Are you exploiting your assumptions?

Towards expressive priors for biomarker discovery and functional prediction

Melanie F. Pradier

Harvard University

HDSI | Harvard Data Science Initiative

CRCS Center for Research on Computation and Society at Harvard John A. Paulson School of Engineering and Applied Sciences

Are you exploiting your assumptions?

Towards expressive priors for biomarker discovery and functional prediction

Melanie F. Pradier

Harvard University

HDSI | Harvard Data Science Initiative

CRCS Center for Research on Computation and Society at Harvard John A. Paulson School of Engineering and Applied Sciences

Are you exploiting your assumptions?

Towards expressive priors for biomarker discovery and functional prediction

Research in collaboration with...




HARVARD

MGH 1811

Roche

Universidad Carlos III de Madrid

University of Cambridge






ML systems are transforming our society

Google

machine learning

All | News | Images | Videos

About 3,000,000,000 results (0.57 seconds)

MAGAZINE • ARTIFICIAL INTELLIGENCE

A.I. breakthroughs in natural-language processing are big for business

BY JEREMY KAHN

January 20, 2020 4:30 AM EST

The American Journal of Medicine

Volume 132, Issue 7, July 2019, Pages 795-801

ELSEVIER

Review

Artificial Intelligence Transforms the Future of Health Care

Nariman Noorbakhsh-Sabet MD, Ramin Zand MD, MPH, Yanfei Zhang PhD, Vida Abedi PhD

Unveiling Biology with Deep Microscopy

An AI-inspired Revolution in the Life Sciences

Brian Hilbush | Follow

Dec 5, 2019 · 6 min read

Article | Open Access | Published: 05 February 2020

Inferring structural variant cancer cell fraction

Nature Communications **11**, Article number: 730 (2020) | Cite this article

Powerful antibiotic discovered using machine learning for first time

Thu 20 Feb 2020 11.28 EST

Team at MIT says halicin kills some of the world's most dangerous strains

Research

26 February 2020 | Open Access

Impact of a deep learning assistant on the histopathologic classification of liver cancer

Amirhossein Kiani, Bora Uyumazturk [...] Jeanne Shen

npj Digital Medicine **3**, 23

nature

Subscribe

NEWS FEATURE · 09 OCTOBER 2019

Why deep-learning AIs are so easy to fool

Artificial-intelligence researchers are trying to fix the flaws of neural networks.

Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation

K Johnson, F Pasquale, J Chapman - Fordham L. Rev., 2019 - HeinOnline

Over the last decade, a growing number of digital startups launched bids to lure business

RESEARCH-ARTICLE

Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?

Publication: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems • May 2019

HDSR

Published Jan 31, 2020

Should We Trust Algorithms?

TECHNOLOGY AND IIOT > DIGITAL TOOLS

Making AI Work with Small Data

It's a true manufacturing dilemma: 'teaching' AI when defects are a rarity.

Alejandro Betancourt

FEB 12, 2020

nature machine intelligence

Perspective | Published: 13 May 2019

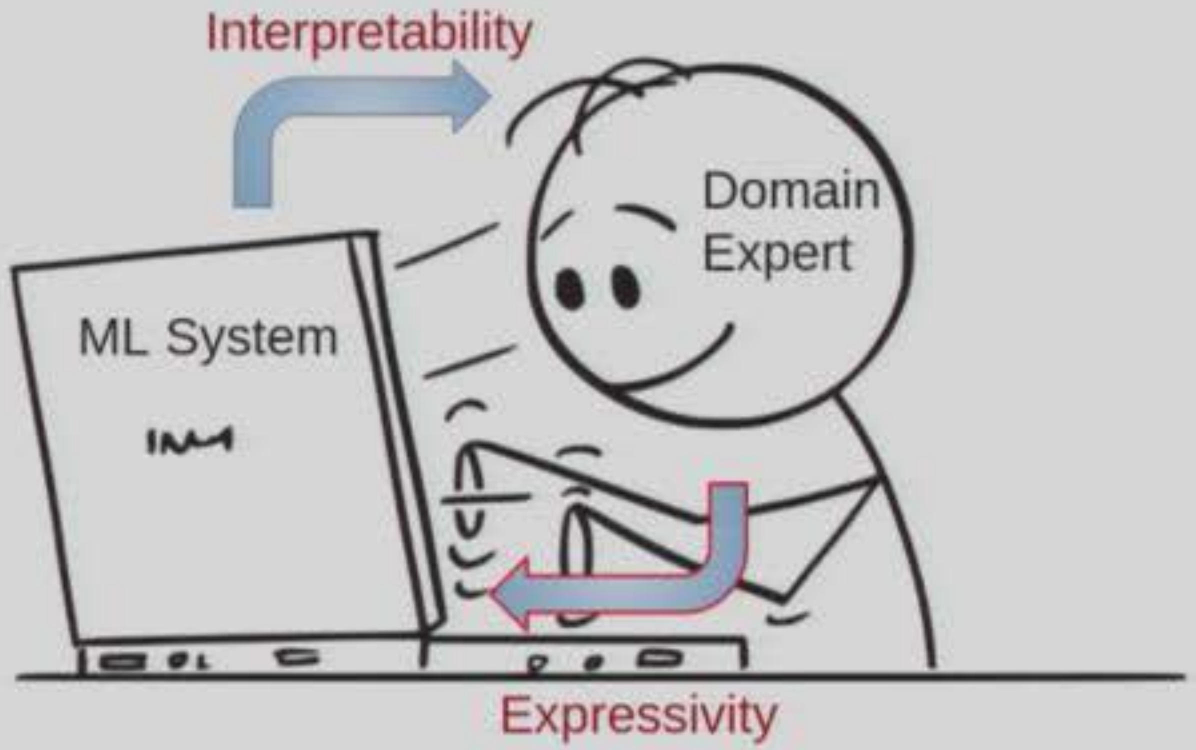
Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

The real test of an AI machine is when it can admit to not knowing something

John Naughton

Interpretability-Expressivity Loop



Interpretability

- "ability to present in understandable terms to a human" (Doshi-Velez et.al, 2017; 2018 EU GDPR)

My contributions

Machine Learning
[UAI, 2020]* [JMLR, 2020]
[DBL@NeurIPS, 2019]
[BNP@NeurIPS, 2018]
[AABI, 2019]

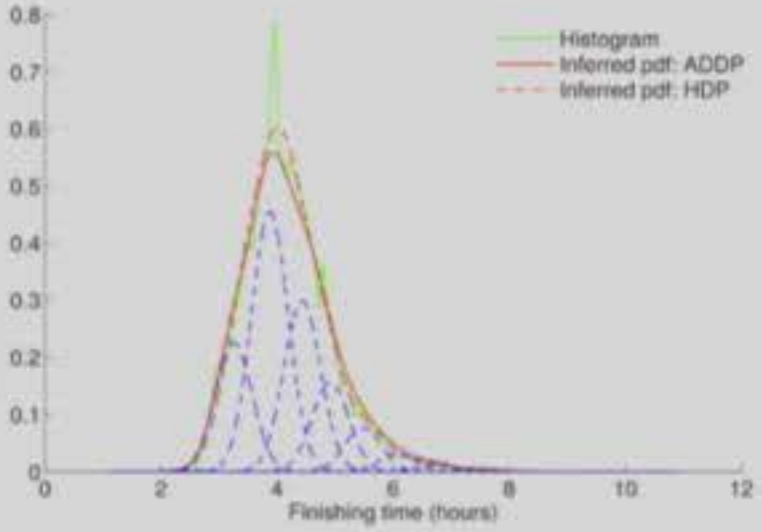
Information Theory
[Entropy, 2016]

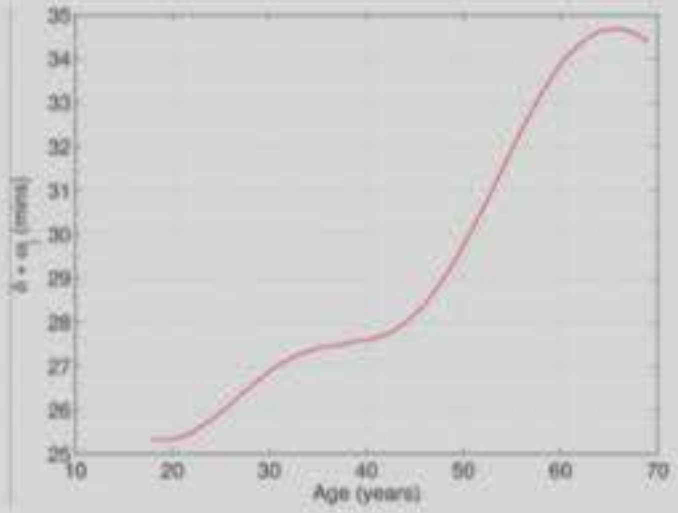
Sport Sciences
[PlosOne, 2016]

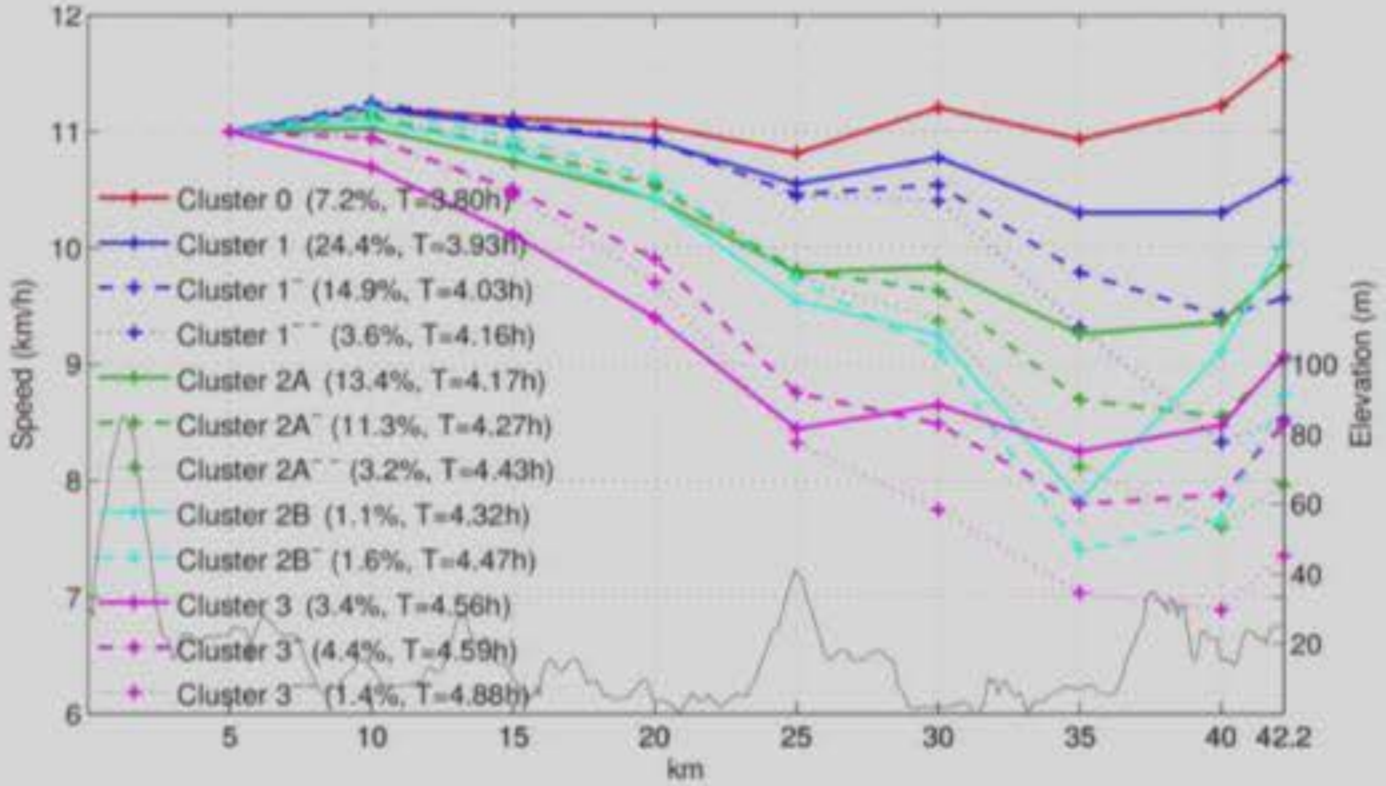
Economics
[PlosOne, 2018]
[ICASSP, 2018]

Cancer Research
[BMC Cancer, 2019]
[Nature Com, 2019]*

Healthcare
[JAMA Psych, 2020]
[Nature Trans. Psych, 2020]
[AJP, 2019]*



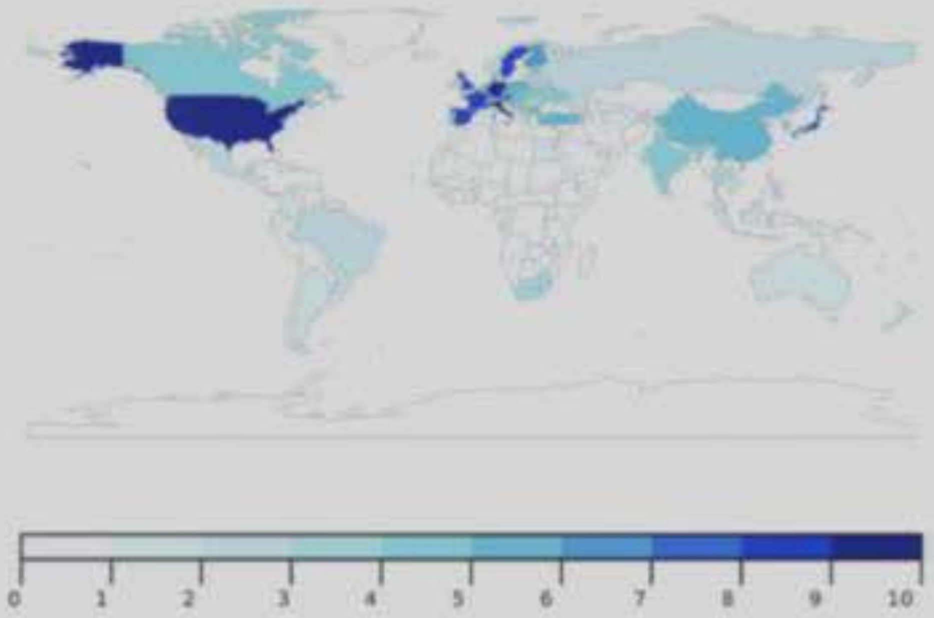


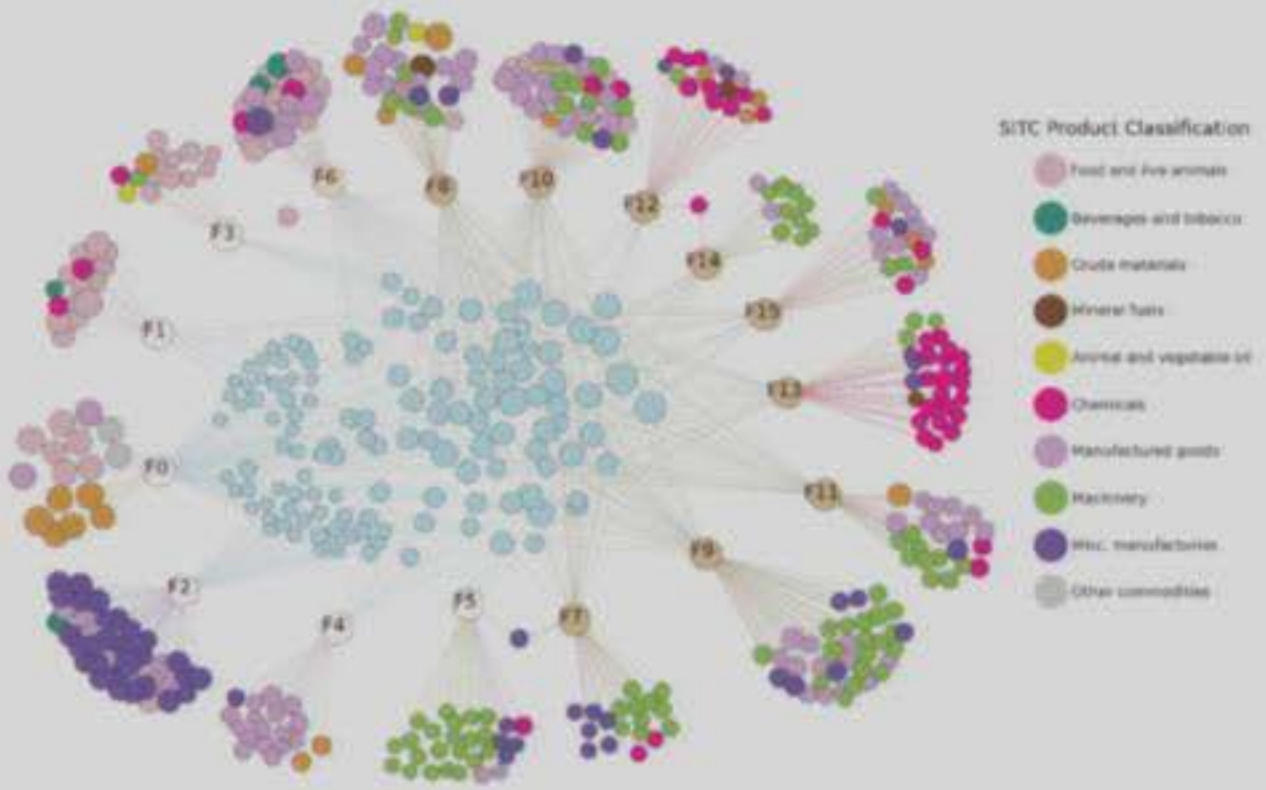


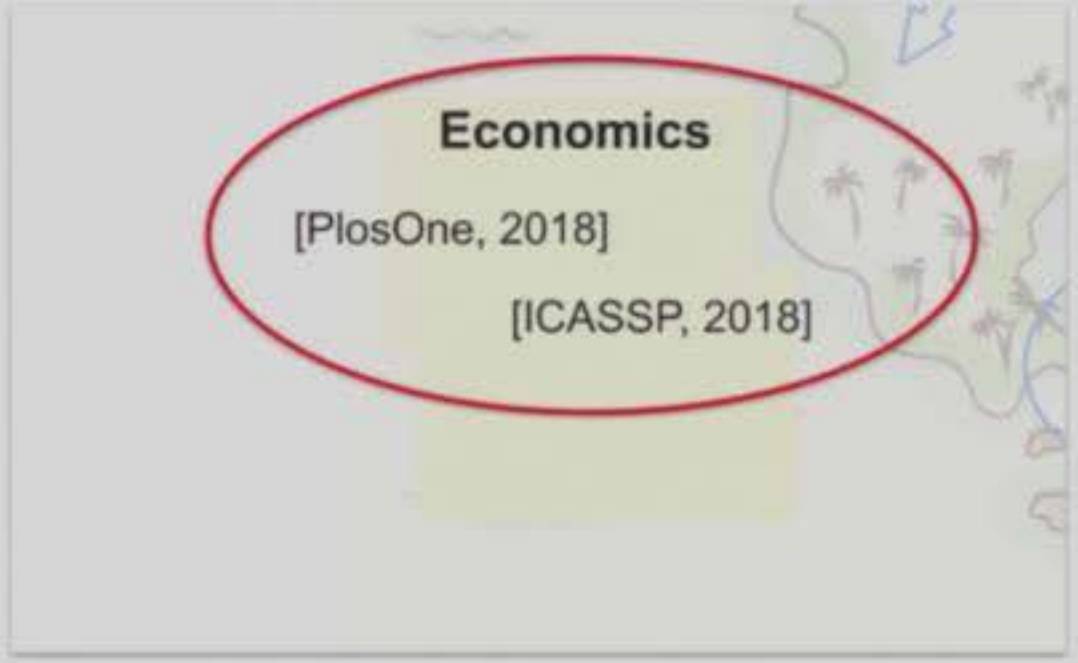
How does aging impact our athletic performance?

