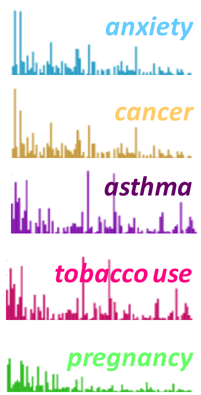
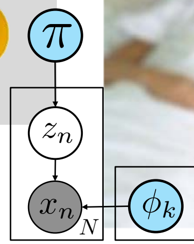
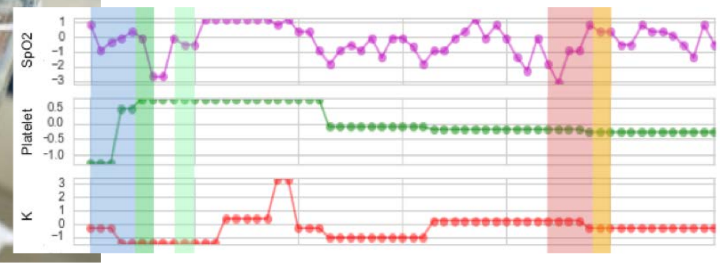
Probabilistic Models and Optimization Algorithms for Personalized Medicine







Mike Hughes

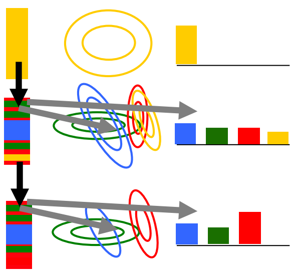
Assistant Professor of Computer Science

joint work with
Finale Doshi-Velez (Harvard)
Erik Sudderth & Gabe Hope (UC Irvine)
Tom McCoy, M.D. and Roy Perlis, M.D. (MGH)
Marzyeh Ghassemi (Toronto), Mike Wu (Stanford)

slides / papers / code
www.michaelchughes.com

My Research Mission:

Reliable Training of Interpretable Models for Complex Data

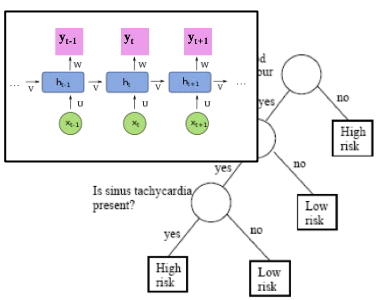


Inference Algorithms for Probabilistic Models:

Adapt Model Size to Data (Bayesian Nonparametrics)

- Dirichlet process mixtures — ***Hughes*** *& Sudderth NeurIPS 2013*
- Topic models of news articles — ***Hughes****, Kim & Sudderth AISTATS 2015*
- HMMs for epigenomics — ***Hughes****, Stephenson, & Sudderth NeurIPS 2015*

BNPy open-source software | `pip install bnpy | github.com/bnpy/bnpy`



Explainable AI:

Optimize Deep NNs for Interpretability

- Find alternate explanations — *Ross,* ***Hughes****, Doshi-Velez IJCAI 2017*
- Find tree-like neural nets — *Wu,* ***Hughes****, Parbhoo, et al. AAAI 2018*

Applications: Personalized Treatments in Medicine

- Predict interventions in ICU — *Ghassemi, Wu,* ***Hughes****, et al. AMIA CRI 2017*
- Discovering subtypes and treatments for depression — ***Hughes*** *et al. (AISTATS 2018)*; ***Hughes*** *et al. (in prep. for JAMA Psychiatry)*

Problem: When will ICU patient **need intervention**?

*Ghassemi, Wu, **Hughes**, et al. AMIA CRI 2017*

Interventions:

- Ventilators to assist breathing
- blood pressure drugs

Early prediction helps:

prepare patient
plan staffing
try less aggressive options early

Cohort from MIMIC-III dataset

mimic.physionet.org

(Johnson et al. Sci. Data 2016)

36,050 patients

- from Beth-Israel Deaconess in Boston 2001-2012
- kept all adults with record within 6-360 hours

Intervention	Training Num Positive	Training Num Control	Heldout Num Positive	Heldout Num Control
Vasopressor	6987	21865	1737	5461
Red blood cell transfusion	19171	9681	4776	2422
Fresh frozen plasma transfusion	2759	26093	620	6578
Platelet transfusion	27818	1034	6944	254
Mechanical Ventilation	13710	15142	3393	3805

