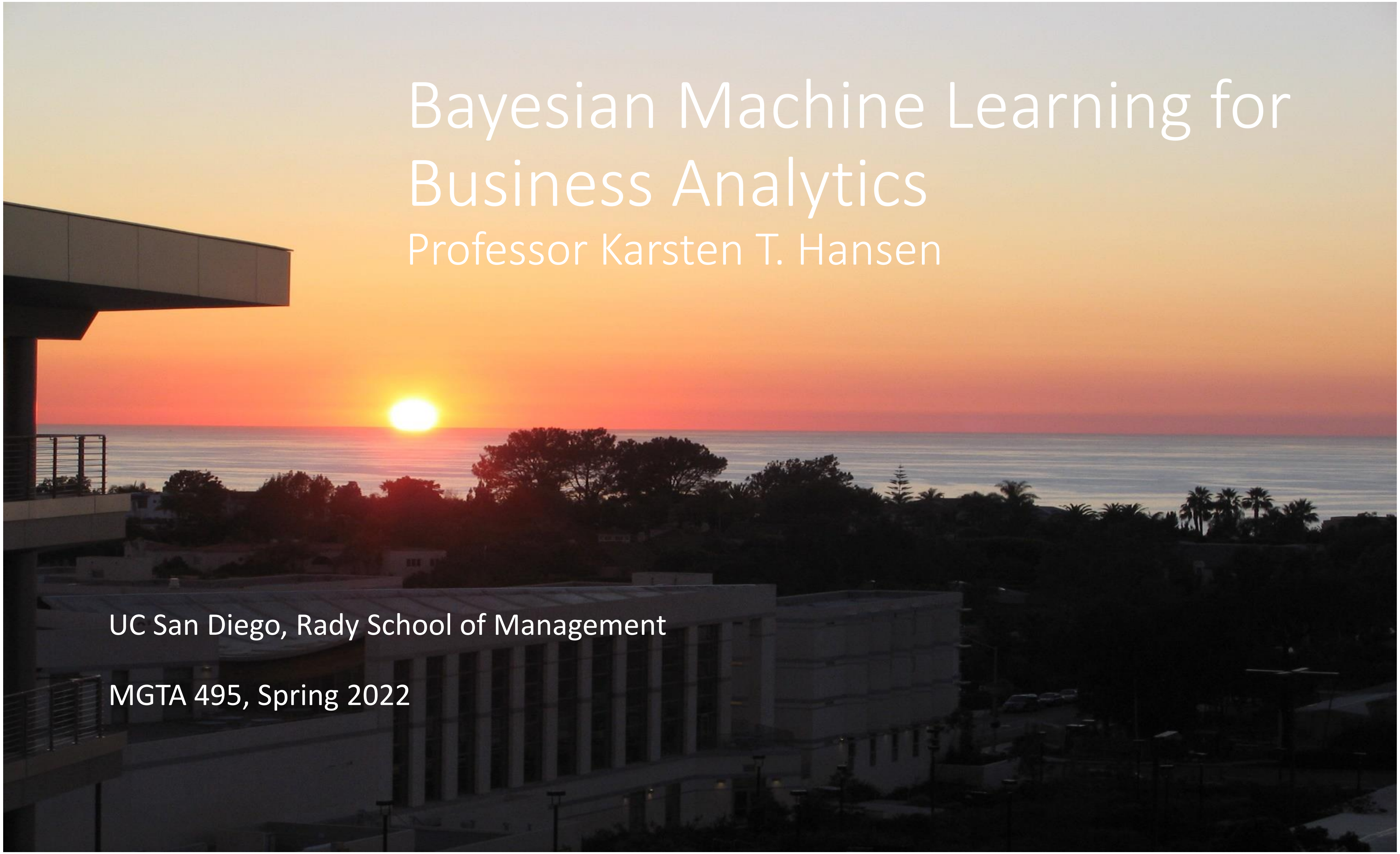


Bayesian Machine Learning for Business Analytics

Professor Karsten T. Hansen

UC San Diego, Rady School of Management

MGTA 495, Spring 2022

What have we done so far…

- Trained some model

$$y = f(x, \theta)$$

  where x are features and θ are weights/biases.
- Performed various unsupervised learning tasks such as clustering, data reduction etc.
- This is all good – and is sometimes all we need – but sometimes we require more