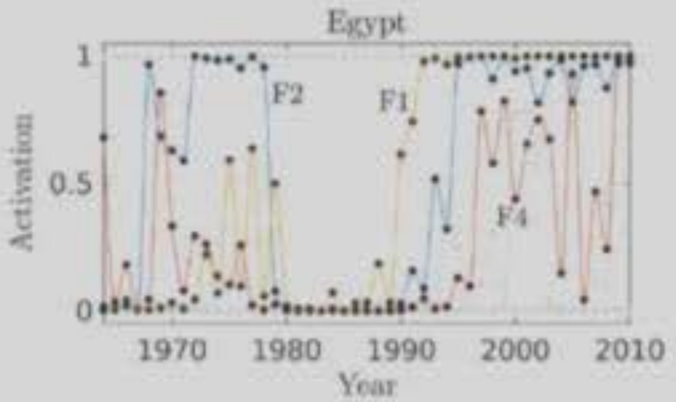
M. F. Pradier, F. J. R. Ruiz, and F. Perez-Cruz. **Prior Design for Dependent Dirichlet Processes: An Application to Marathon Modeling**. *PlosONE*. 2016.

Which factors make countries wealthier than others?



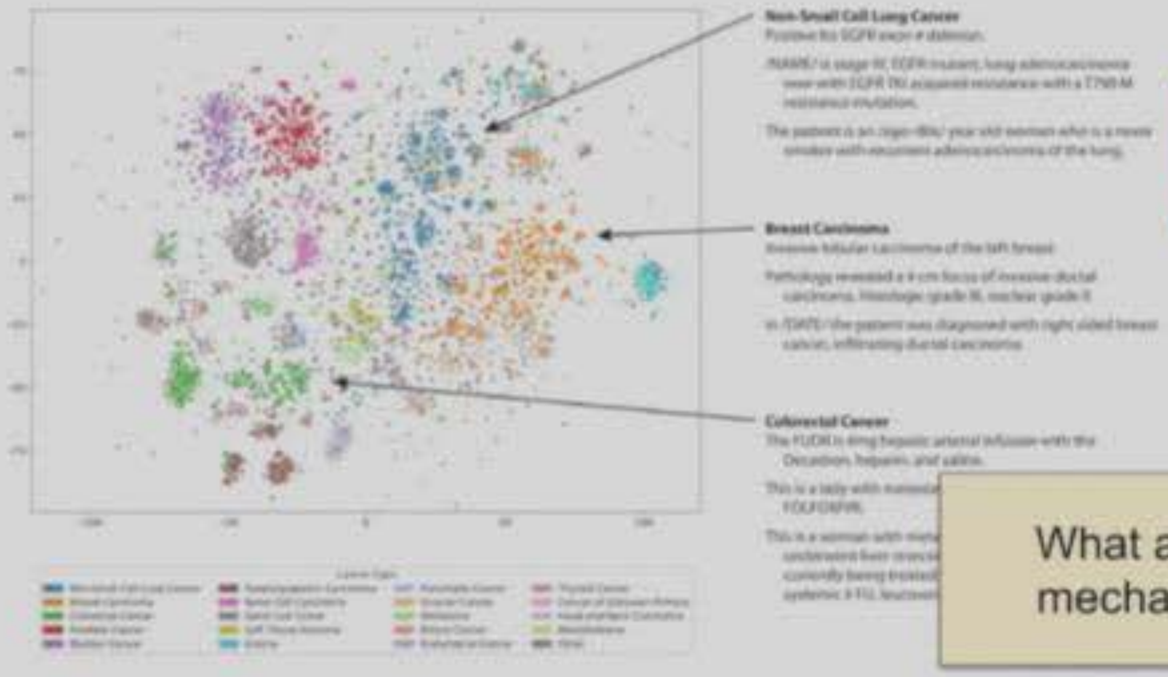






Which factors make countries wealthier than others?

M. F. Pradier*, Z. Utkovski*, V. Stojkoski, L. Kocarev and F. Perez-Cruz. **Economic Complexity Unfolded: An Interpretable Model for the Productive Structure of Economies**. *PlosONE*. 2018.



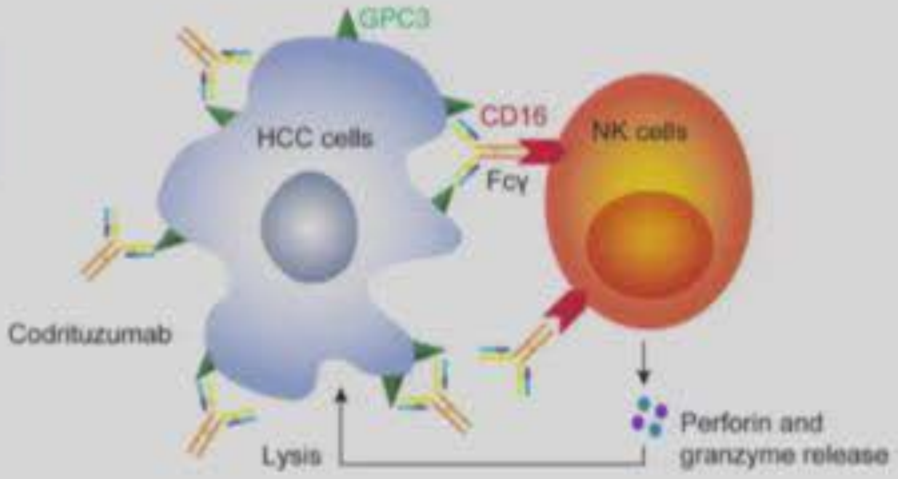
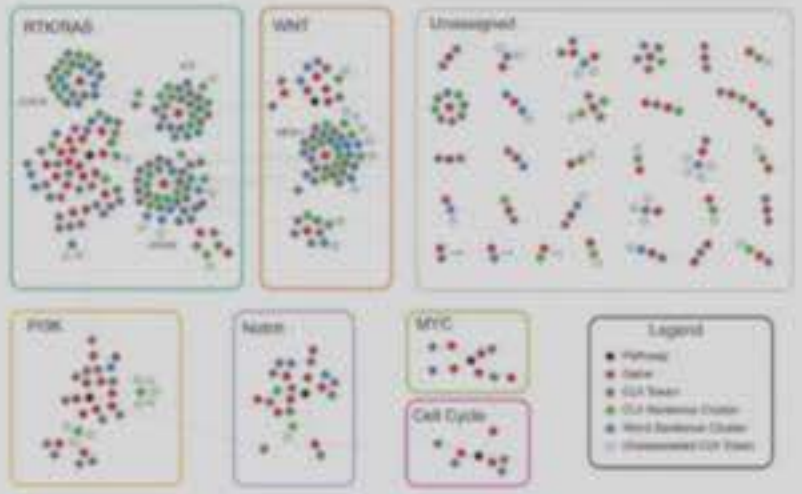
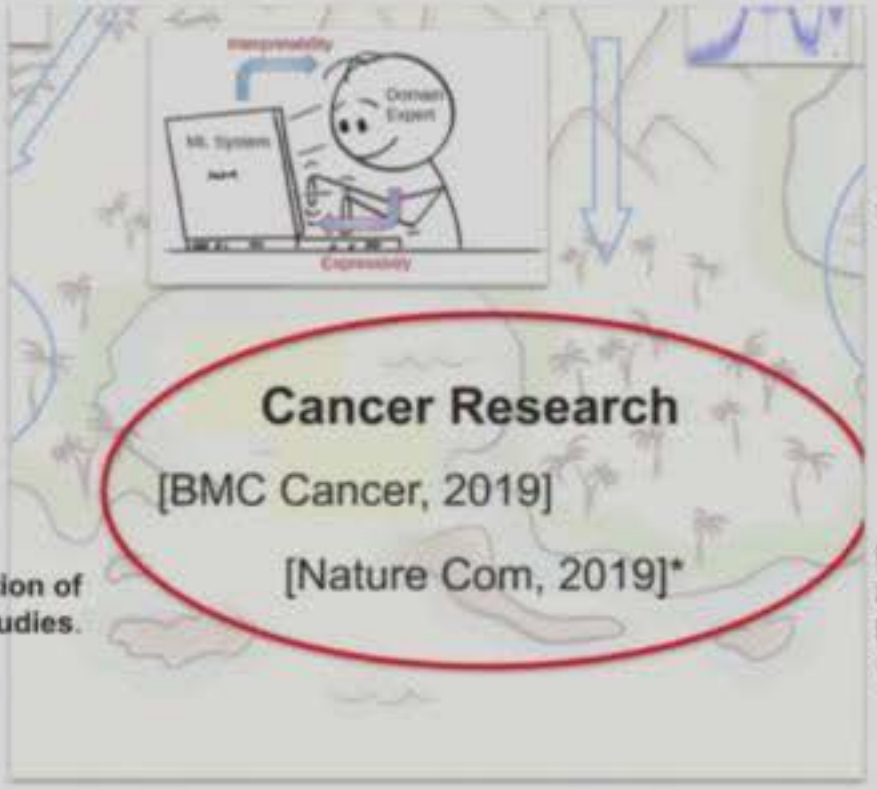


Diagram: Mechanism of codrituzumab-induced antibody-dependent cytotoxicity through the interaction of Fc CD16 in NK cells

What are the underlying mechanisms of cancer?

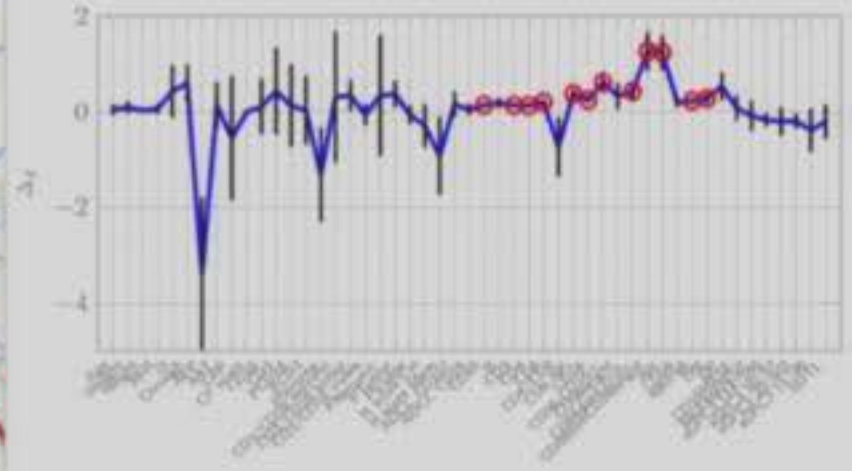




Cancer Research

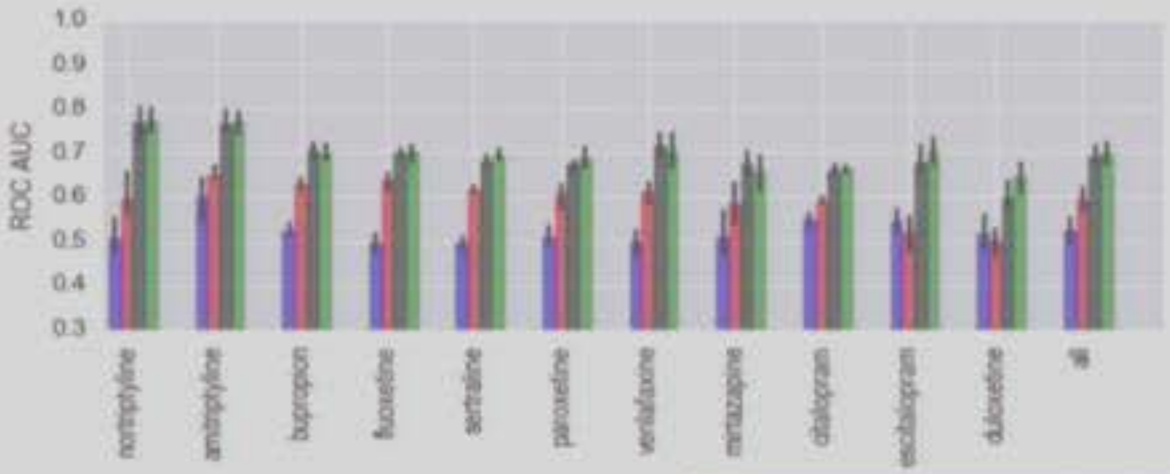
[BMC Cancer, 2019]

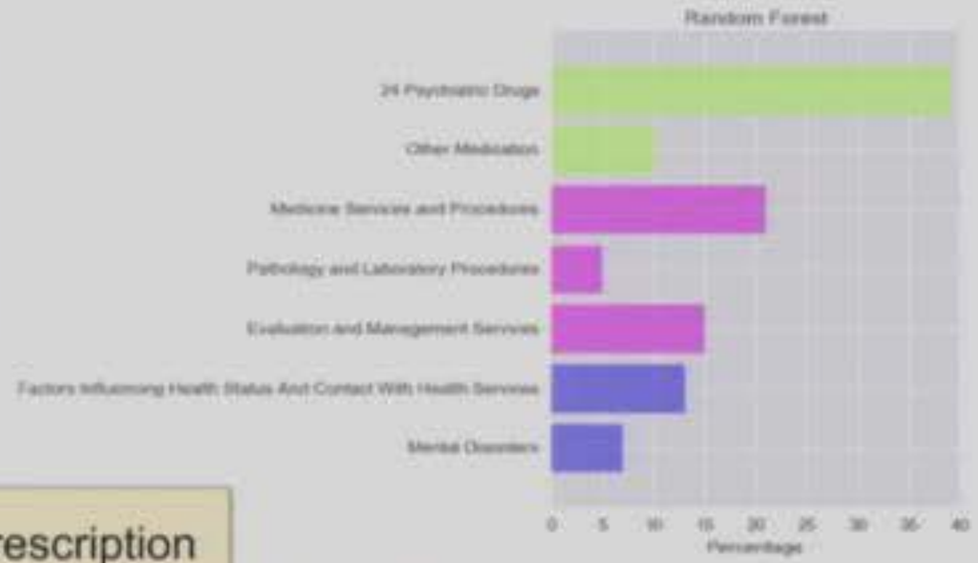
[Nature Com, 2019]*



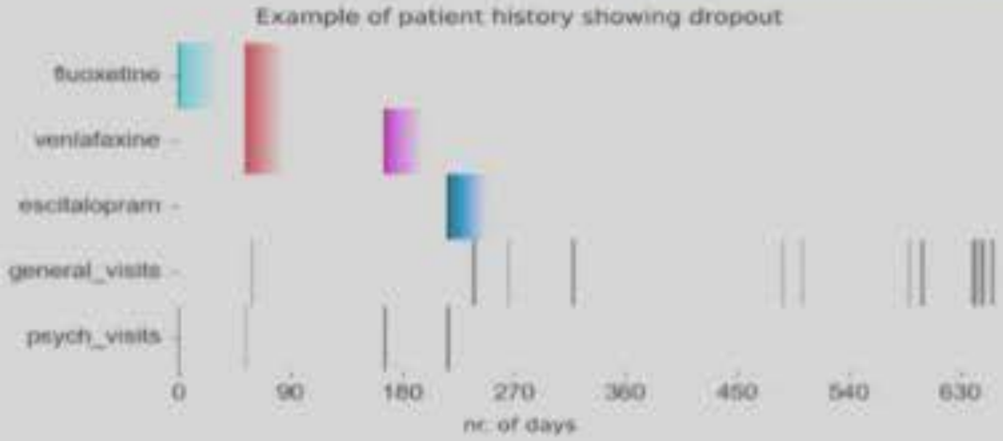
S. Stark, S. L. Hyland, M. F. Pradier, K. Lehmann, A. Wicki, F. Perez-Cruz, J. E. Vogt, and G. Ratsch. **Unsupervised Extraction of Phenotypes from Cancer Clinical Notes for Association Studies**. *In submission to Nature Communications*. 2019.