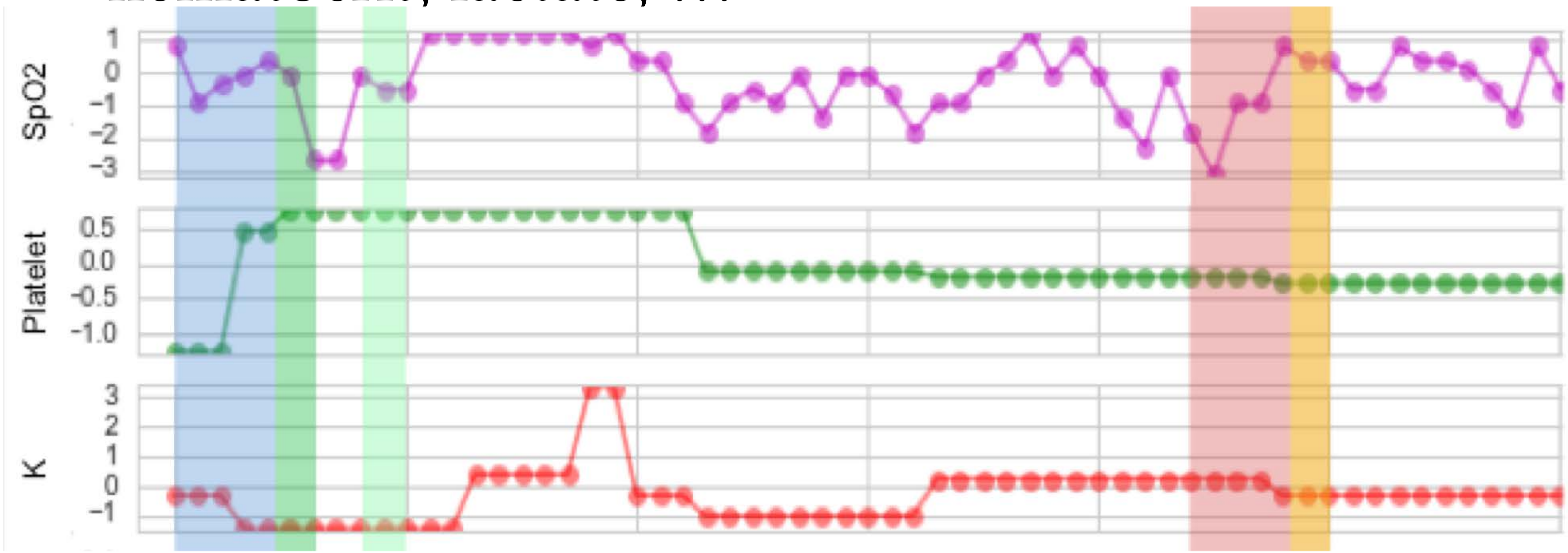
Observed data

7 nurse-validated vital signs (hourly)

heart rate, blood pressure, temp., SpO2, …

11 lab measurements (much less than hourly)