Getting richer insights

- How do we account for uncertainty? We usually just train the model by searching for the "best" architecture and weights $\hat{f}(x, \hat{\theta})$
- Then we plug in some x* to get a single prediction:

$$y^* = \hat{f}(x^*, \hat{\theta})$$

- Here y* is a single number – but how sure is the model/machine of this prediction?
- Answering this question requires that the model also outputs some measure of uncertainty

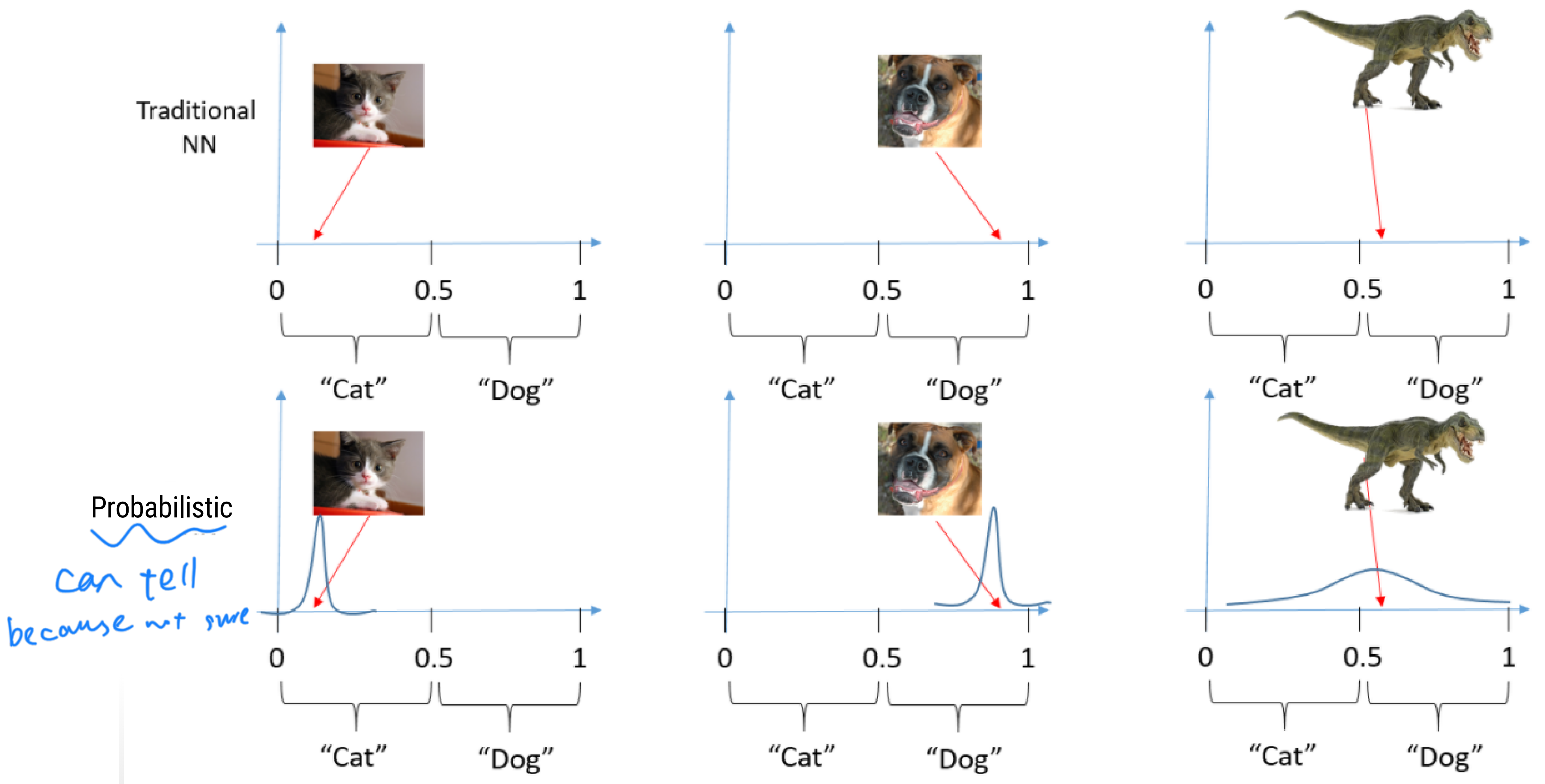
Getting richer insights

- If we are to use our model outputs for decision making, it is crucial that we characterize the associated uncertainty of the outputs
- How can we get a model to tell us "I don't know!" ? How ?
- How can we get a model to tell us "I'm really sure of this!" ?

- How can we add information we may already have to get better outputs?

Example

show sth out of the class



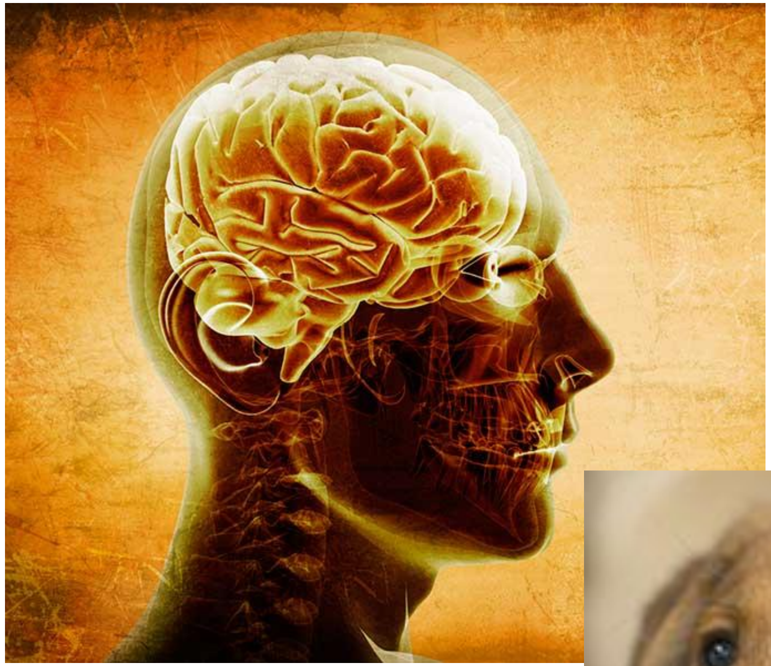
What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?

Alex Kendall
University of Cambridge
agk34@cam.ac.uk

Yarin Gal
University of Cambridge
yg279@cam.ac.uk

Understanding what a model does not know is a critical part of many machine learning systems. Today, deep learning algorithms are able to learn powerful representations which can map high dimensional data to an array of outputs. However these mappings are often taken blindly and assumed to be accurate, which is not always the case. In two recent examples this has had disastrous consequences. In May 2016 there was the first fatality from an assisted driving system, caused by the perception system confusing the white side of a trailer for bright sky [1]. In a second recent example, an image classification system erroneously identified two African Americans as gorillas [2], raising concerns of racial discrimination. If both these algorithms were able to assign a high level of uncertainty to their erroneous predictions, then the system may have been able to make better decisions and likely avoid disaster.

Reasoning



Prior Knowledge + New Data

Updated Knowledge

How do we formalize this process?