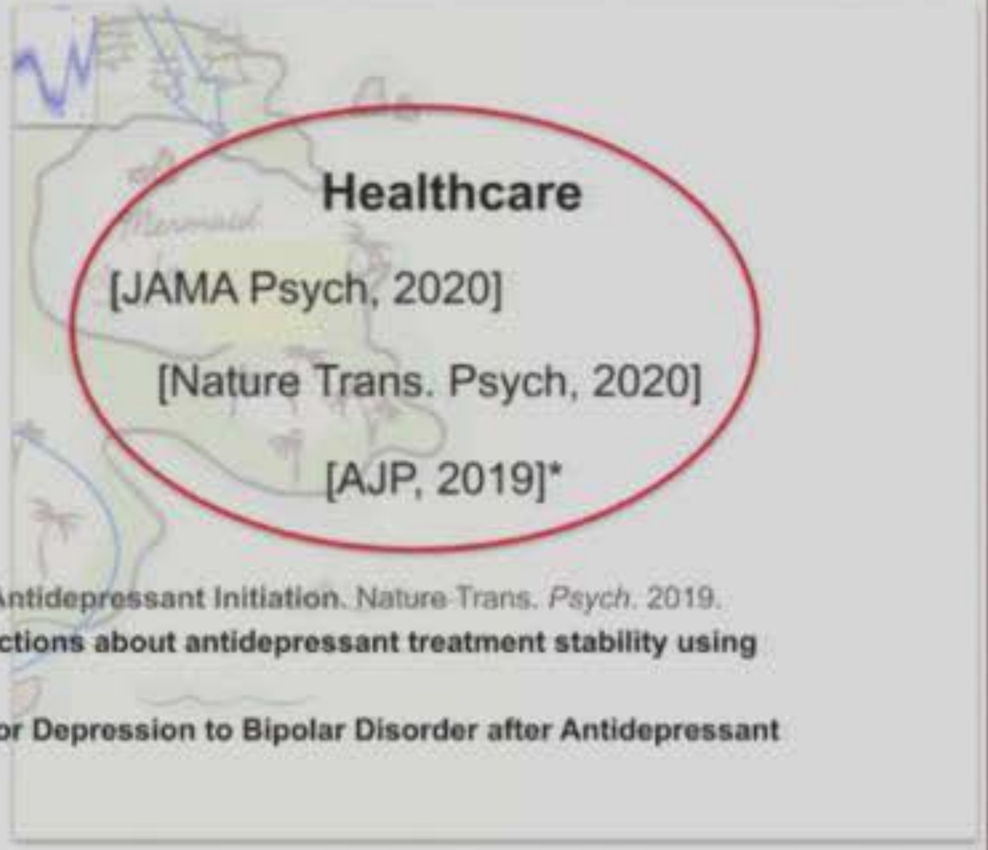
M. F. Pradier, B. Reis, L. Jukofsky, F. Milletti, T. Ohtomo, F. Perez-Cruz, and O. Puig. **Case-control Indian Buffet Process identifies biomarkers of response to Codrituzumab**. *BMC Cancer*. 2019.





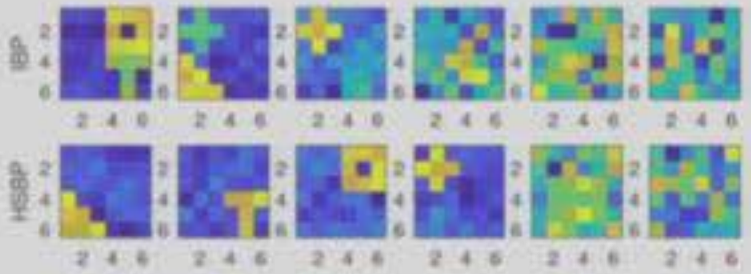
Can we personalize prescription of antidepressants?





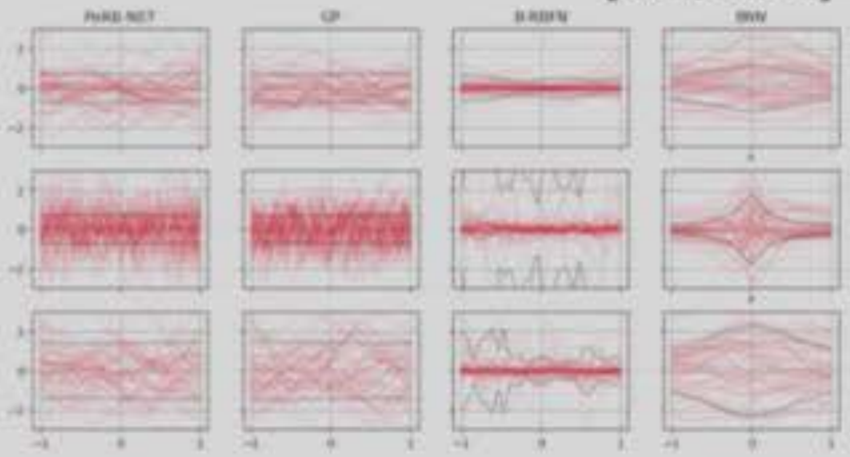
M. F. Pradier, T. H. McCoy, M. Hughes, R. H. Perlis and F. Doshi-Velez. **Predicting Treatment Discontinuation after Antidepressant Initiation.** Nature Trans. *Psych.* 2019.
M. C. Hughes, M. F. Pradier, A. S. Ross, T. H. McCoy, R. H. Perlis and F. Doshi-Velez. **Generating interpretable predictions about antidepressant treatment stability using supervised topic models**. *Accepted JAMA Psychiatry*. 2019.
M. F. Pradier, M. Hughes, T. H. McCoy, S. Barroilhet, F. Doshi-Velez and R. H. Perlis. **Predicting Transition from Mayor Depression to Bipolar Disorder after Antidepressant Initiation**. *Submitted to American Journal of Psychiatry*. 2019.

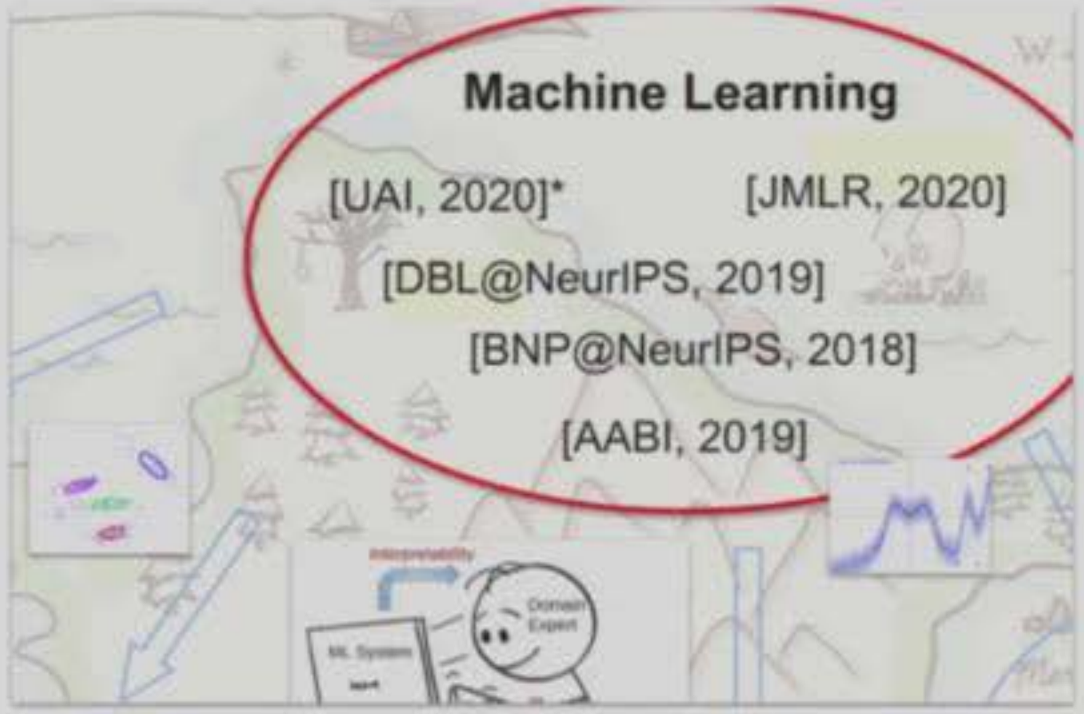
[BNP@NeurIPS, 2018]



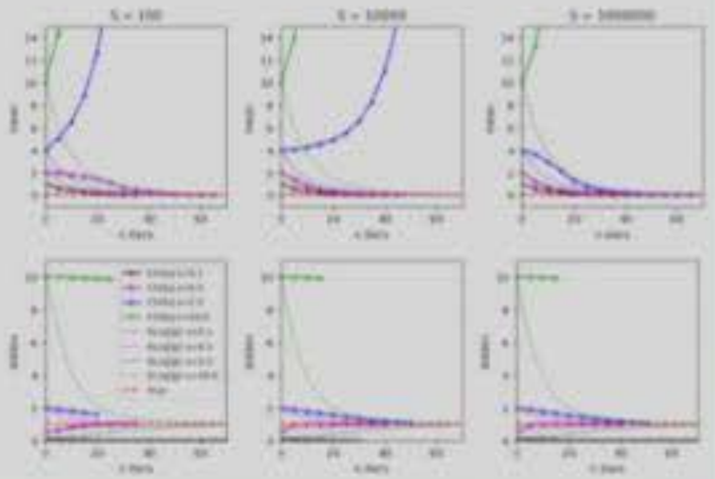
Can we design expressive priors?

[UAI, 2020]*



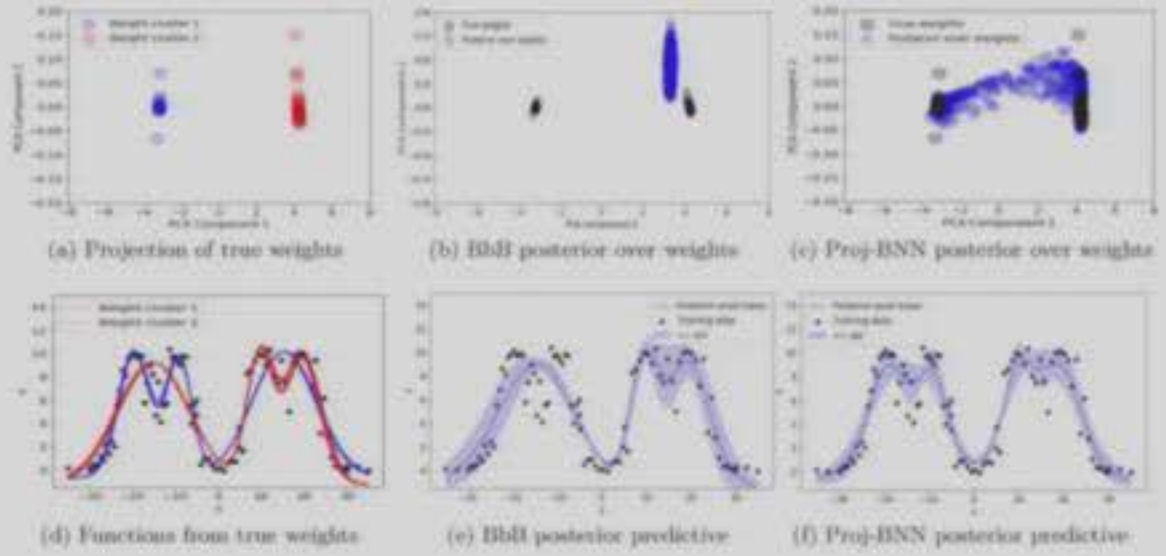


[AABI, 2019]



Can we improve inference?

[DBL@NeurIPS, 2019]



(a) Projection of true weights (b) BbB posterior over weights (c) Proj-BNN posterior over weights

(d) Functions from true weights (e) BbB posterior predictive (f) Proj-BNN posterior predictive

My contributions

Machine Learning
[UAI, 2020]* [JMLR, 2020]
[DBL@NeurIPS, 2019]
[BNP@NeurIPS, 2018]
[AABI, 2019]

Information Theory
[Entropy, 2016]

Sport Sciences
[PlosOne, 2016]

Economics
[PlosOne, 2018]
[ICASSP, 2018]

Cancer Research
[BMC Cancer, 2019]
[Nature Com, 2019]*

Healthcare
[JAMA Psych, 2020]
[Nature Trans. Psych, 2020]
[AJP, 2019]*

My contributions

Machine Learning
[UAI, 2020]* [JMLR, 2020]
[DBL@NeurIPS, 2019]
[BNP@NeurIPS, 2018]
[AABI, 2019]

Information Theory
[Entropy, 2016]

Sport Sciences
[PlosOne, 2016]

Healthcare
[JAMA Psych, 2020]
[Nature Trans. Psych, 2020]
[AJP, 2019]*

Economics
[PlosOne, 2018]
[ICASSP, 2018]

Cancer Research
[BMC Cancer, 2019]
[Nature Com, 2019]*

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

Application: Phase II clinical trial for immunotherapy for liver cancer

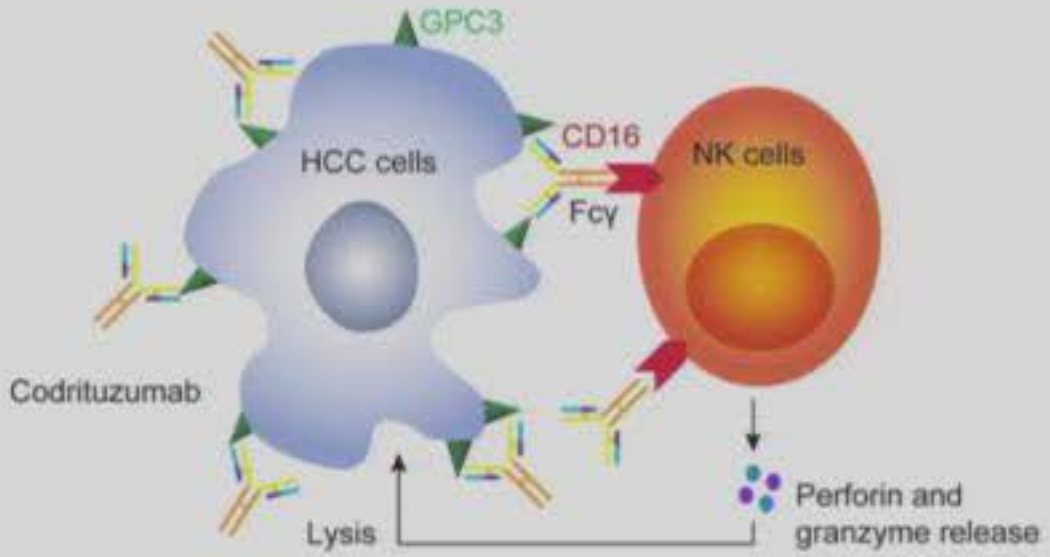


Diagram: Mechanism of codrituzumab-induced antibody-dependent cytotoxicity through the interaction of Fc CD16 in NK cells

[Abou-Alfa et.al, 2016]

- **Method**: classical statistical tests
- **Conclusion**: no evidence for treatment effectiveness

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

Application: Phase II clinical trial for immunotherapy for liver cancer

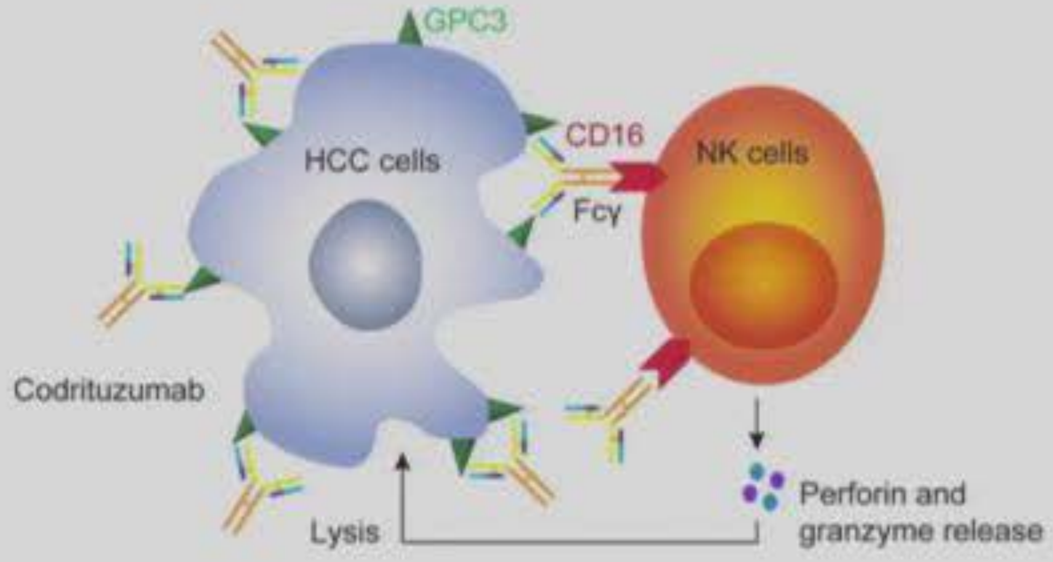
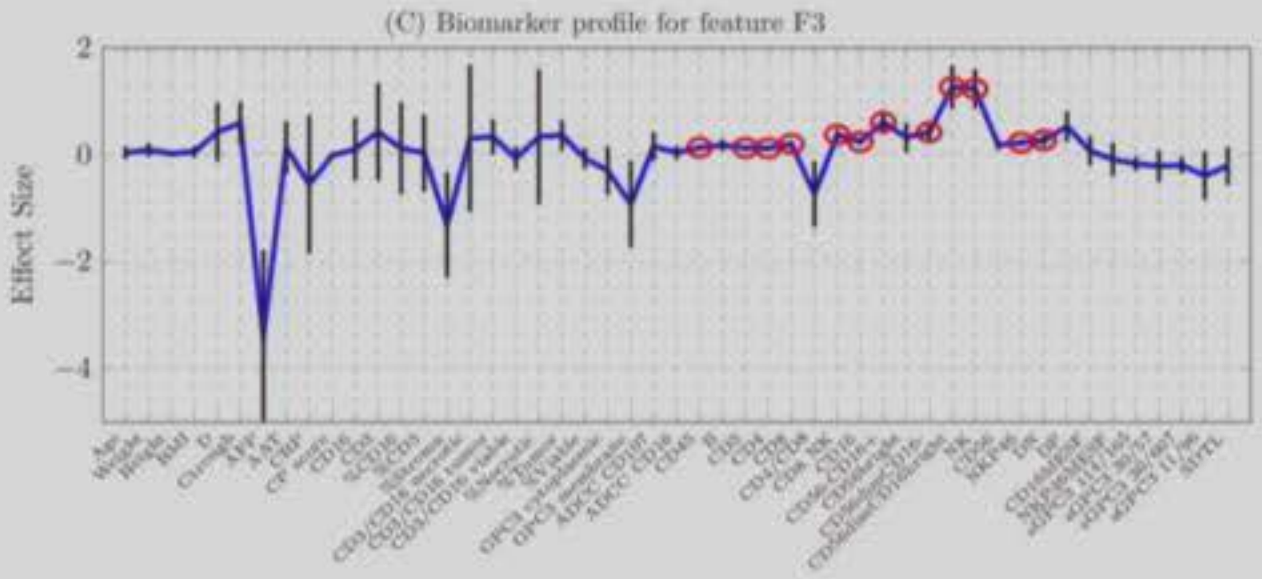


Diagram: Mechanism of codrituzumab-induced antibody-dependent cytotoxicity through the interaction of Fc CD16 in NK cells



[Abou-Alfa et.al, 2016]

- **Method**: classical statistical tests
- **Conclusion**: no evidence for treatment effectiveness

[BMC Cancer, 2019]

- **Method**: novel Bayesian approach
- **Conclusion**: treatment effective for subgroup

Why is this problem hard?