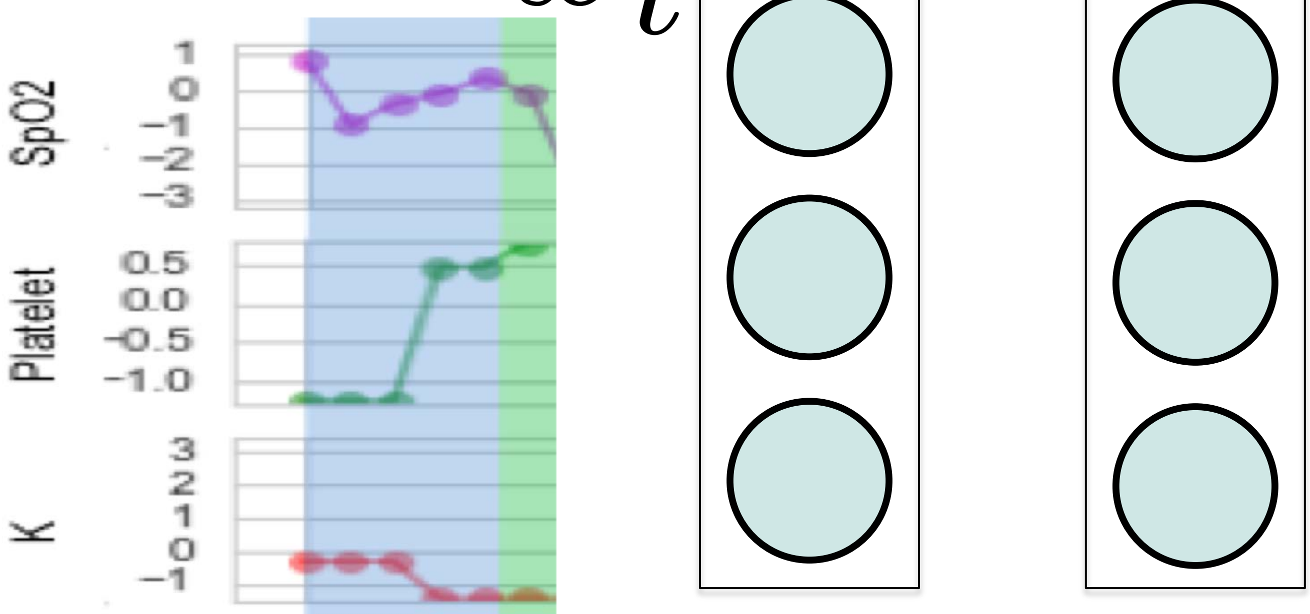
hematocrit, lactate, …



Challenge: how to build models that effectively handle irregular data arrival times

Probabilistic time-series model

Observed Feature Vector x_t



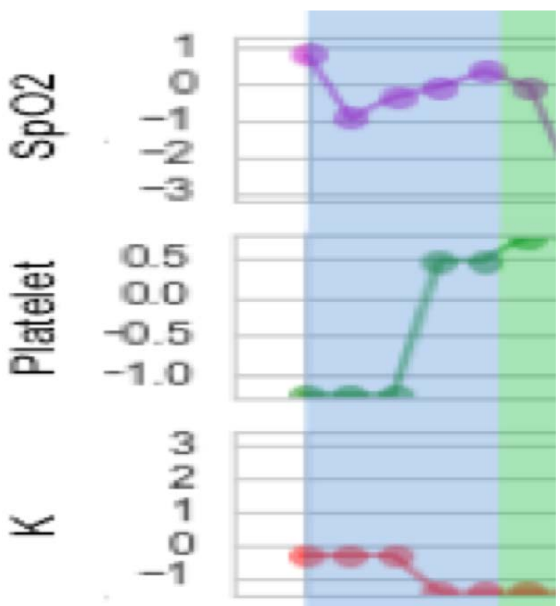
.... hour t hour t+1 ...