Artificial Intelligence

If we want to design complex artificial systems, we need these systems to be able to (1) form beliefs about the real world and (2) express the strength of these beliefs

- "There is a car approaching"
- "This email is spam"
- "That person is lying"

Probabilistic Reasoning

- How can we formalize this process?
- If we are to teach a machine how to do this, we need a specific set of rules that we can give to the machine
- Consider an event E about which there is uncertainty
  - How do we encode this uncertainty mathematically?
  - Suppose the machine currently have information I1 related to E, and then learn some new information I2. How should we instruct the machine to update beliefs about E?

(Very) Simple Example

- Reasoning about a coin
- Initially, suppose you are pretty sure that it is a fair coin but you are not fully confident in this
  - How to express this precisely?
- You see one result of a coin toss. It is "Tails" – should you update? How?
- Suppose the next is also "Tails". What now?

Probabilistic Reasoning = Bayesian Reasoning

Bayesian Reasoning

- Provides a specific set of instructions on how to reason with incomplete information
- Compare to deductive reasoning: If A then B
- Bayesian Reasoning: If A then B becomes more plausible
- Bayesian reasoning = using the laws of probability to drive our thinking
- This requires a complete probabilistic model

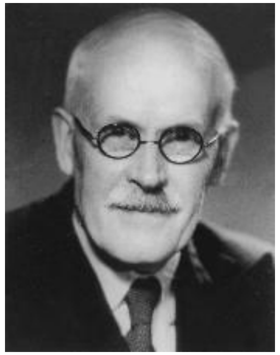
models are pretty large, never very far

when 1980s, computer cheap

Thomas Bayes, 1701-1761

parallel

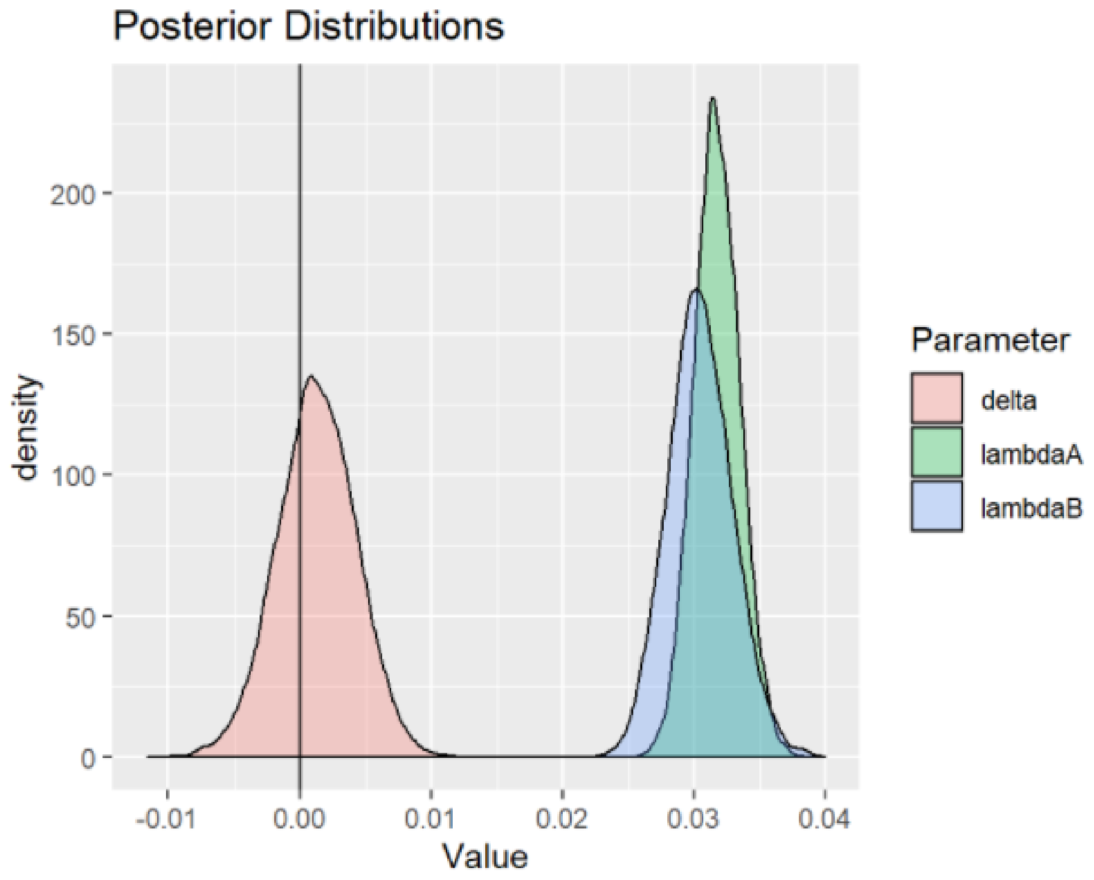
Pierre-Simon Laplace (1749-1827)