Challenges

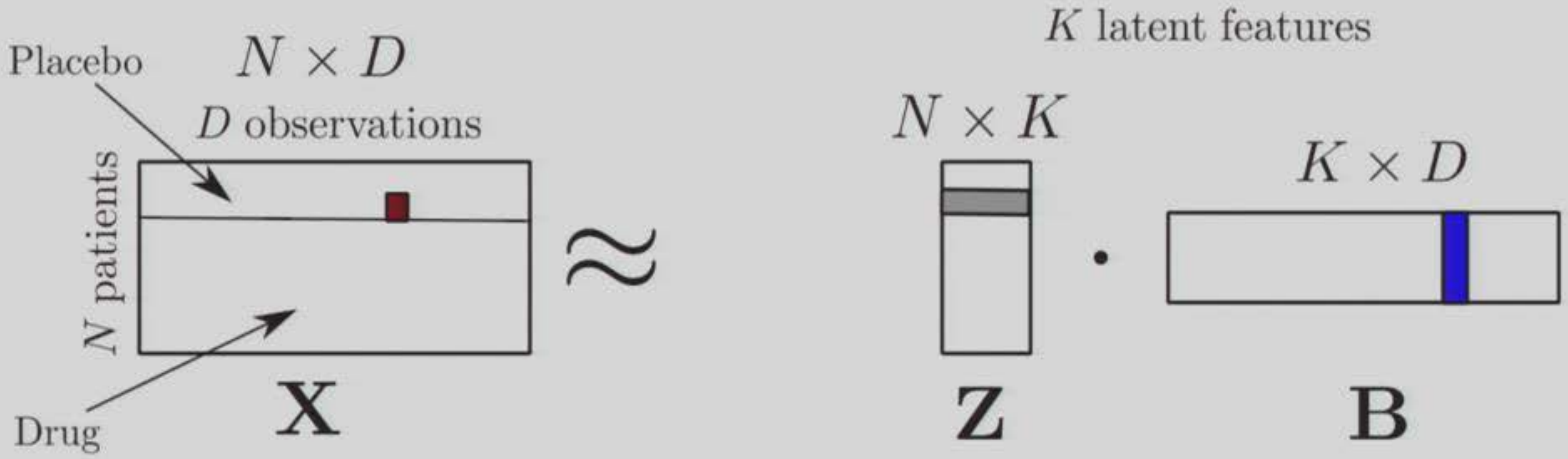
Why is this problem hard?

Challenges

1. complex correlations
2. patient heterogeneity
3. few data available
4. very different data types
5. drug effect vs natural response

Challenge 1: how to deal with complex correlations?

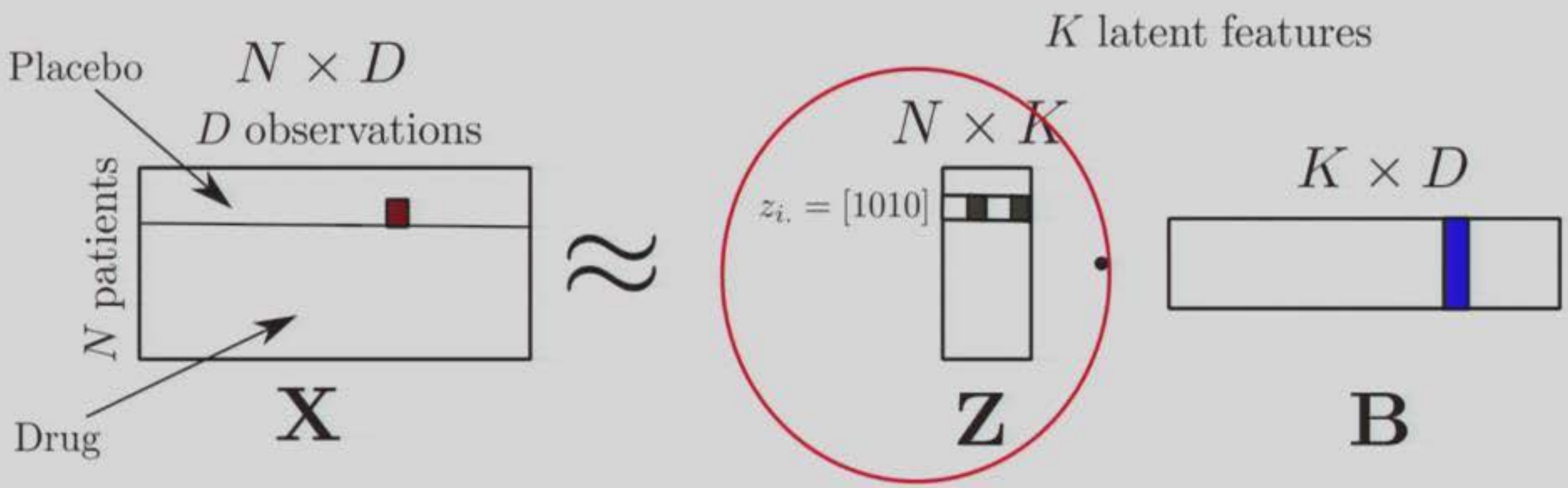
Our approach: latent feature model



$$\blacksquare\ x_{id} = 173\,\mathrm{ml/dL} = 73 + 0 + 100\,\mathrm{ml/dL}$$

Challenge 2: how to deal with patient heterogeneity?

Our approach: binary latent feature model



- Binary latent feature active: correlation pattern applies to patient

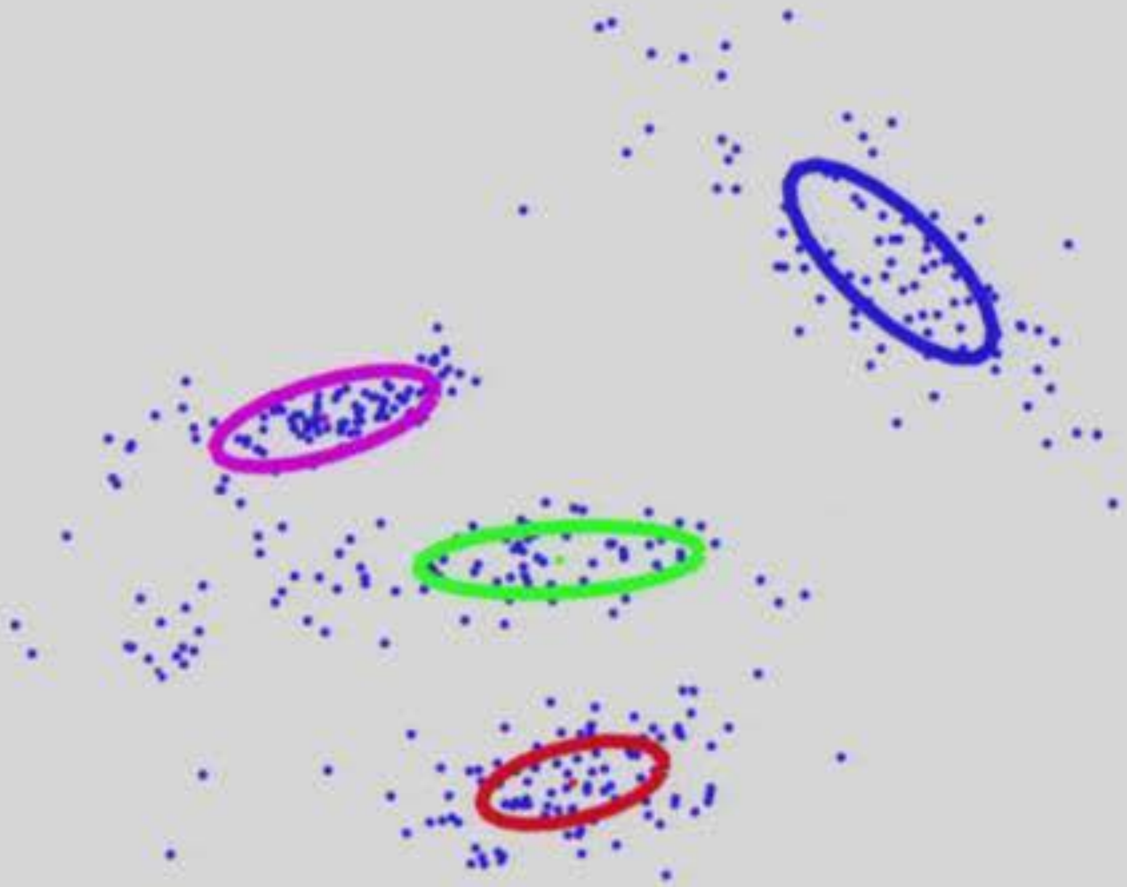
Challenge 3: how to deal with little data?

Our approach: Bayesian nonparametrics

- **Bayesian**: to handle uncertainty

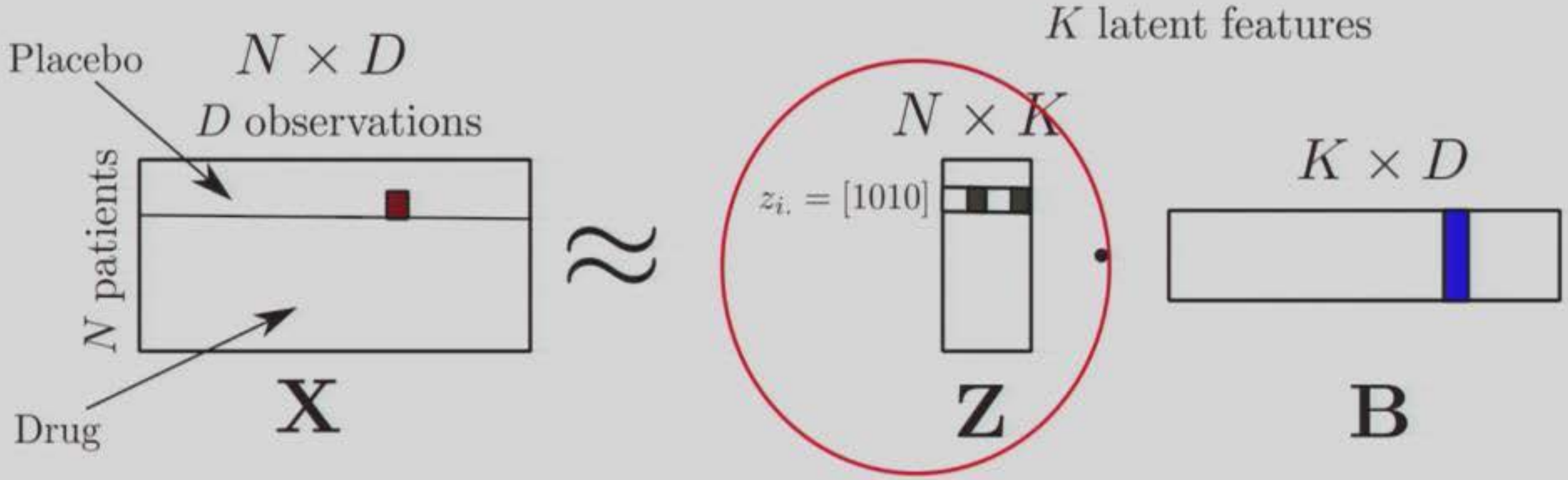
$$p(\text{parameters}|\text{data}) \propto p(\text{data}|\text{parameters})p(\text{parameters})$$

- **Nonparametric**: to adapt model complexity (hypothesis generation)



Challenge 2: how to deal with patient heterogeneity?

Our approach: binary latent feature model



- Binary latent feature active: correlation pattern applies to patient

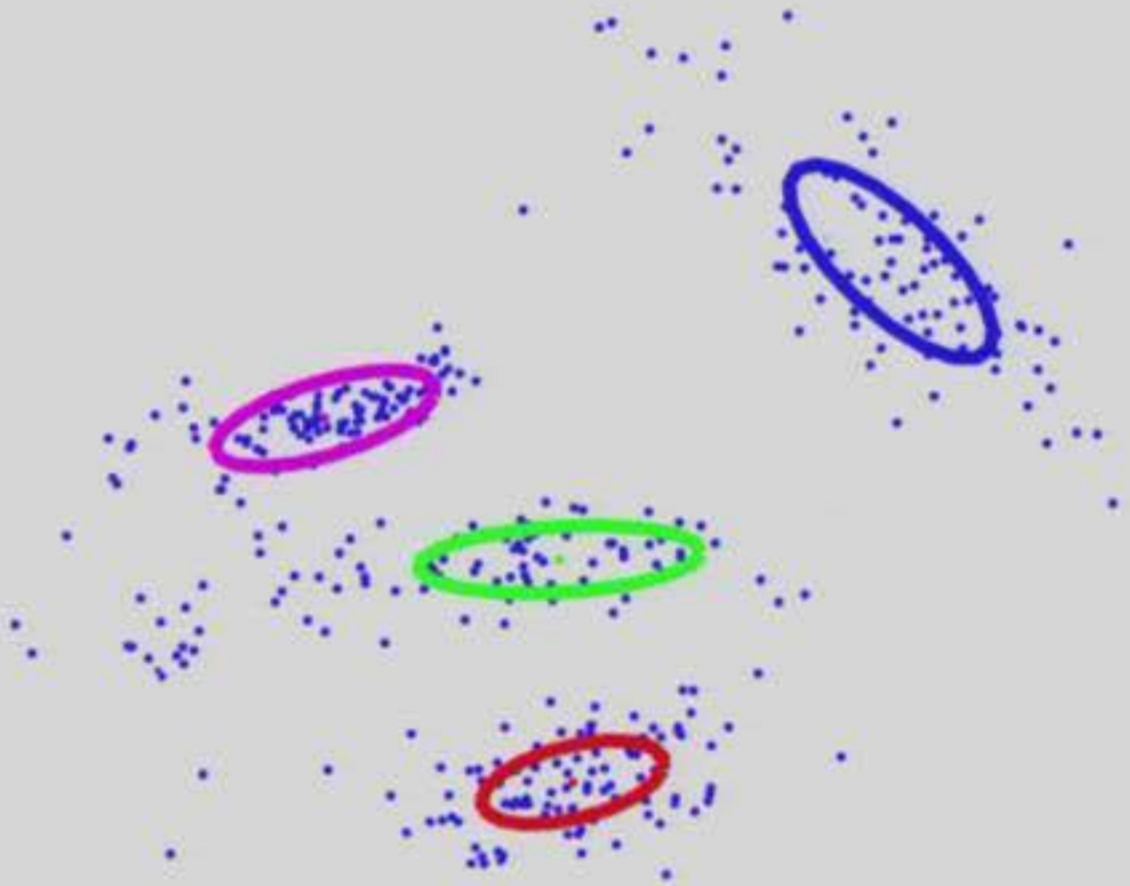
Challenge 3: how to deal with little data?

Our approach: Bayesian nonparametrics

- **Bayesian**: to handle uncertainty

$$p(\text{parameters}|\text{data}) \propto p(\text{data}|\text{parameters})p(\text{parameters})$$

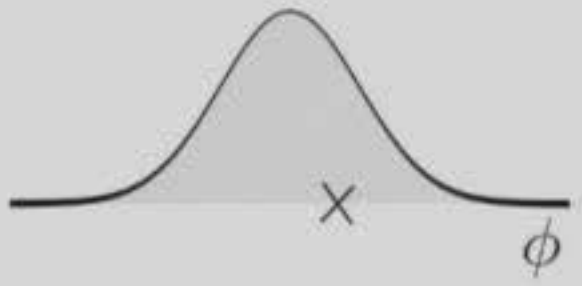
- **Nonparametric**: to adapt model complexity (hypothesis generation)



Challenge 3: how to deal with little data?

Our approach: Bayesian nonparametrics

- Indian Buffet Process [Ghahramani et.al, 2006]

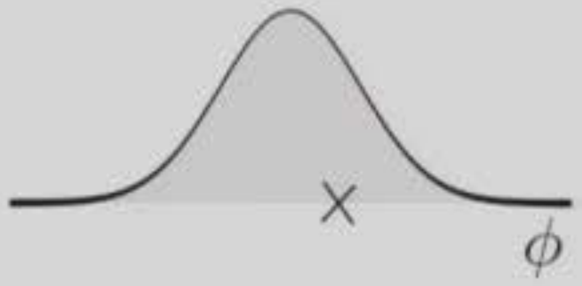


$x \sim \mathcal{N}(\mu_0, \sigma_0^2)$ $\longrightarrow$ $x = 0.356$

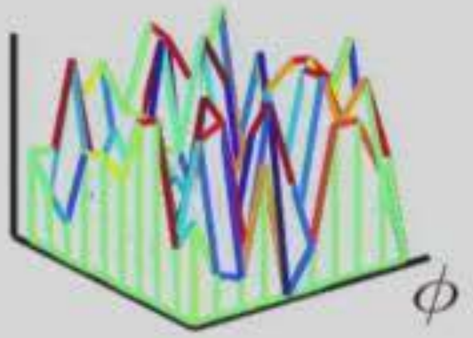
Challenge 3: how to deal with little data?