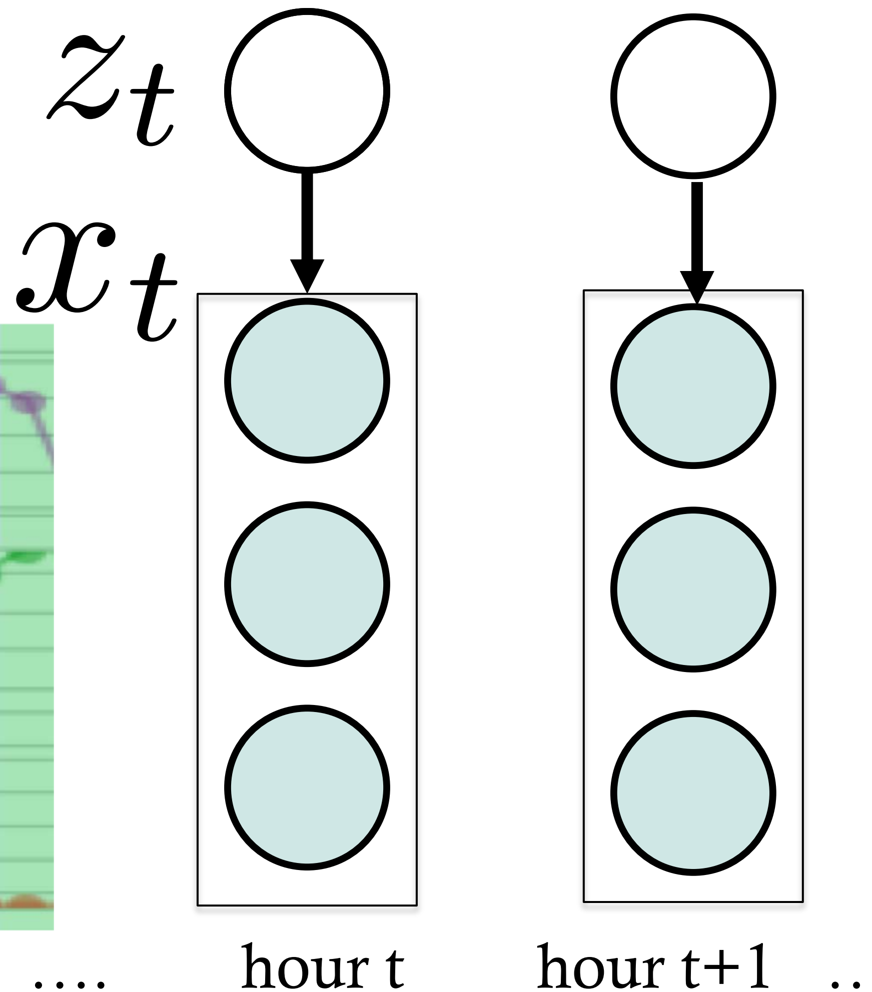
Probabilistic time-series model

Hidden Patient State

one of K possible values

Observed Feature Vector

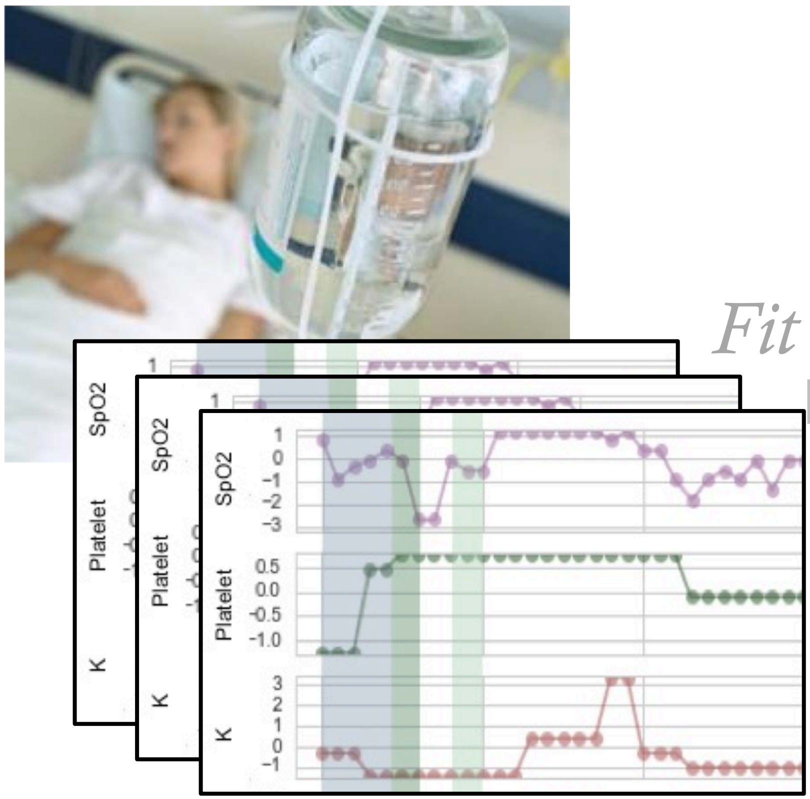




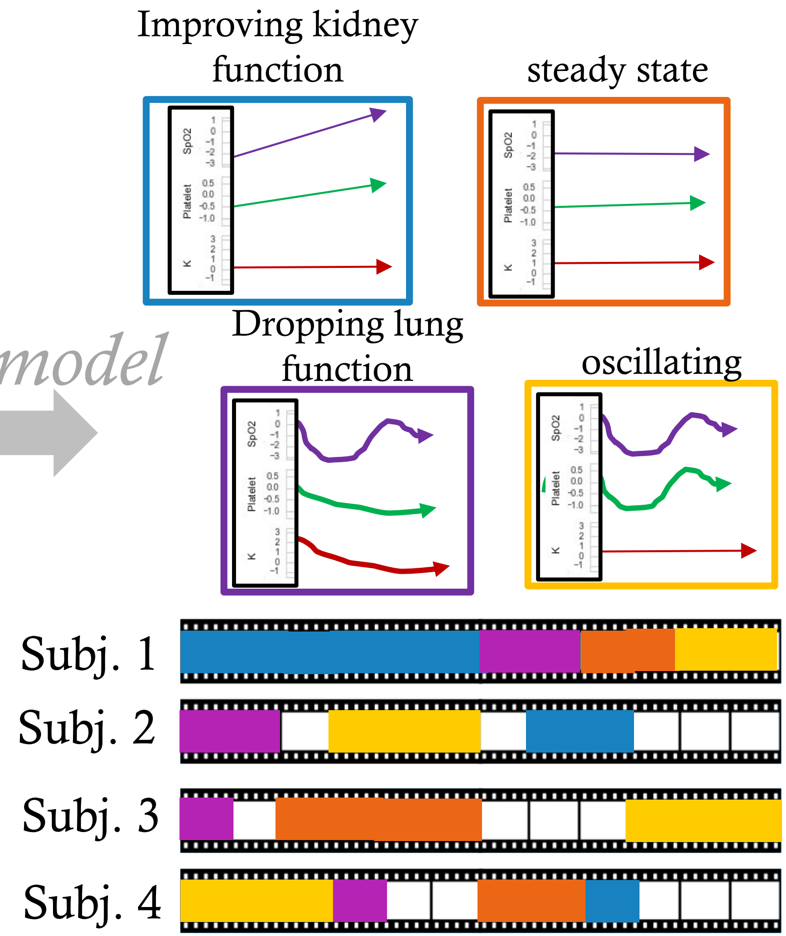