Harold Jeffreys (1891-1989)

Week 1

- Introduction to Bayesian Thinking
  - How to encode knowledge as probabilities?
  - How to update knowledge when new information arrives?
- Simple Examples
  - How to update knowledge analytically?



Week 2

- Hierarchical (or Multilevel) Models
- Shrinkage
- How to compare many individual estimates?
- Example:
  - Player 1 has played 6 matches and won 3
  - Player 2 has played 300 matches and won 150
  - Who is the better player?
- Bayesian Decision Theory

Microsoft | Research Our research Programs & events More Sign up: Research News

TrueSkill™ Ranking System

Established: November 18, 2005

Overview Publications Groups

The TrueSkill ranking system is a skill based ranking system for Xbox Live developed at Microsoft Research. The purpose of a ranking system is to both identify and track the skills of gamers in a game (mode) in order to be able to match them into competitive matches. TrueSkill has been used to rank and match players in many different games, from Halo 3 to Forza Motorsport 7.

An improved version of the TrueSkill ranking system, named TrueSkill 2, launched with Gears of War 4 and was later incorporated into Halo 5.

Week 3

Project Meetings!

NO EXAMS

Done week 10

May 31st is final day

~~20~~ 20min per team

Little bit more technical

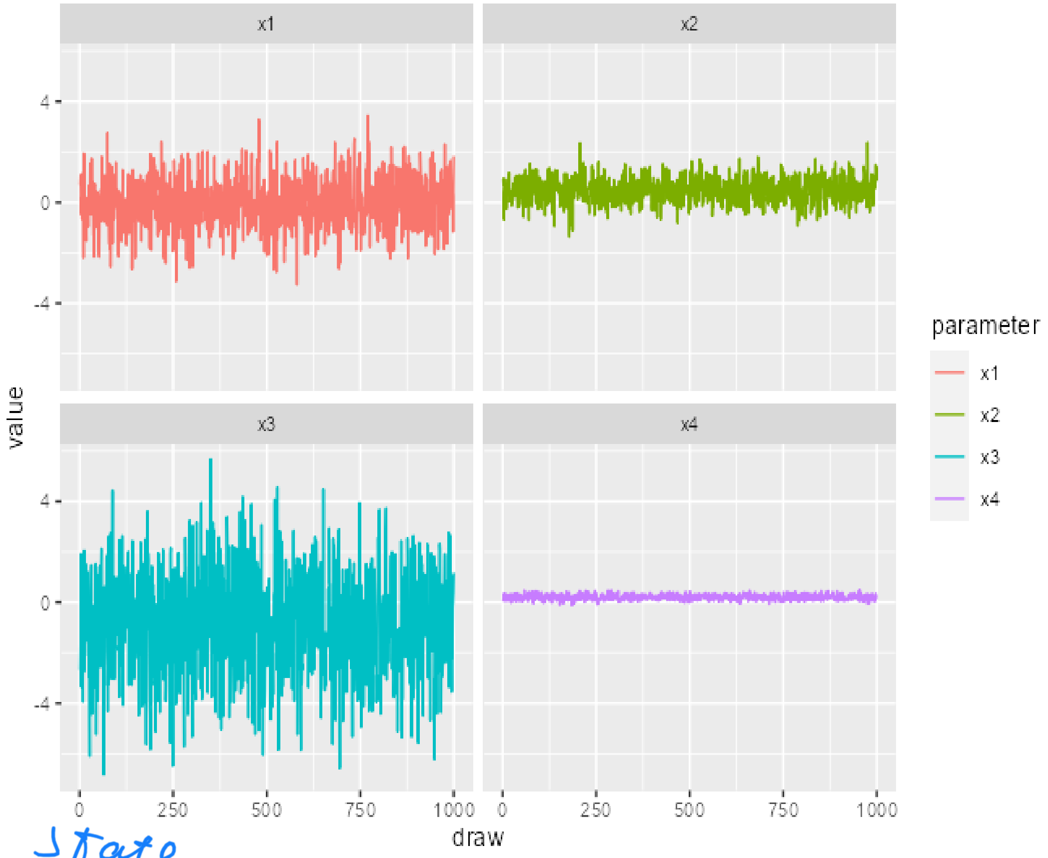
use prob models for ~~sth~~

formly start in week 3

main workload

Week 4

- Algorithms for probabilistic/Bayesian models
- Deterministic vs Probabilistic Algorithms
- Probabilistic programming
- Probabilistic algorithms are based on generating a stream of pseudo random numbers Discrete time
    - Markov Chain Monte Carlo continuous state
    - Hybrid Monte Carlo

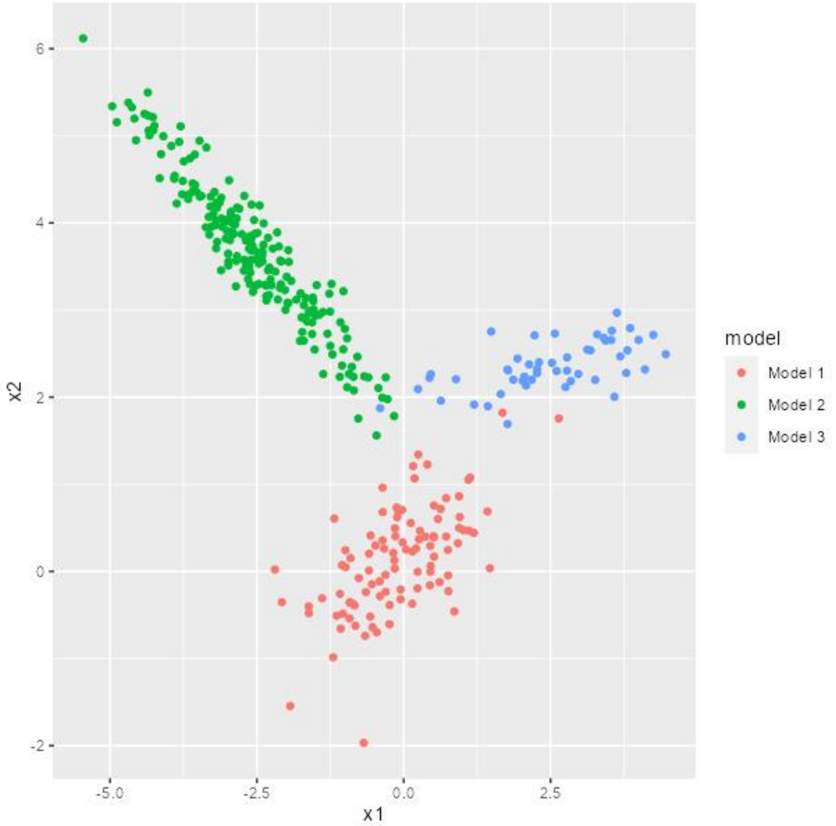


Week 5

- Bayesian Classification
  - Logit Models
  - Multinomial Logit Models
  - Hierarchical Classification
- Naïve Bayes