Our approach: Bayesian nonparametrics

- Indian Buffet Process [Ghahramani et.al, 2006]



$$\xrightarrow[x \sim \mathcal{N}(\mu_0, \sigma_0^2)]{} \quad x = 0.356$$



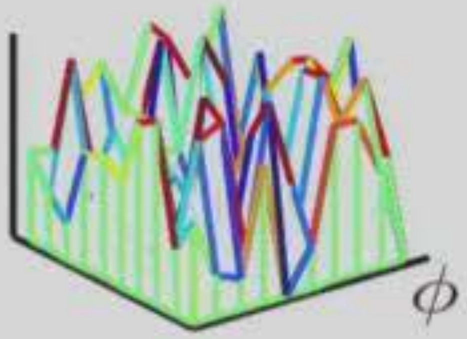
$$\xrightarrow[Z \sim \mathrm{IBP}(\alpha)]{} \quad Z = \begin{bmatrix} 1 & 0 & 1 & \\ 0 & 1 & 0 & \vdots \\ 1 & 1 & 0 & \\ 1 & 0 & 1 & \end{bmatrix}$$

Challenge 3: how to deal with little data?

Our approach: Bayesian nonparametrics

- Indian Buffet Process [Ghahramani et.al, 2006]

$$\xrightarrow[x \sim \mathcal{N}(\mu_0, \sigma_0^2)]{} \quad x = 0.356$$

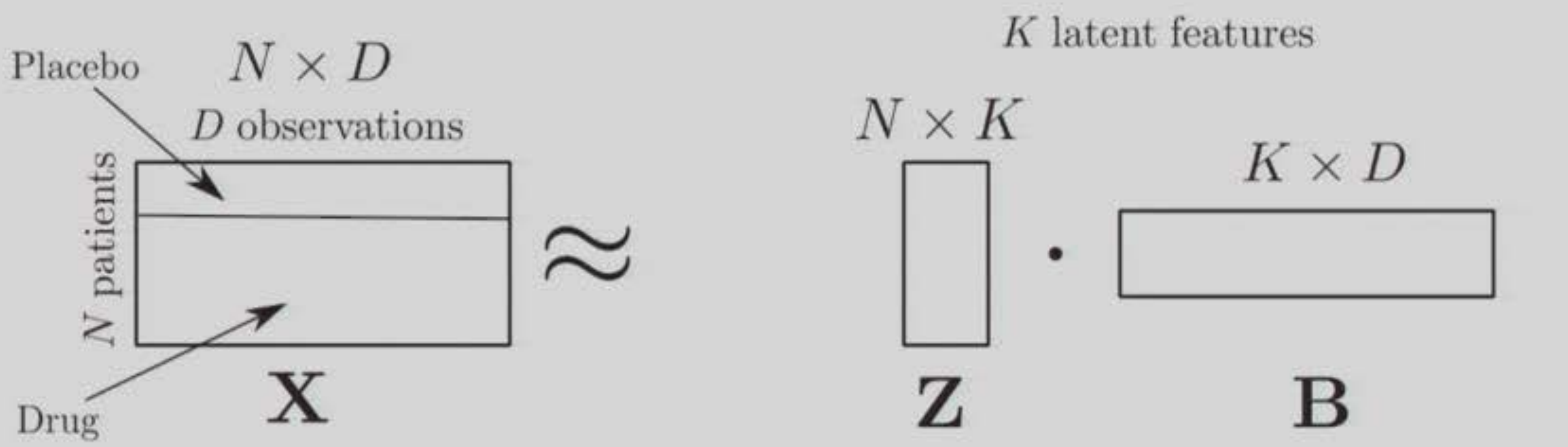


$$\xrightarrow[Z \sim \mathrm{IBP}(\alpha)]{} \quad Z = \begin{bmatrix} 1 & 0 & 1 & \\ 0 & 1 & 0 & \vdots \\ 1 & 1 & 0 & \\ 1 & 0 & 1 & \end{bmatrix}$$

- Prior over binary matrices with infinite number of columns
- $\mathbf{Z} \sim \mathrm{IBP}(\alpha)$, where α: concentration parameter

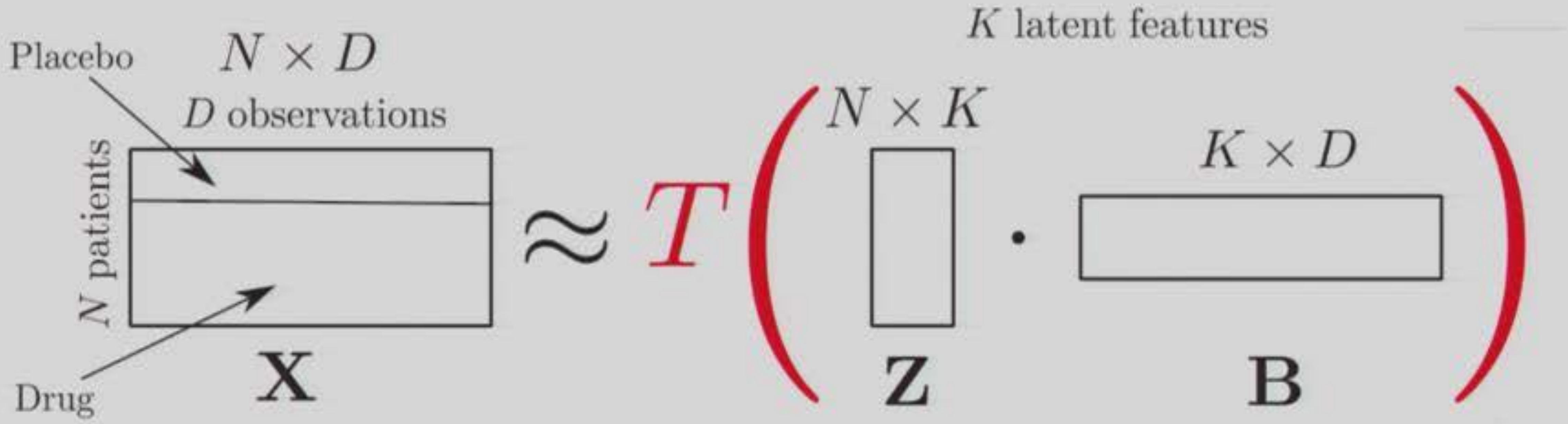
Challenge 4: how to deal with different data types?

Our approach: use link functions



Challenge 4: how to deal with different data types?

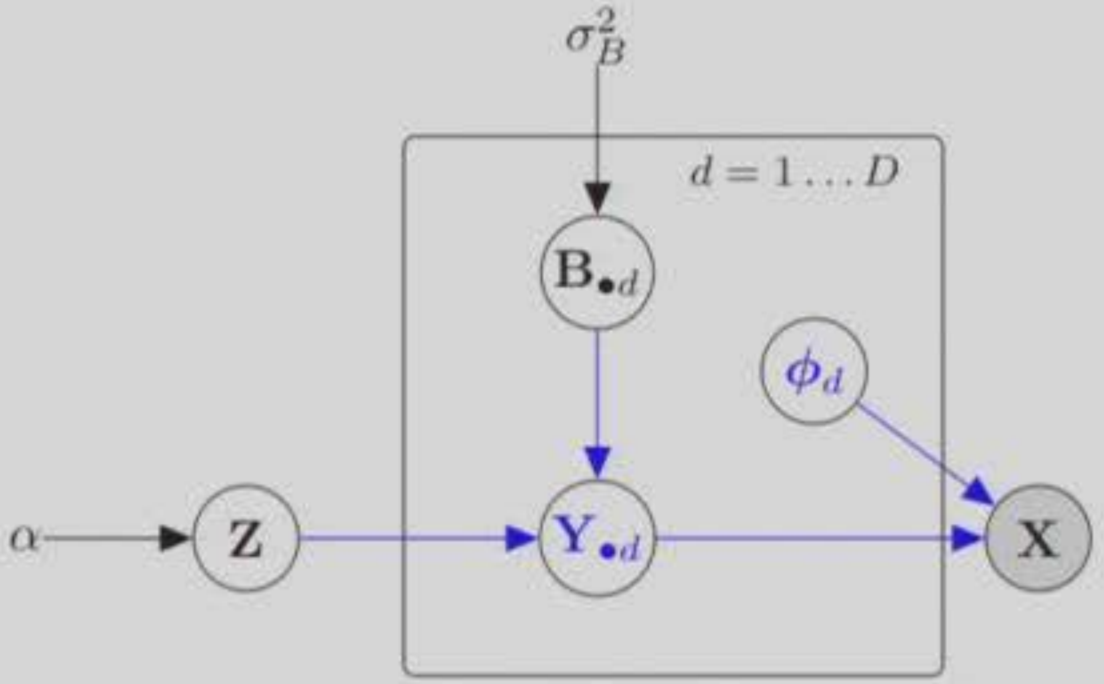
Our approach: use link functions



- Link functions T_d for each observed feature d
- Each T_d depends on type of data

Solving all 4 challenges together

General latent feature model (GLFM) [JMLR, 2020]



$$
\begin{aligned}
x_{nd} &= T_d(y_{nd}; \boldsymbol{\phi_d}) \\
y_{nd}|\mathbf{Z},\mathbf{B} &\sim \mathcal{N}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d}, \sigma_y^2) \\
B_{kd} &\sim \mathcal{N}(0, \sigma_B^2) \\
\mathbf{Z} &\sim \mathrm{IBP}(\alpha)
\end{aligned}
$$

Infinite latent feature model for heterogeneous datasets

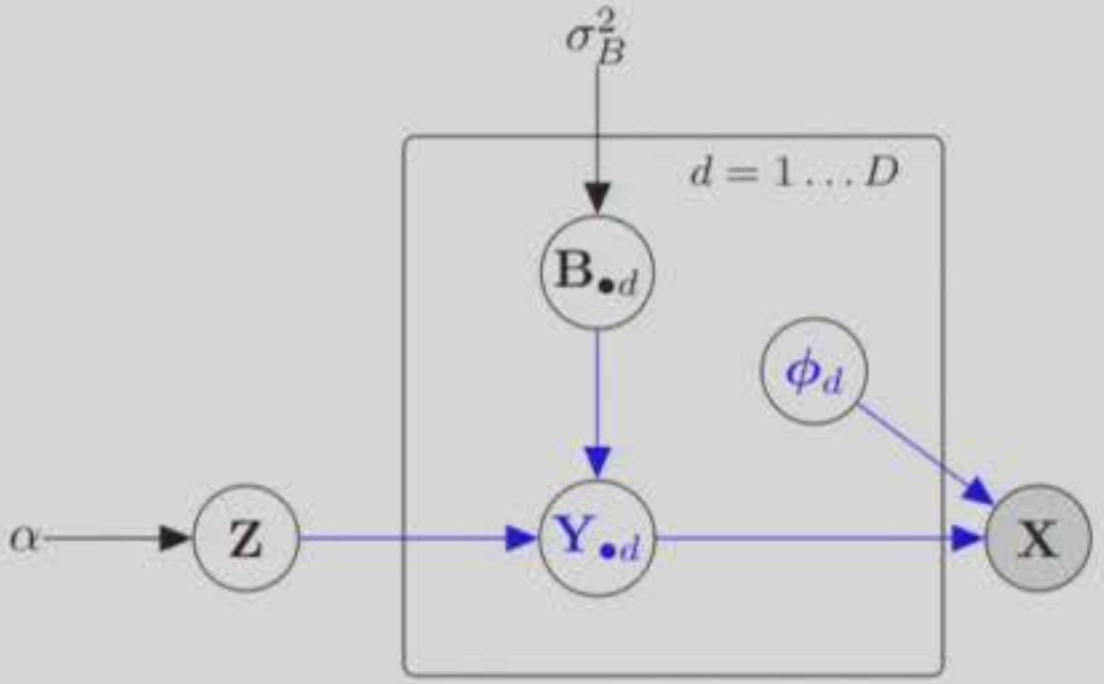
GLFM package

- Open-source python/matlab/R code
- Discovers complex correlations
- Is able to deal with data heterogeneity, noisy observations, missing data, different data types
- Adjusts model complexity

https://github.com/ivaleraM/GLFM

Solving all 4 challenges together

General latent feature model (GLFM) [JMLR, 2020]



$$
\begin{aligned}
x_{nd} &= T_d(y_{nd}; \boldsymbol{\phi_d}) \\
y_{nd}|\mathbf{Z}, \mathbf{B} &\sim \mathcal{N}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d}, \sigma_y^2) \\
B_{kd} &\sim \mathcal{N}(0, \sigma_B^2) \\
\mathbf{Z} &\sim \mathrm{IBP}(\alpha)
\end{aligned}
$$

Infinite latent feature model for heterogeneous datasets

GLFM package

- Open-source python/matlab/R code
- Discovers complex correlations
- Is able to deal with data heterogeneity, noisy observations, missing data, different data types
- Adjusts model complexity

https://github.com/ivaleraM/GLFM

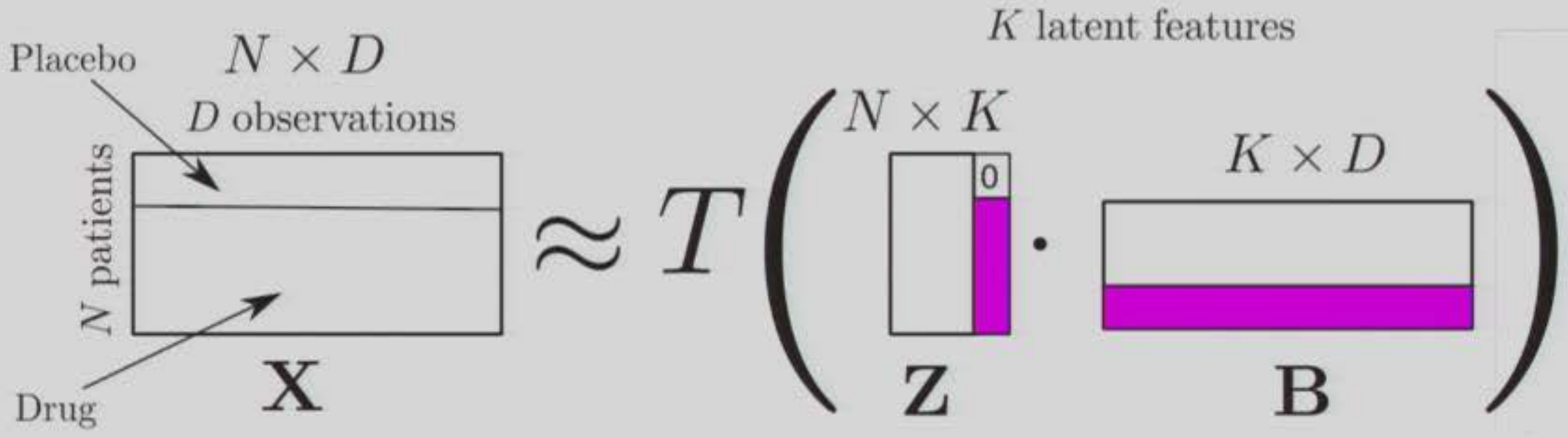
Revisiting list of challenges

Challenges

1. complex correlations
2. patient heterogeneity
3. few data available
4. very different data types
5. drug effect vs natural response

Challenge 5: drug effect vs natural response?

Our approach: structured prior

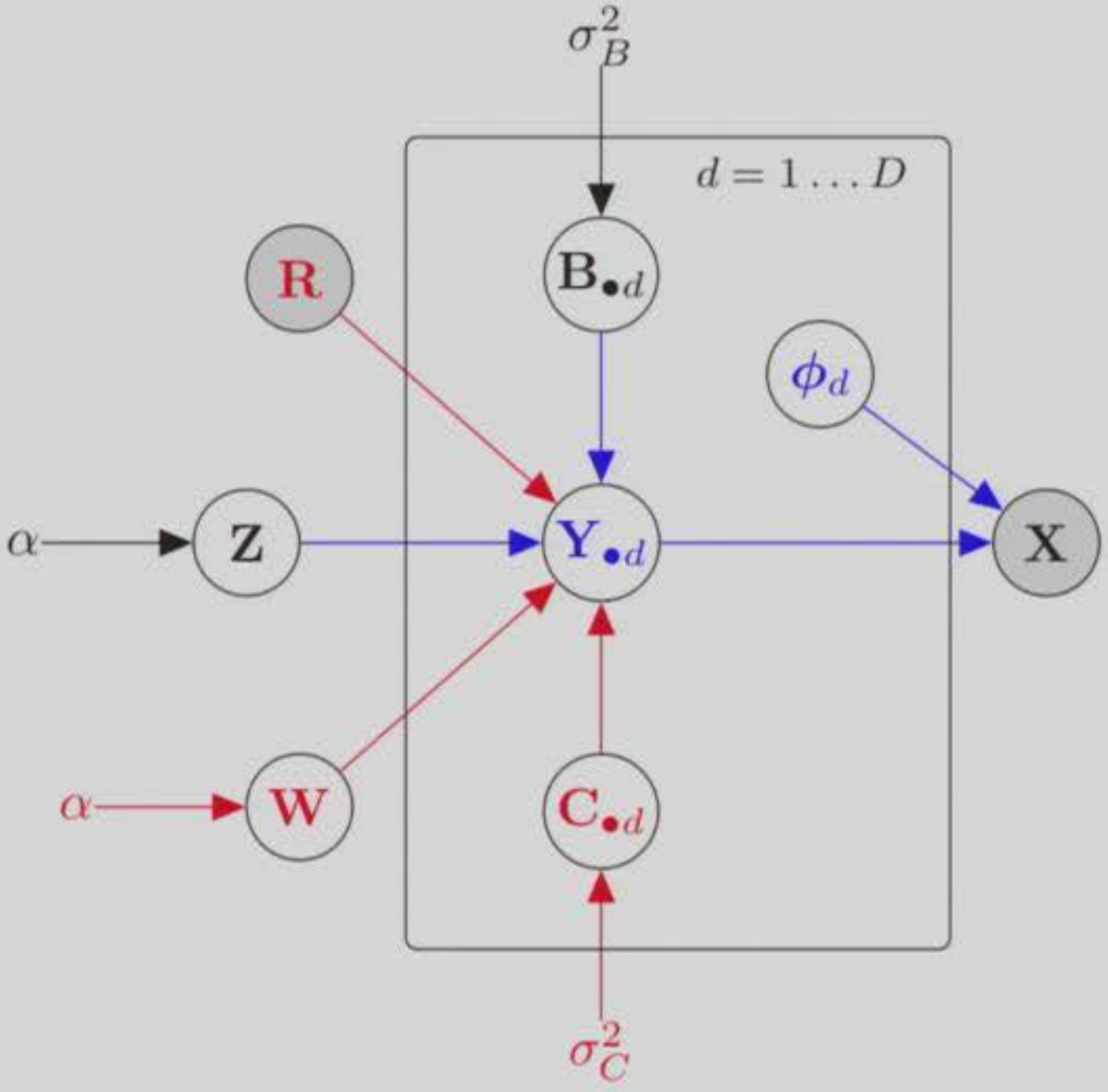


Treatment-specific latent features

- can only activate for patients in treatment arm
- number learned automatically

How to distinguish drug effect vs natural response?

