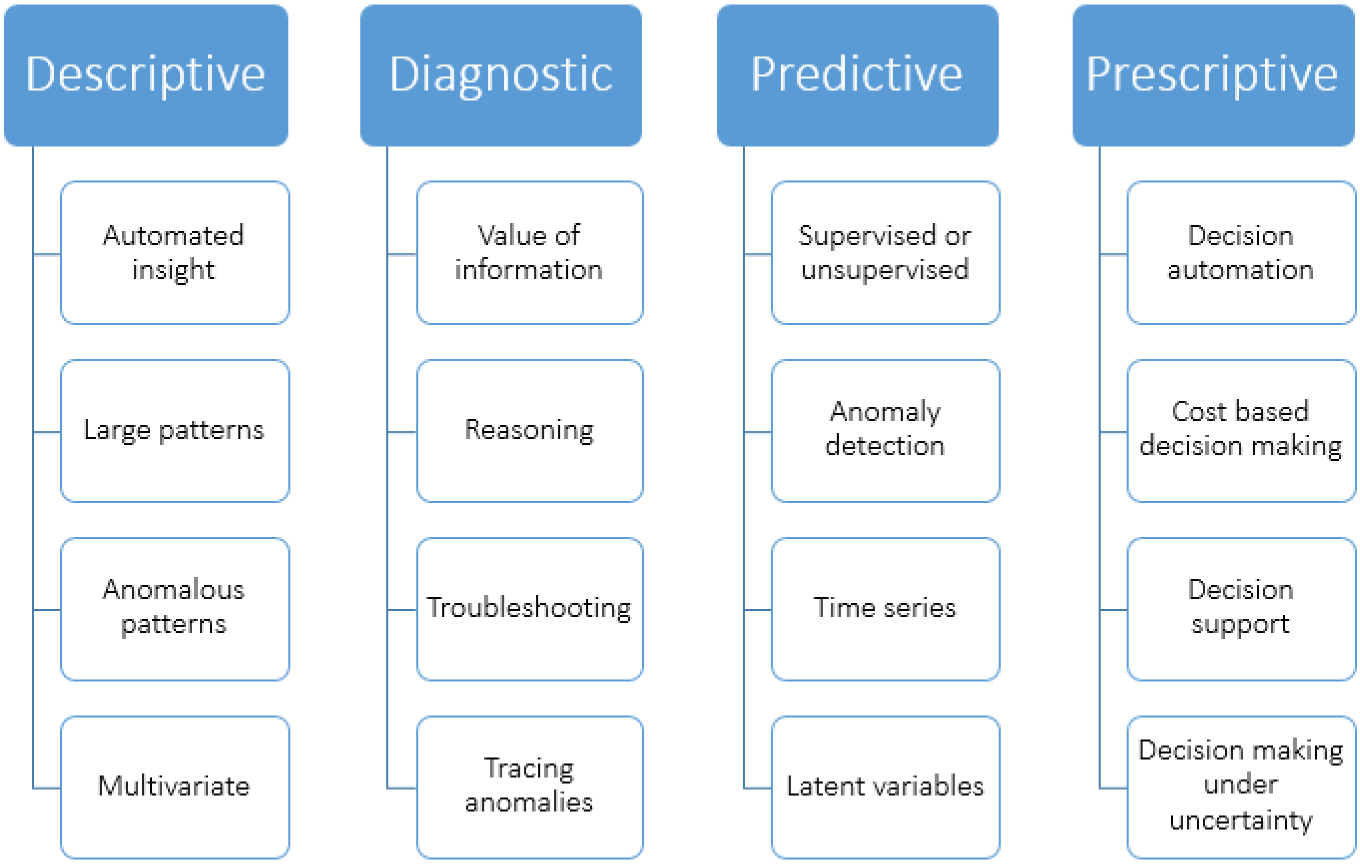
Bayesian Belief Network(BBN)

Bayesian Network

Bayesian networks are a type of `Probabilistic Graphical Model` that can be used to build models from data and/or expert opinion.
They are also commonly referred to as `Bayes nets, Belief networks and sometimes Causal networks`.