Goal: Summaries of Health

ICU signals from many patients

Health state trajectories
z_1 …z_T

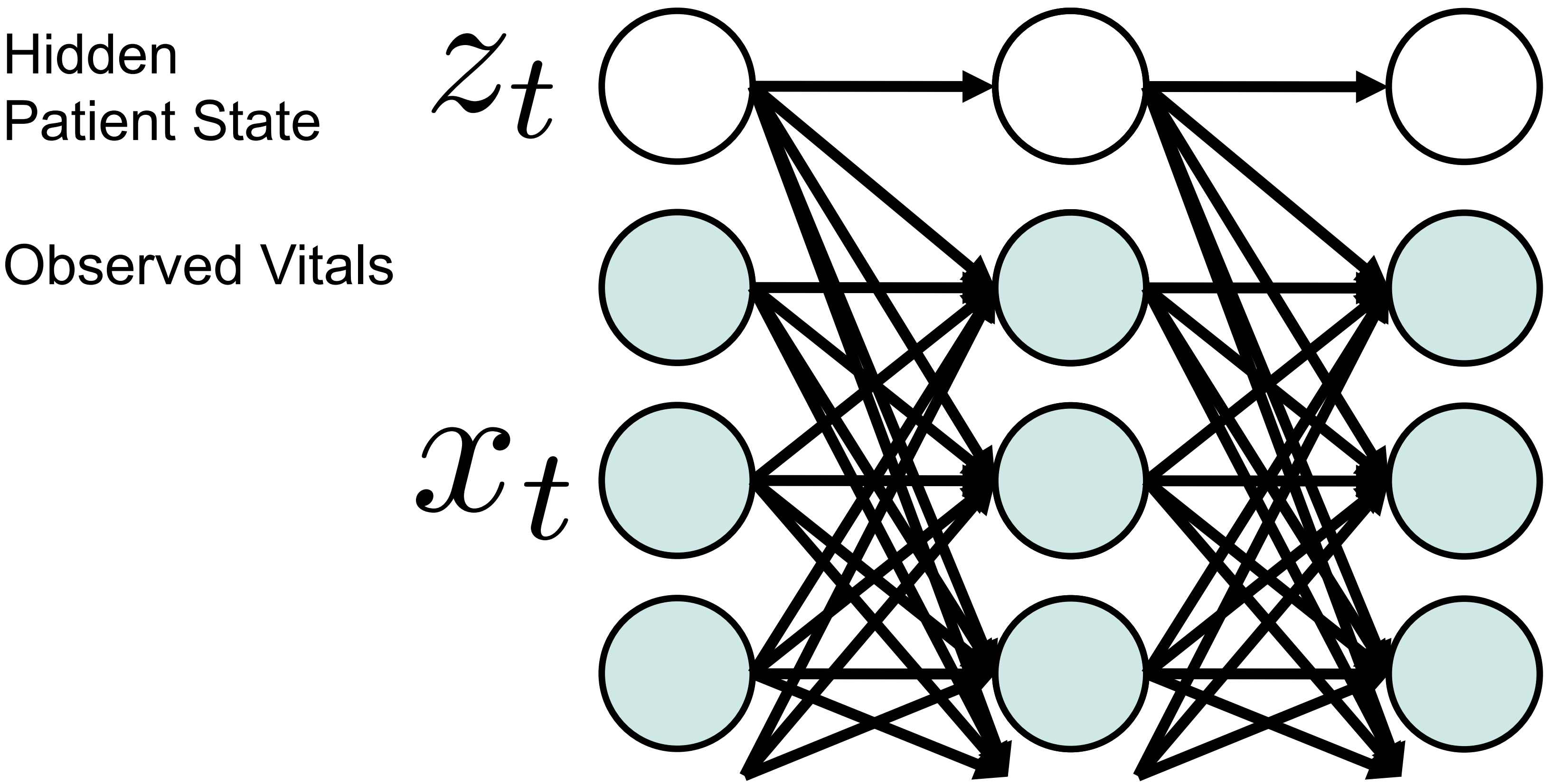


Fit the model



(Ghassemi, Wu, Hughes, et al AMIA CRI '17)