Case-control Indian Buffet Process (C-IBP) [BMC Cancer, 2019]



R_n: drug indicator por patient n

$$x_{nd} = T_d(y_{nd}; \boldsymbol{\phi_d})$$

$$y_{nd} | \mathbf{Z}, \mathbf{W}, \mathbf{B}, \mathbf{C}, \mathbf{R} \sim$$

$$\mathcal{N}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d} + \mathbb{1}[\mathbf{R}_n = 1]\mathbf{W}_{n\bullet}\mathbf{C}_{\bullet d}, \sigma_y^2)$$

$$B_{kd} \sim \mathcal{N}(0, \sigma_B^2)$$

$$\mathbf{Z} \sim \mathrm{IBP}(\alpha)$$

$$C_{kd} \sim \mathcal{N}(0, \sigma_C^2)$$

$$\mathbf{W} \sim \mathrm{IBP}(\alpha)$$

- **Inference**: MCMC approach with accelerated Gibbs sampling
- **Biomarker discovery**: statistical multiple hypothesis testing

Results: subpopulations

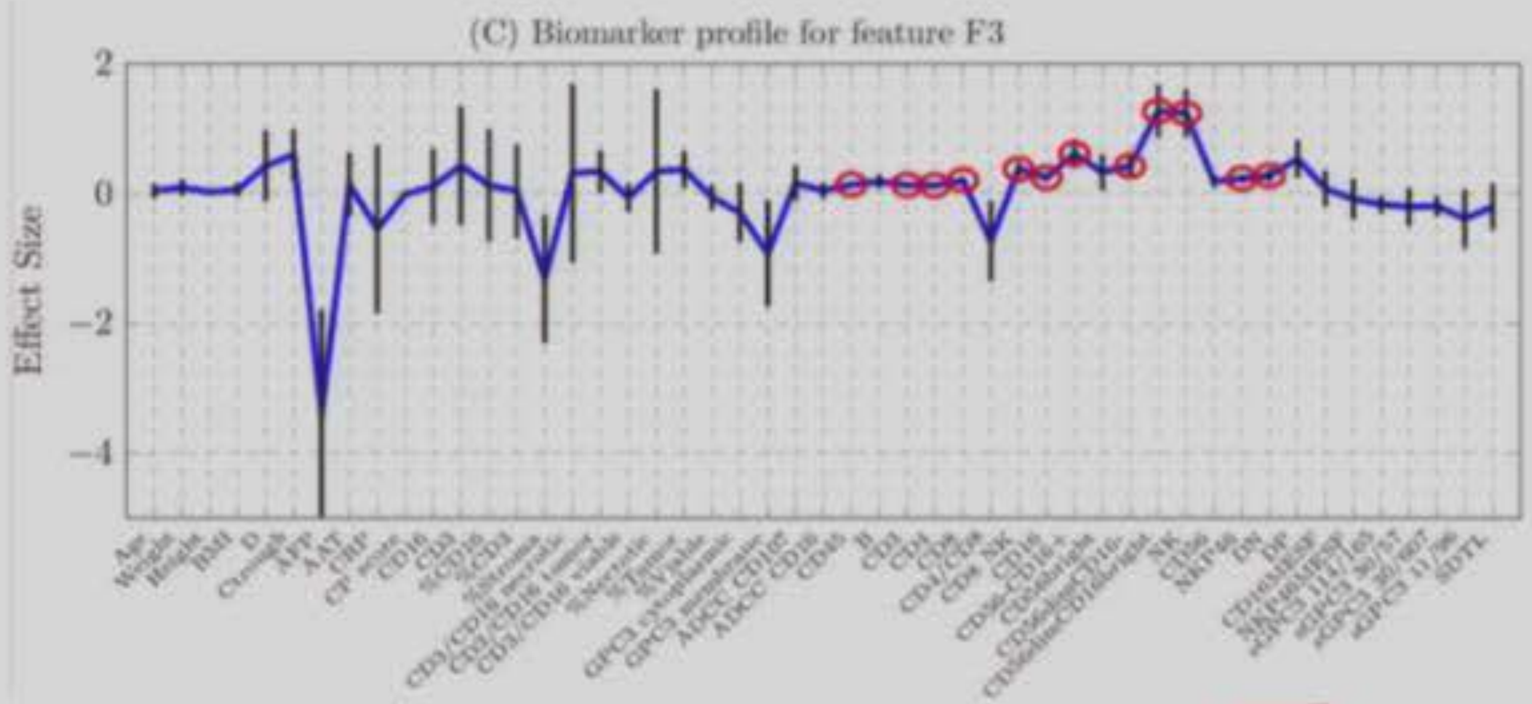
GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Results: biomarker discovery

Treatment-specific feature F3



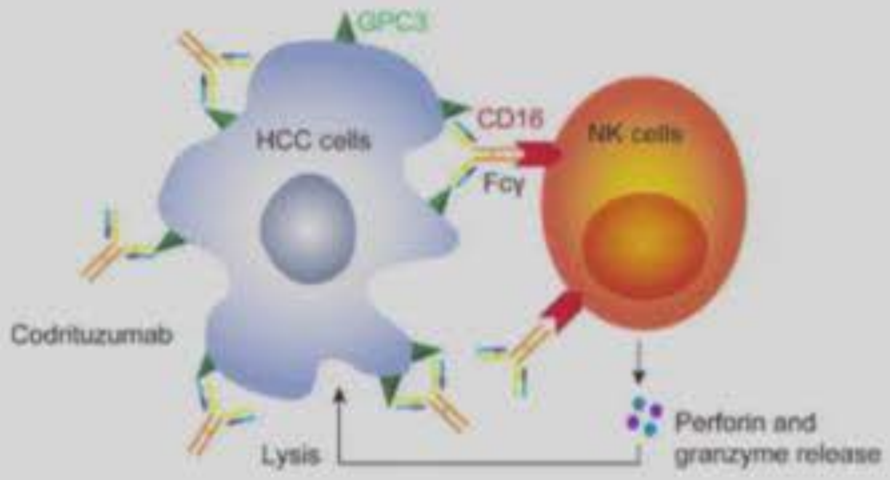


Diagram: Mechanism of codrituzumab-induced antibody-dependent cytotoxicity through the interaction of Fc CD16 in NK cells

M. F. Pradier, B. Reis, L. Jukofsky, F. Milletti, T. Ohtomo, F. Perez-Cruz, and O. Puig. **Case-control Indian Buffet Process identifies biomarkers of response to Codrituzumab**. *BMC Cancer*. 2019.

What did we find?

- Subgroup for which treatment is especially effective
- Relevant biomarkers (drug acting as expected)

Results: subpopulations

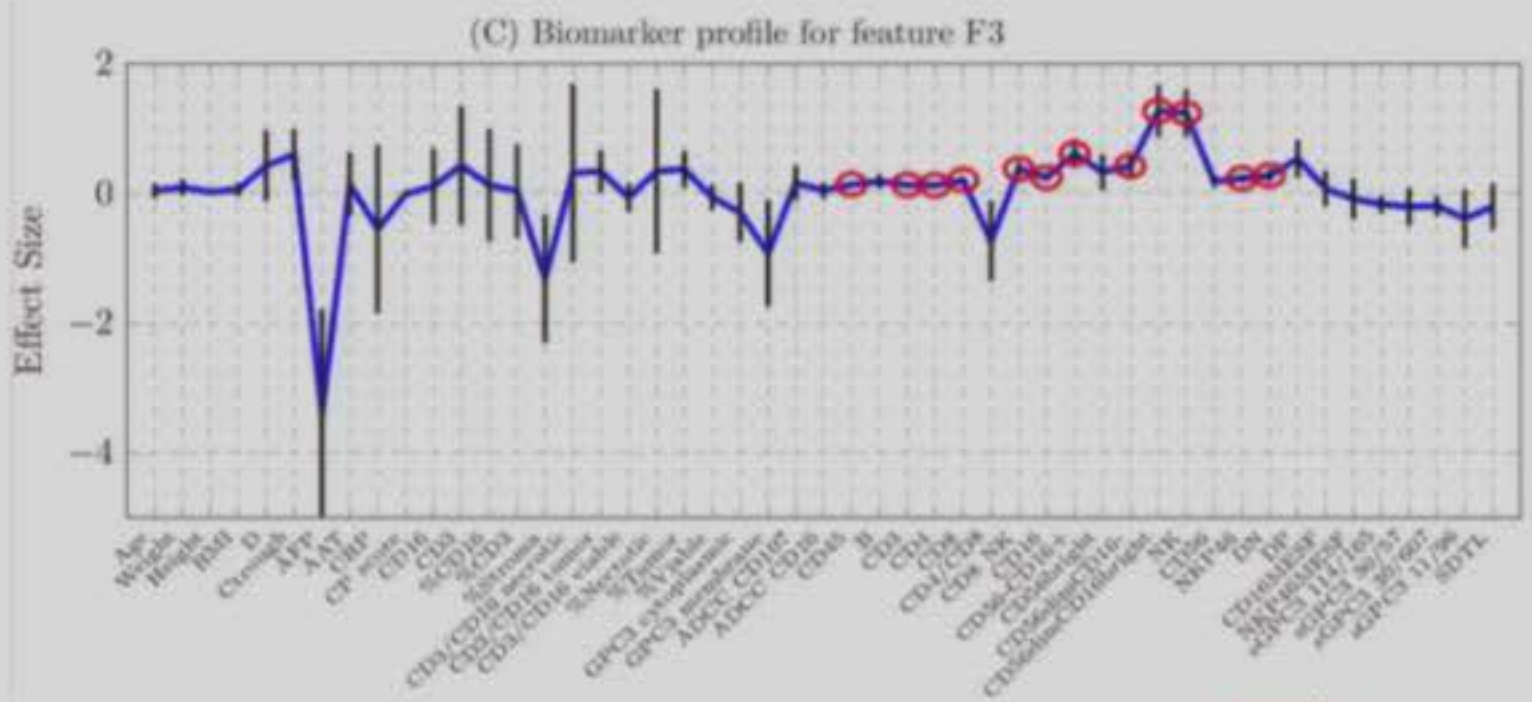
GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Results: biomarker discovery

Treatment-specific feature F3



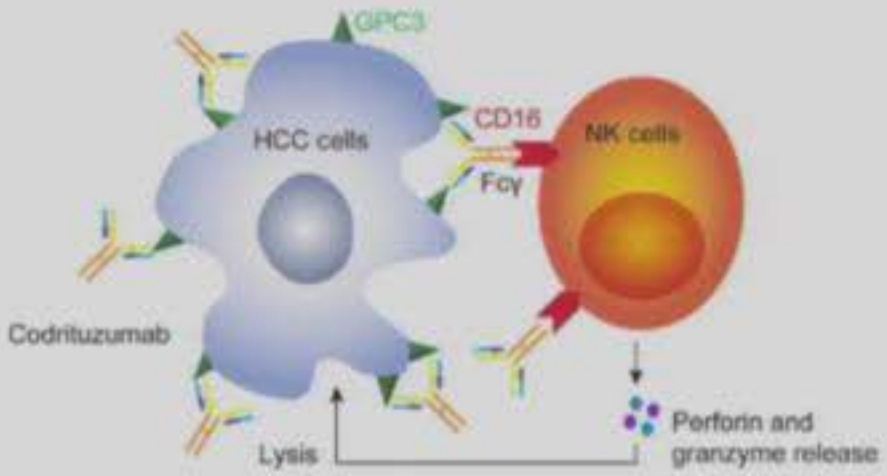


Diagram: Mechanism of codrituzumab-induced antibody-dependent cytotoxicity through the interaction of Fc CD16 in NK cells

M. F. Pradier, B. Reis, L. Jukofsky, F. Milletti, T. Ohtomo, F. Perez-Cruz, and O. Puig. **Case-control Indian Buffet Process identifies biomarkers of response to Codrituzumab**. *BMC Cancer*. 2019.

What did we find?

- Subgroup for which treatment is especially effective
- Relevant biomarkers (drug acting as expected)

Results: subpopulations

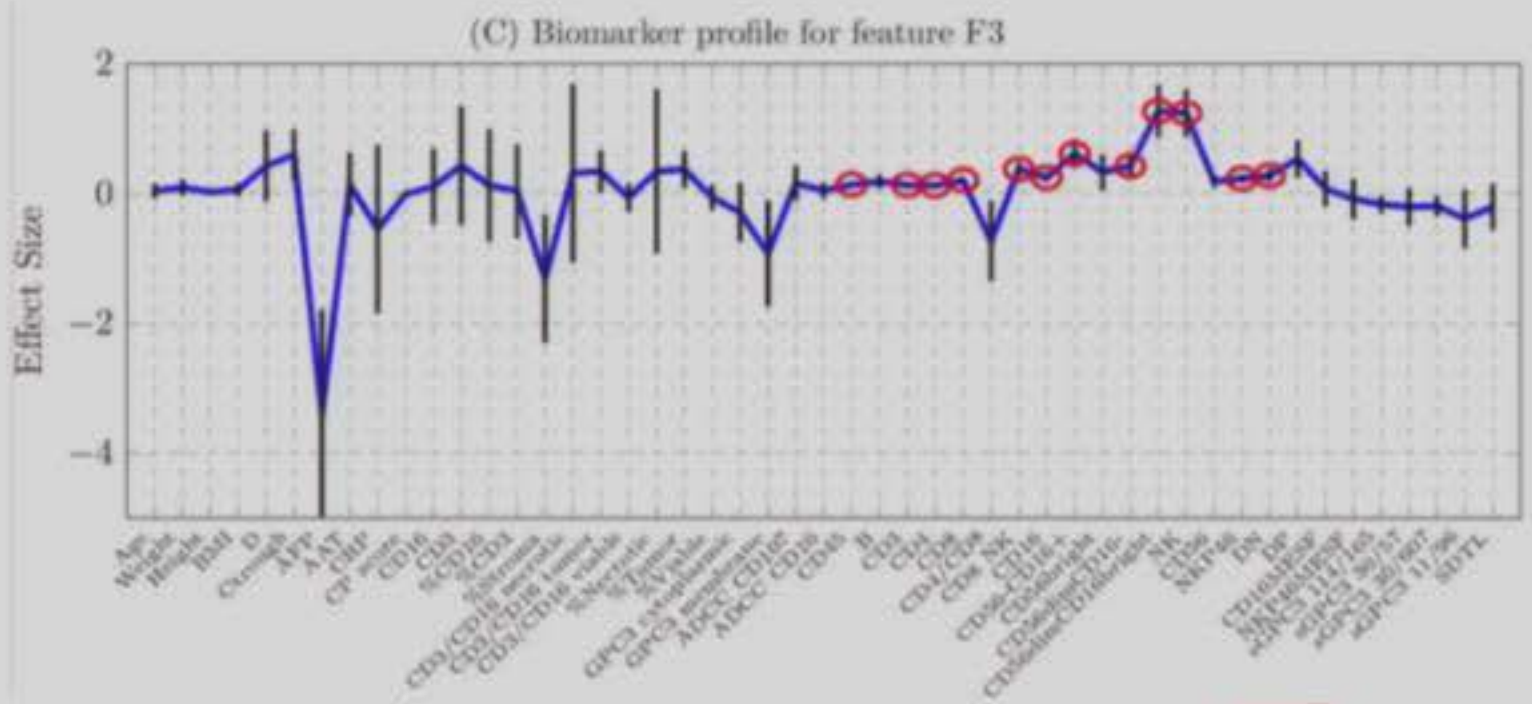
GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Results: biomarker discovery

Treatment-specific feature F3



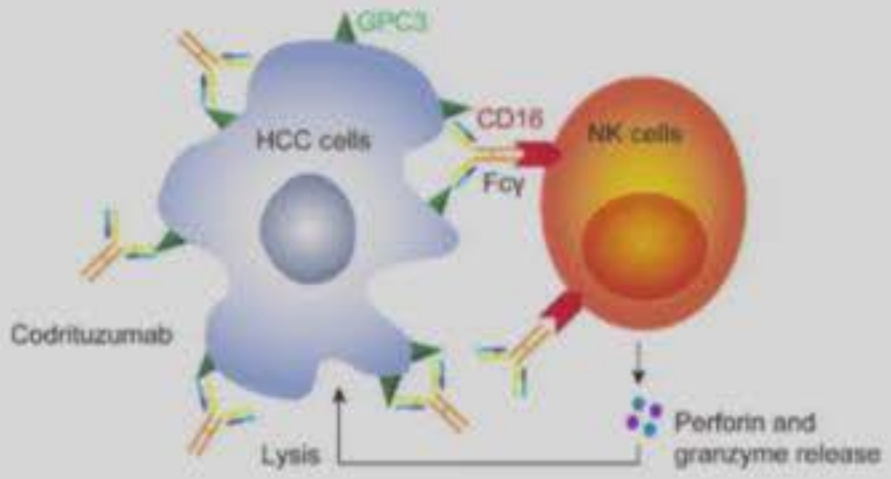


Diagram: Mechanism of codrituzumab-induced antibody-dependent cytotoxicity through the interaction of Fc CD16 in NK cells

M. F. Pradier, B. Reis, L. Jukofsky, F. Milletti, T. Ohtomo, F. Perez-Cruz, and O. Puig. **Case-control Indian Buffet Process identifies biomarkers of response to Codrituzumab**. *BMC Cancer*. 2019.