Bayesian Network (BN) is an intuitive, graphical representation of a joint probability distribution of a set of random variables with a possible mutual causal relationship.

It is of wide application in many fields such as NLP, medical image analysis.



- Bayesian Network Repository
- Bayesian Networks by João Neto
- Additive Bayesian Network Modelling in R
- https://silo.ai/bayesian-networks-for-fast-troubleshooting/
- Bayesian networks - an introduction
- Bayesian Networks: Introductory Examples
- Bayesian Network – Brief Introduction, Characteristics & Examples
- Bayesian Networks(Part I)
- Bayesian Networks(Part II)
- pomegranate is a Python package that implements fast and flexible probabilistic models.
- http://robsonfernandes.net/bnviewer/
- https://www.hugin.com/

Hidden Markov Models

- Hidden Markov Model (HMM) Markov Processes and HMM
- https://www.maths.lancs.ac.uk/~fearnhea/GTP/

- https://web.stanford.edu/~jurafsky/slp3/A.pdf
- https://pomegranate.readthedocs.io/en/latest/index.html

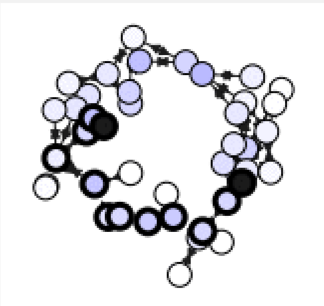
Probabilistic Programming

Probabilistic graphical models provide a formal lingua franca for modeling and a common target for efficient inference algorithms. Their introduction gave rise to an extensive body of work in machine learning, statistics, robotics, vision, biology, neuroscience, artificial intelligence (AI) and cognitive science. However, many of the most innovative and useful probabilistic models published by the AI, machine learning, and statistics community far outstrip the representational capacity of graphical models and associated inference techniques. Models are communicated using a mix of natural language, pseudo code, and mathematical formulae and solved using special purpose, one-off inference methods. Rather than precise specifications suitable for automatic inference, graphical models typically serve as coarse, high-level descriptions, eliding critical aspects such as fine-grained independence, abstraction and recursion.

PROBABILISTIC PROGRAMMING LANGUAGES aim to close this representational gap, unifying general purpose programming with probabilistic modeling; literally, users specify a probabilistic model in its entirety (e.g., by writing code that generates a sample from the joint distribution) and inference follows automatically given the specification. These languages provide the full power of modern programming languages for describing complex distributions, and can enable reuse of libraries of models, support interactive modeling and formal verification, and provide a much-needed abstraction barrier to foster generic, efficient inference in universal model classes.

We believe that the probabilistic programming language approach within AI has the potential to fundamentally change the way we understand, design, build, test and deploy probabilistic systems. This approach has seen growing interest within AI over the last 10 years, yet the endeavor builds on over 40 years of work in range of diverse fields including mathematical logic, theoretical computer science, formal methods, programming languages, as well as machine learning, computational statistics, systems biology, probabilistic AI.

Graph Nets library



- Probabilistic Programming Primer
- PROBABILISTIC-PROGRAMMING.org
- Programming Languages and Logics Fall 2018
- Graph Nets library
- Hakaru a simply-typed probabilistic programming language
- PROBPROG 2018 -- The International Conference on Probabilistic Programming
- PyMC: Probabilistic Programming in Python
- Stan is a state-of-the-art platform for statistical modeling and high-performance statistical computation.
- Welcome to Pyro Examples and Tutorials!

- A quick update: Edward, and some motivations
- Monadic probabilistic programming in Scala with Rainier
- CSCI E-82A Probabilistic Programming and Artificial Intelligence Stephen Elston, PhD, Principle Consultant, Quantia Analytics LLC
- http://probcomp.csail.mit.edu/
- https://www.beast2.org/

- https://www.stat.cmu.edu/~cshalizi/uADA/17/
- https://www.cs.cmu.edu/~aarti/Class/10701/lecs.html