Week 6

- Mixture Models
- Bayesian Clustering



$$p(y|\theta) = \sum_{j=1}^{K} \lambda_j \, f(y|\gamma_j)$$

Week 7

Project Meetings 2

Week 8,9

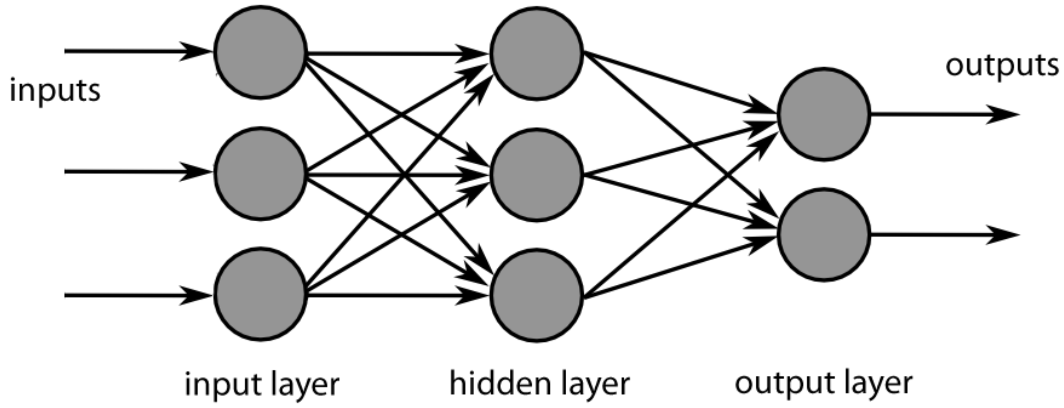
- Probabilistic Matrix Factorization
- Applications
  - Topic Models
  - Recommender Systems
  - Clustering of non-negative data

Bayesian Nonparametric Poisson Factorization for Recommendation Systems

Prem Gopalan Princeton University
Francisco J. R. Ruiz University Carlos III in Madrid
Rajesh Ranganath Princeton University
David M. Blei Princeton University

Other Topics….

- Bayesian Deep Learning
- Neural Nets with uncertainty
- Hard but promising area…



- Variational Autoencoders
- Generative Models
- Representing high dimensional data in smaller dimensions

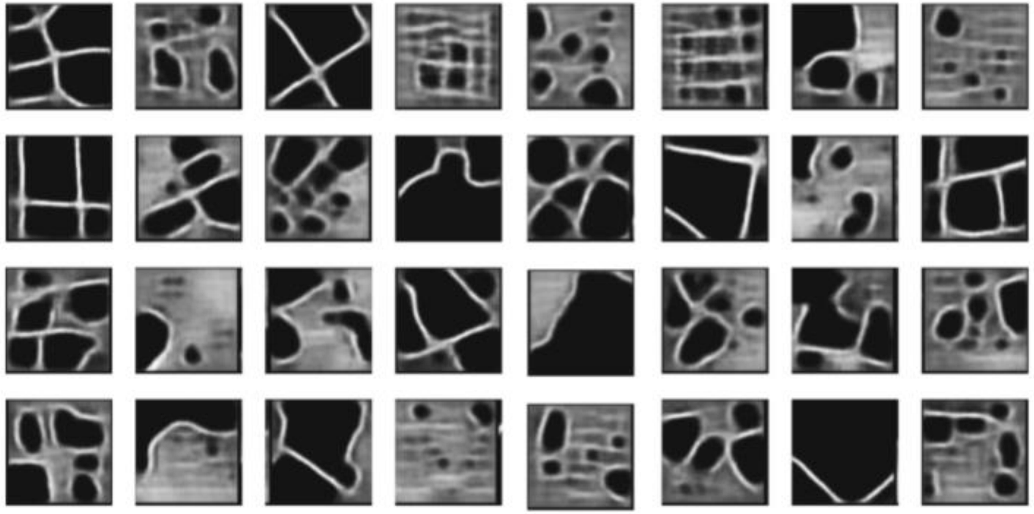


Fig. 8 Examples of synthetic urban street forms generated by passing a randomly sampled latent code *z* through the decoder network