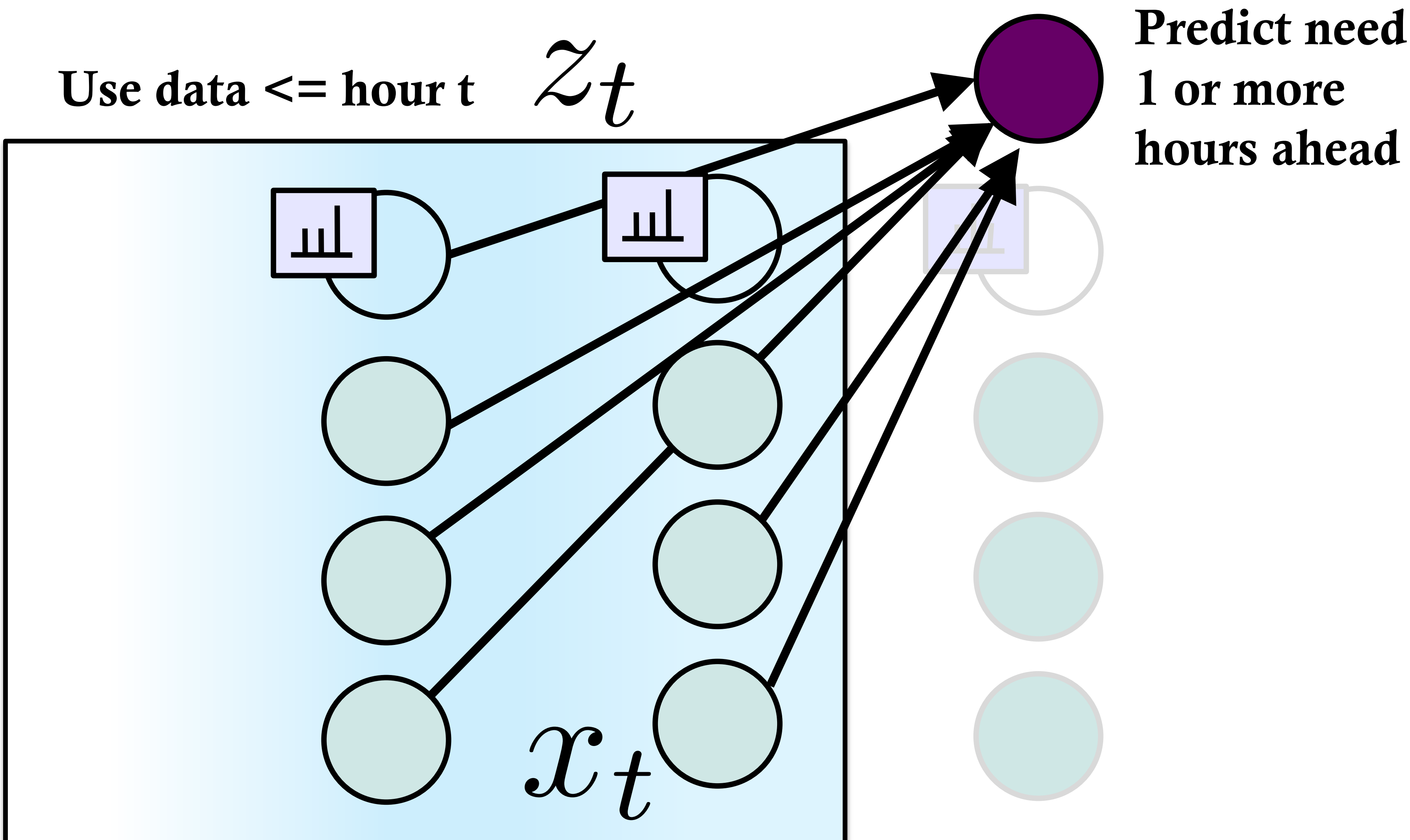
Autoregressive time-series model



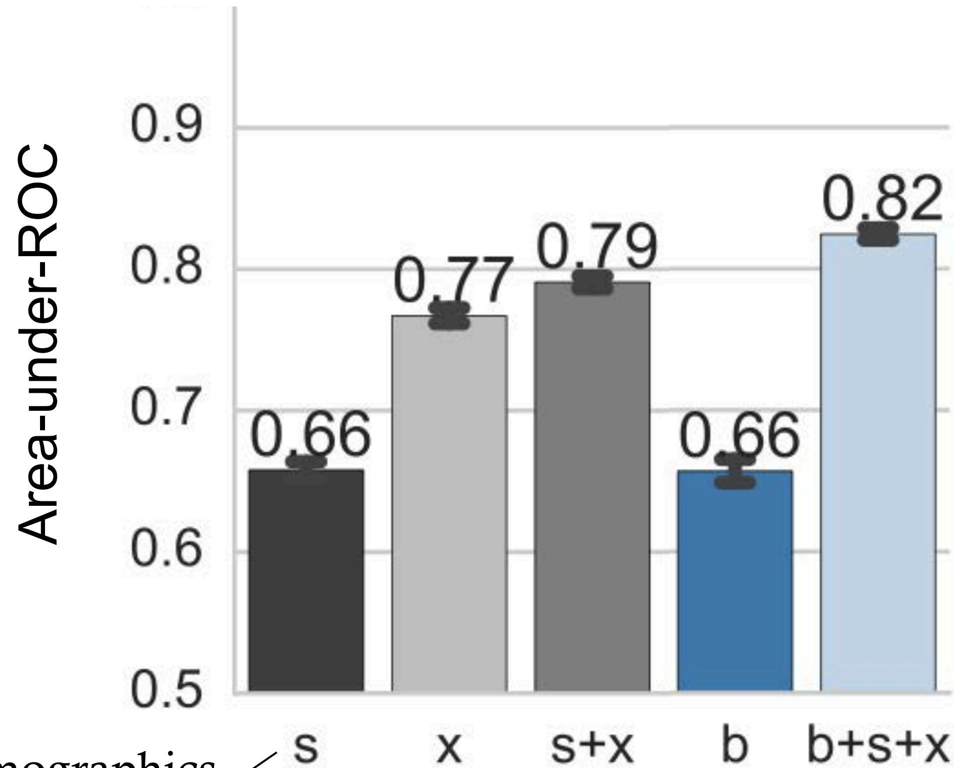
$$x_t | z_t = k \sim \mathcal{N}(A_k x_{t-1} + \mu_k, \Sigma_k)$$

autoregressive Gaussian allows modeling trajectories/trends in vitals

Task: predict need **in advance**



Vasopressor prediction : 1 hr ahead



static demographics
(age, race, etc)

patient vitals at time t

Learned
"patient state embedding"
using time-series model

Medical Data is **Exciting**

Good models can improve

- **personalized treatments** for patients
- **scientific knowledge** about disease
  - subtypes
  - co-occurring conditions

Medical Data is **Challenging**

Challenges

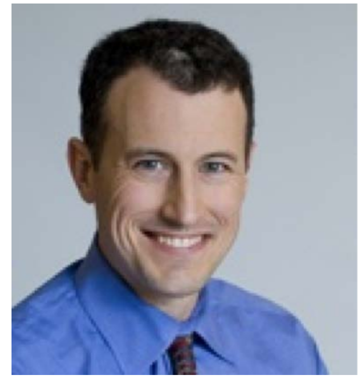
- how to interpret and trust model?
- labeled data hard to get, lots of unlabeled data
- causality

How Can ML Help Psychiatrists?