What did we find?

- Subgroup for which treatment is especially effective
- Relevant biomarkers (drug acting as expected)

Results: subpopulations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

My contributions

Machine Learning
[UAI, 2020]* [JMLR, 2020]
[DBL@NeurIPS, 2019]
[BNP@NeurIPS, 2018]
[AABI, 2019]

Information Theory
[Entropy, 2016]

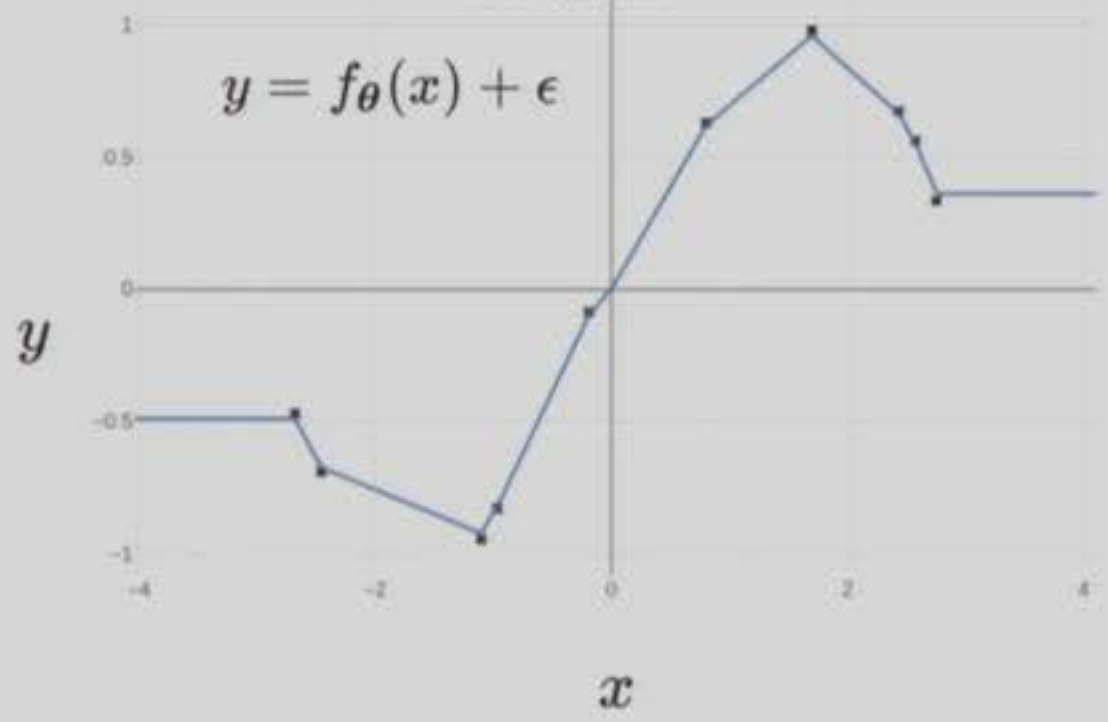
Sport Sciences
[PlosOne, 2016]

Economics
[PlosOne, 2018]
[ICASSP, 2018]

Cancer Research
[BMC Cancer, 2019]
[Nature Com, 2019]*

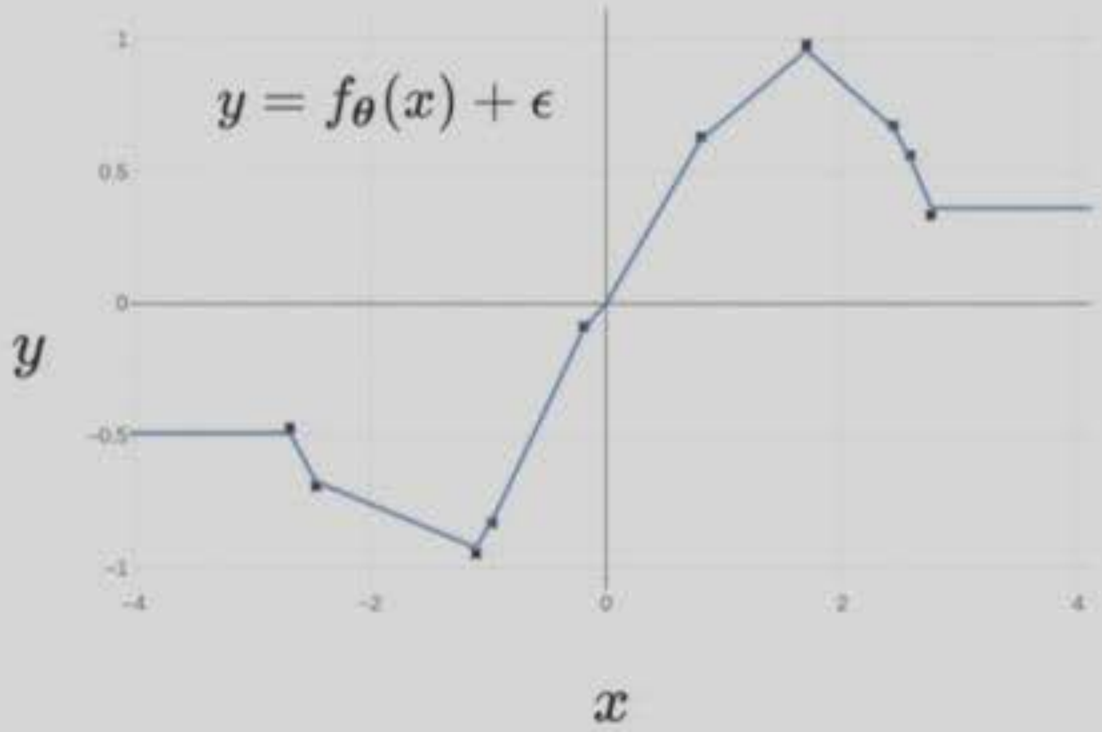
Healthcare
[JAMA Psych, 2020]
[Nature Trans. Psych, 2020]
[AJP, 2019]*

Neural networks (NNs) can fit complicated trends in data

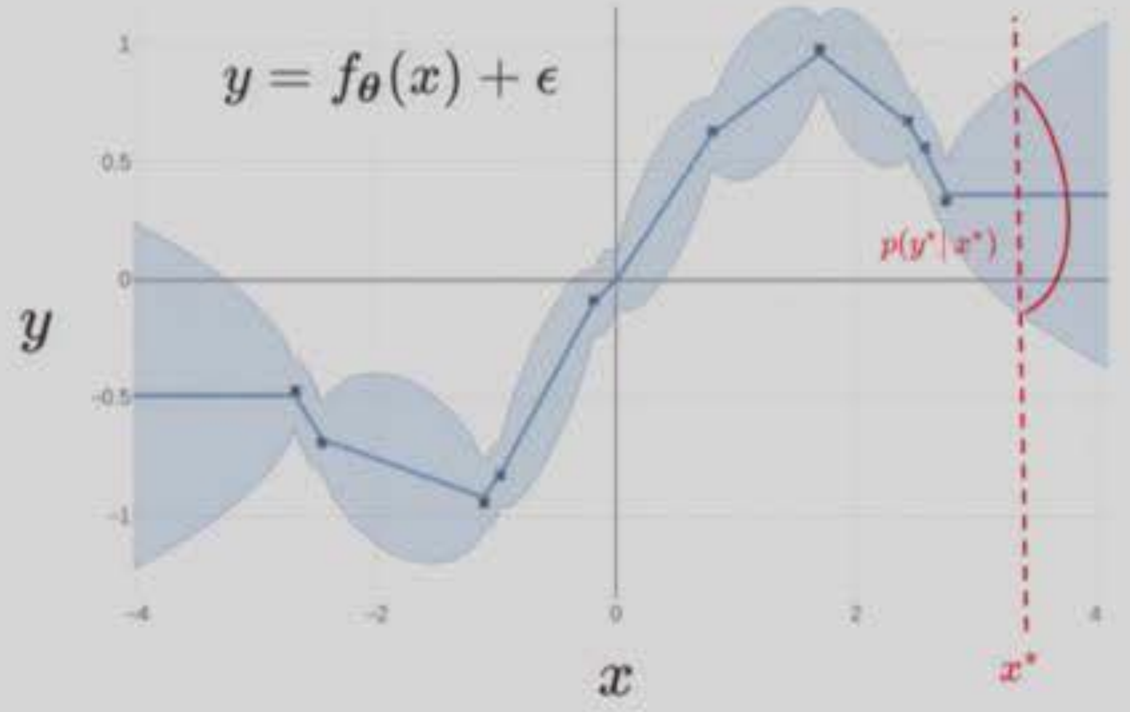


Several success stories...

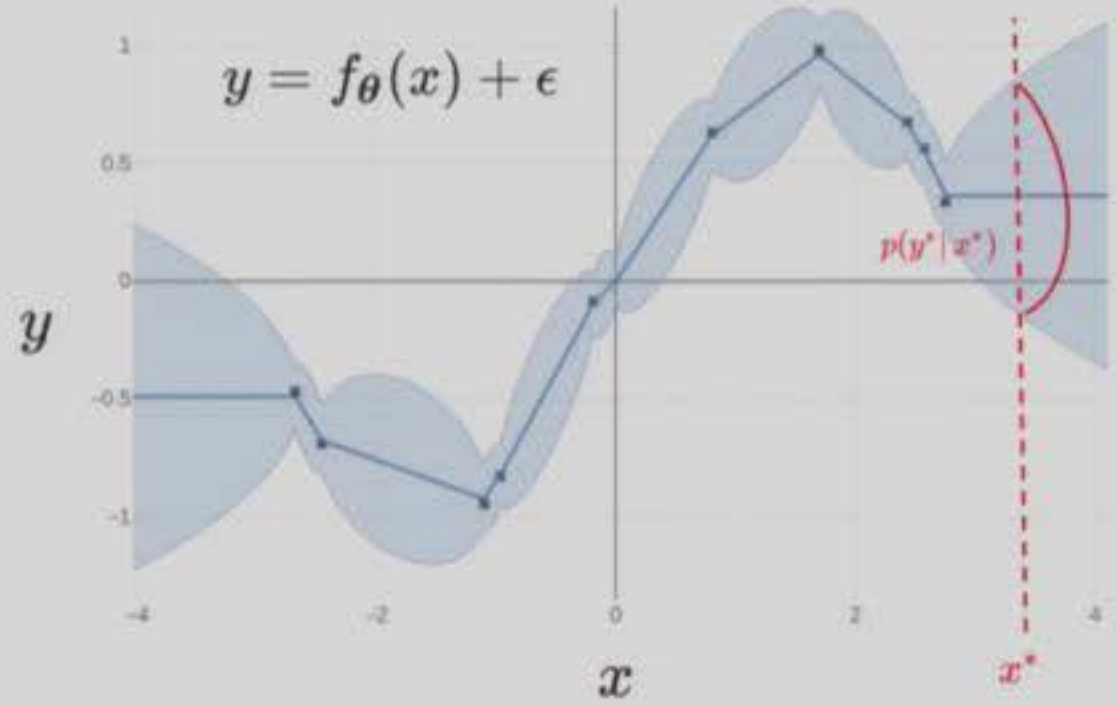
But what if NN is wrong in a high-stake situation?



But what if NN is wrong in a high-stake situation?

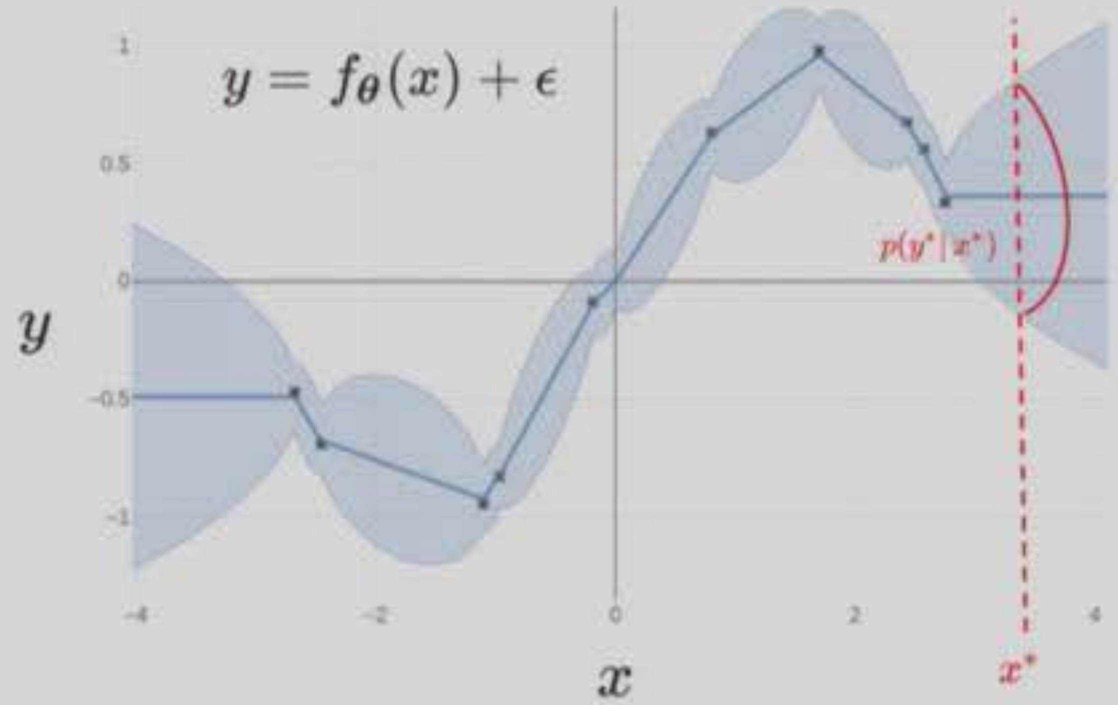


But what if stakes are high?



Uncertainty estimation becomes crucial!

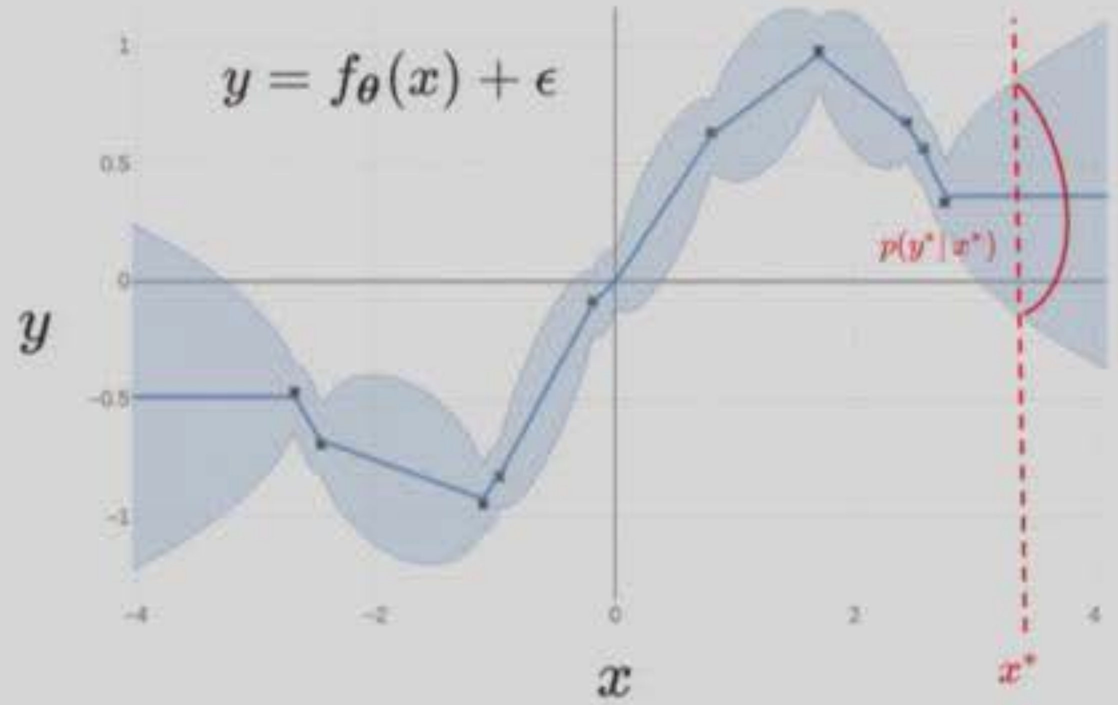
Sometimes we have a priori functional knowledge...



Some examples of assumptions:

- Range of heart rate at rest between 60-100 bpm.
- Slow/fast variation of air pollutant; volatility of stock market

Sometimes we have a priori functional knowledge...