Kempinska and Murcio *Applied Network Science* (2019) 4:114
https://doi.org/10.1007/s41109-019-0234-0
Applied Network Science

RESEARCH Open Access

Modelling urban networks using Variational Autoencoders

Kira Kempinska[1,2*] and Roberto Murcio[1]

Track: User Modeling, Interaction and Experience on the Web WWW 2018, April 23-27, 2018, Lyon, France

Variational Autoencoders for Collaborative Filtering

Dawen Liang
Netflix
Los Gatos, CA
dliang@netflix.com

Rahul G. Krishnan
MIT
Cambridge, MA
rahulgk@mit.edu

Matthew D. Hoffman
Google AI
San Francisco, CA
mhoffman@google.com

Tony Jebara
Netflix
Los Gatos, CA
tjebara@netflix.com

Learning Latent Representations of Bank Customers With The Variational Autoencoder

Rogelio A. Mancisidor[a,b,*], Michael Kampffmeyer [a], Kjersti Aas [c], Robert Jenssen [a]

Logistics

- Regular class hours: Tuesdays, 9:00am-11:50am PST
- I am available on Discord for text/chat sessions
- No exams!
- 3 to 4 individual assignments
- Team project – to be presented on May 31 (week 10)
- Class materials: Slides, code, text book, articles

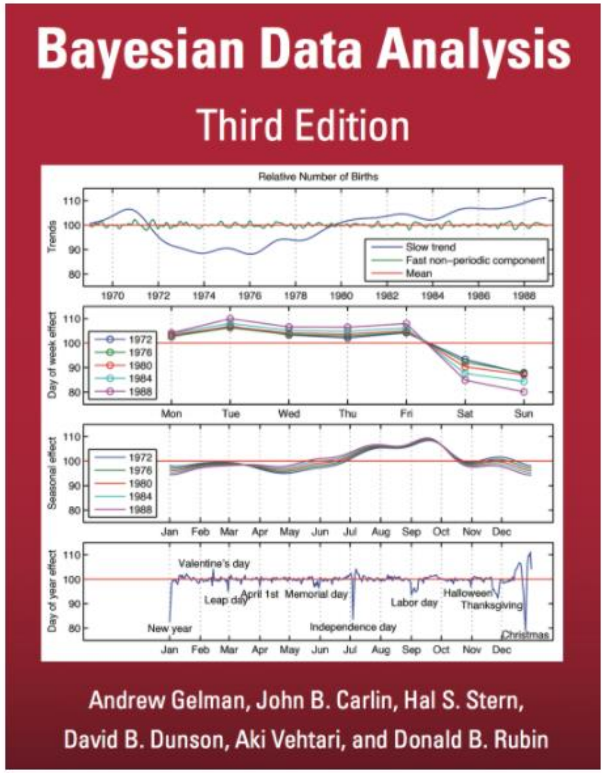
Team Project

- Try to implement a Bayesian model using real data
- You should either code up your own algorithm OR at least make sure you understand the code if you use existing code
- I will serve as your consultant on the project
- Week 3: Project meeting 1
- Teams: self selected
- Output: Presentation in week 10

Project Ideas

- Hierarchical/Multilevel Models for tabular data
  - Regression
  - Classification
  - Ranking project (sports, gaming etc)
- Document classification using topic models
- Project on Decision Theory
- Probabilitistic Non-Negative Matrix Factorization
  - Recommendation systems
  - Count data
- Variational Autoencoders for generative tasks
- Bayesian Neural Nets

Materials



+ articles + slides + code