Prof. Finale Doshi-Velez

Prof. Erik Sudderth

+



Roy Perlis, MD

Tom McCoy, MD

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

THE UNIVERSITY OF CALIFORNIA
IRVINE

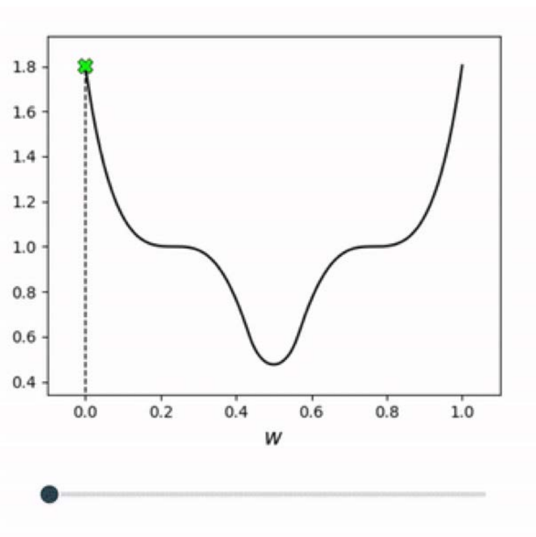
MGH 1811
MASSACHUSETTS
GENERAL HOSPITAL
PSYCHIATRY

Leah Weiner

Gabe Hope

How to optimize?

$$\max_{\phi,\eta} \quad \lambda \log p(y|x,\phi,\eta) + \log p(x|\phi)$$



Credit: Jeremy Watt

Optimize via stochastic gradient descent

- Write objective as Python code
- Automatic gradients from Tensorflow

Compare to Human Doctors

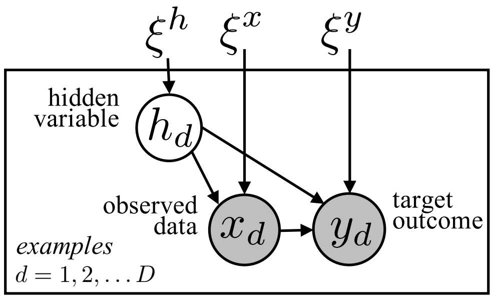
- Evaluation with retrospective data is challenging
- Reasonable attempt:

TOP-3 Accuracy

Given 3 guesses to recommend antidepressants for each patient, what fraction of patients have guess match at least 1 drug on stable list.

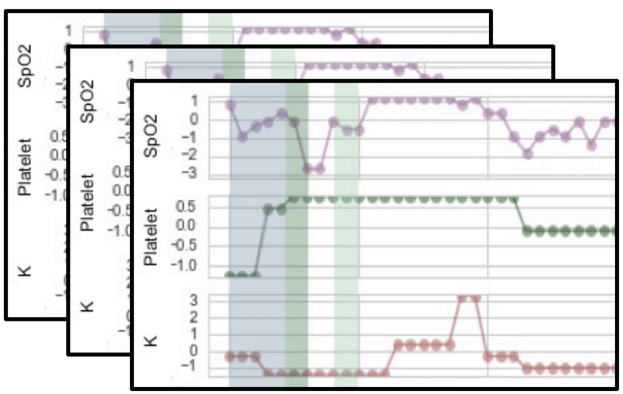
	1st visit	**…**	**3rd visit**
human doctors (observed practice)	**87% (+/- 2)**		40% (+/- 7)
always give most common drugs	54% (+/- 3)		35% (+/- 7)
our method	58% (+/-2)		**46% (+/- 7)**

Future: Time-Series Models to combine data from data



- Mixture models
- Topic models
- Hidden Markov models
- Network models (MMSB)
- PCA or factor analysis
- Non-negative matrix factorization
- Probabilistic encoder/decoder (VAE)

- Disease progression over time



- Models of many data sources