Some examples of assumptions:

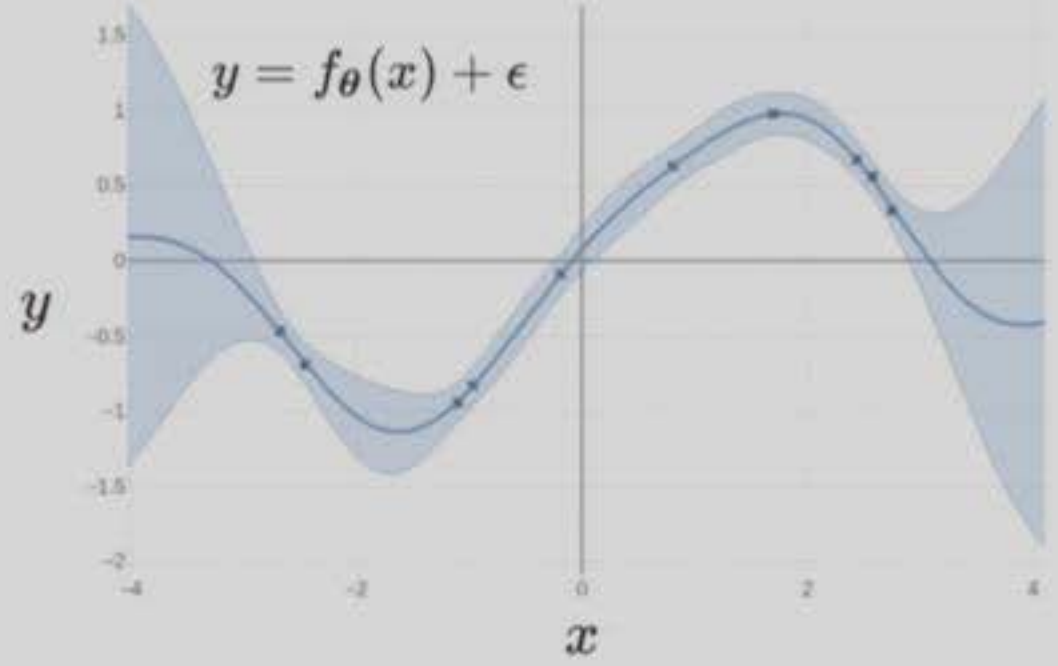
- Range of heart rate at rest between 60-100 bpm.
- Slow/fast variation of air pollutant; volatility of stock market

Key point: All of these translate into restrictions on learned functions.
Question: How can we incorporate such desiderata into the model?

An easy way to specify functional desiderata: Gaussian Processes (GPs)

Definition: *a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.*

$$f \sim \mathrm{GP}\big(\mu(\cdot), k(\cdot,\cdot)\big)$$



Example: RBF kernel as covariance function:

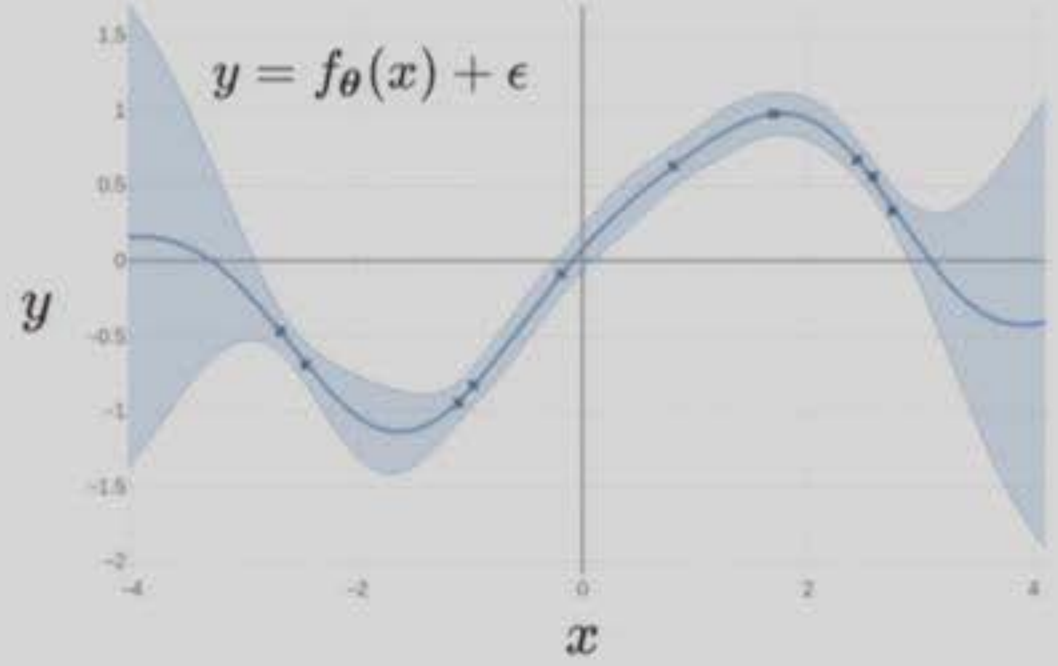
$$k(x,x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\gamma^2}\right)$$

- Stationarity
- Lengthscale
- Amplitude variance

An easy way to specify functional desiderata: Gaussian Processes (GPs)

Definition: *a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.*

$$f \sim \mathrm{GP}\big(\mu(\cdot), k(\cdot,\cdot)\big)$$



Example: RBF kernel as covariance function:

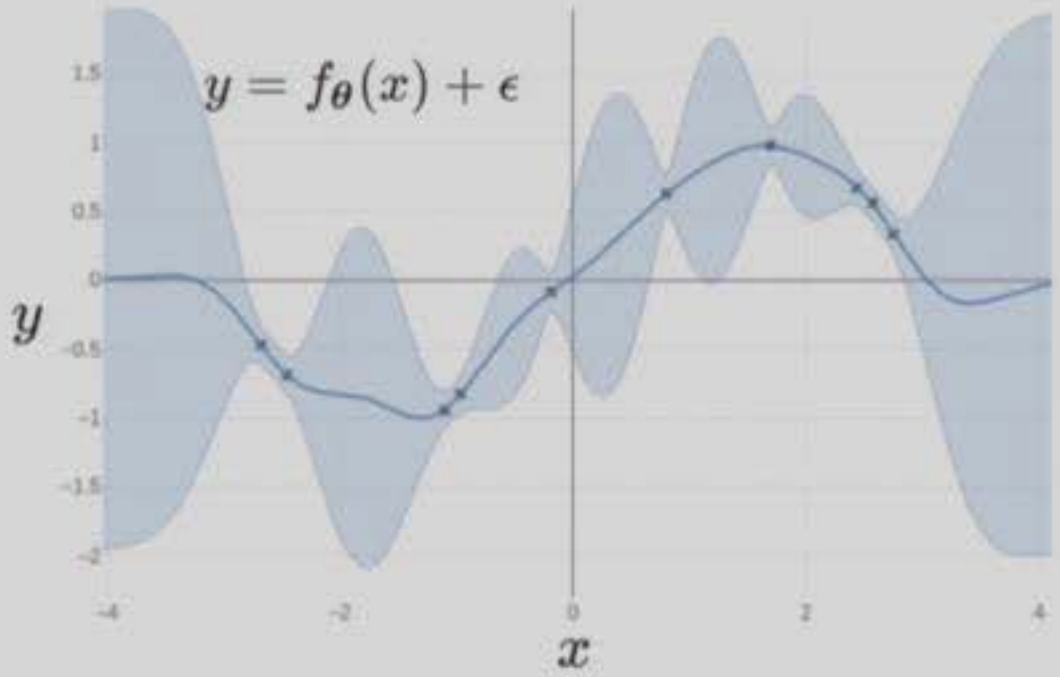
$$k(x,x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\gamma^2}\right)$$

- Stationarity
- Lengthscale
- Amplitude variance

An easy way to specify functional desiderata: Gaussian Processes (GPs)

Definition: *a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.*

$$f \sim \mathrm{GP}\big(\mu(\cdot), k(\cdot,\cdot)\big)$$



Example: RBF kernel as covariance function:

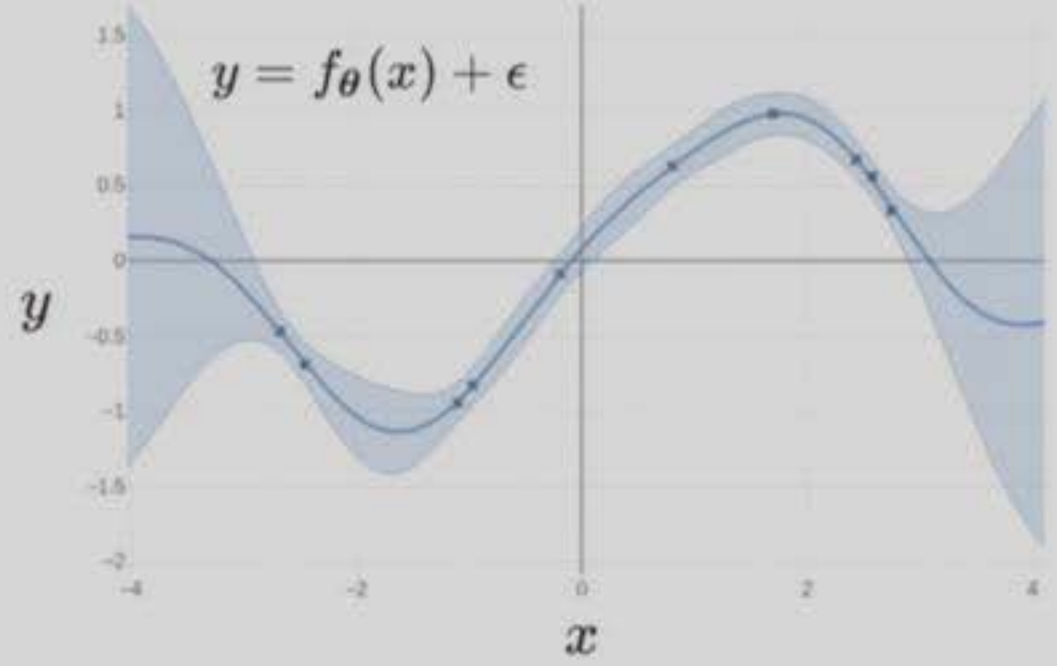
$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\gamma^2}\right)$$

- Stationarity
- Lengthscale
- Amplitude variance

An easy way to specify functional desiderata: Gaussian Processes (GPs)

Definition: *a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.*

$$f \sim \mathrm{GP}\big(\mu(\cdot), k(\cdot,\cdot)\big)$$



Example: RBF kernel as covariance function:

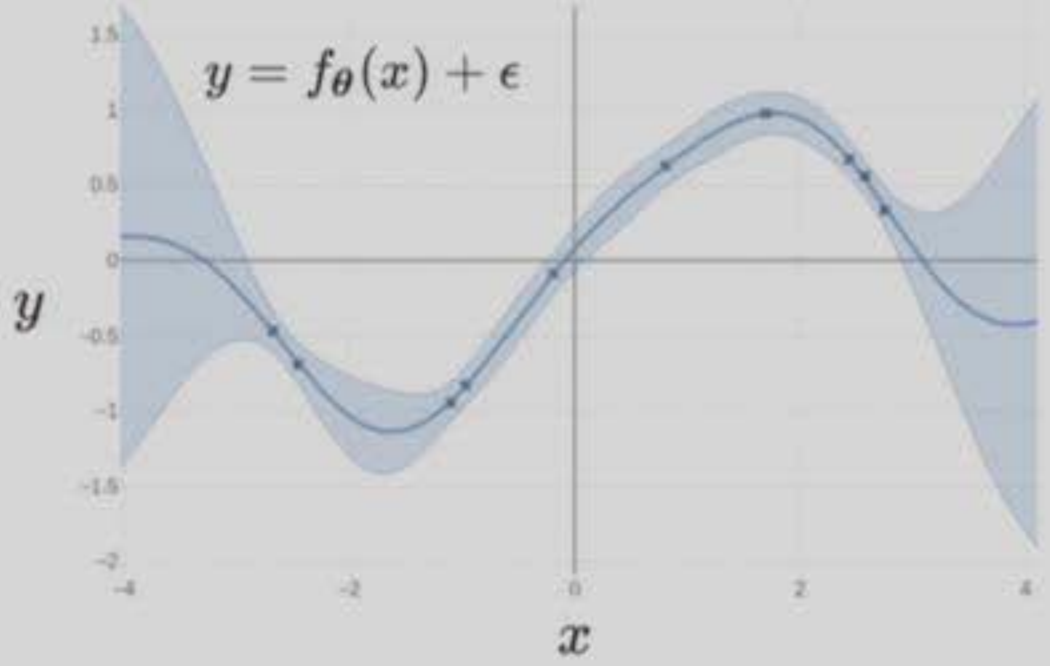
$$k(x,x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\gamma^2}\right)$$

- Stationarity
- Lengthscale
- Amplitude variance

An easy way to specify functional desiderata: Gaussian Processes (GPs)

Definition: *a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.*

$$f \sim \mathrm{GP}\big(\mu(\cdot), k(\cdot,\cdot)\big)$$



Example: RBF kernel as covariance function:

$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\gamma^2}\right)$$

- Stationarity
- Lengthscale
- Amplitude variance

GPs are great, but what if I still want a NN?

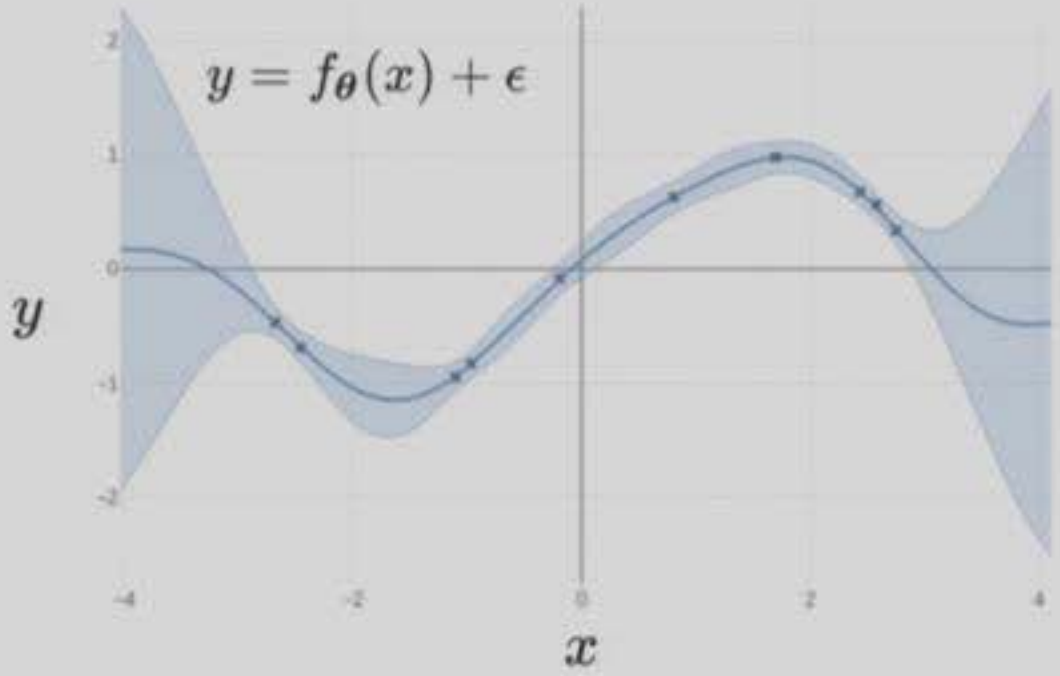
Benefits of NN approaches:

- widely used (many tools available)
- parametric expression
- fast at train and evaluation time

An easy way to specify functional desiderata: Gaussian Processes (GPs)

Definition: *a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.*

$$f \sim \mathrm{GP}\big(\mu(\cdot), k(\cdot,\cdot)\big)$$



Example: RBF kernel as covariance function:

$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\gamma^2}\right)$$

- Stationarity
- Lengthscale
- Amplitude variance

GPs are great, but what if I still want a NN?

Benefits of NN approaches:

- widely used (many tools available)
- parametric expression
- fast at train and evaluation time

GPs are great, but what if I still want a NN?

Benefits of NN approaches:

- widely used (many tools available)
- parametric expression
- fast at train and evaluation time

Key Research Questions:

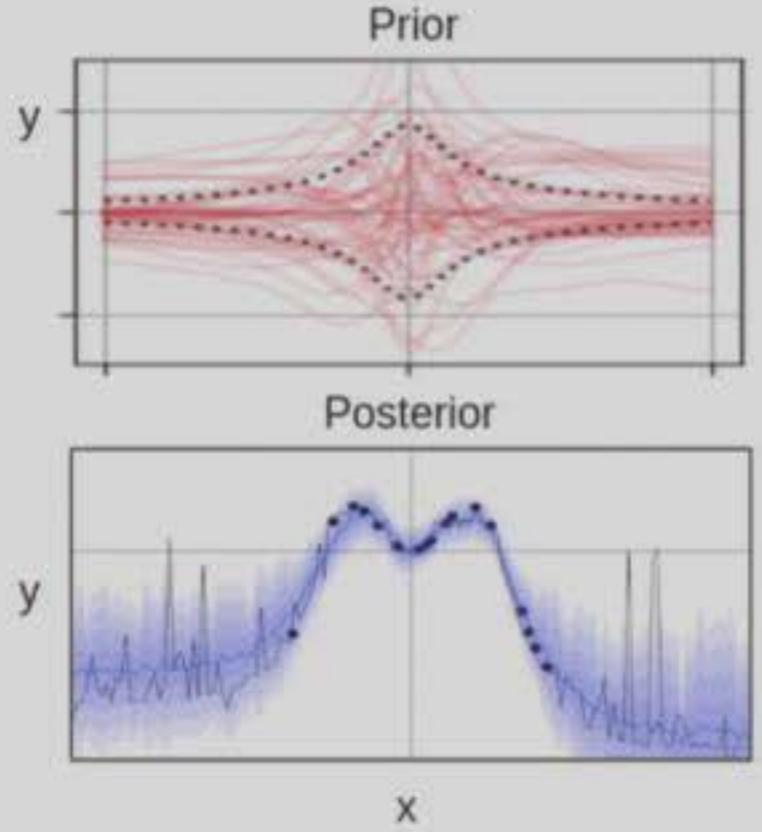
1. Can we design Bayesian NN priors that encode **stationarity properties** like a GP while retaining the benefits of neural networks?
2. Can we easily specify lengthscale and amplitude variance in a **decoupled** fashion?

Background: Bayesian Neural Networks

- Assume prior on network parameters
- Most common, i.i.d Gaussians

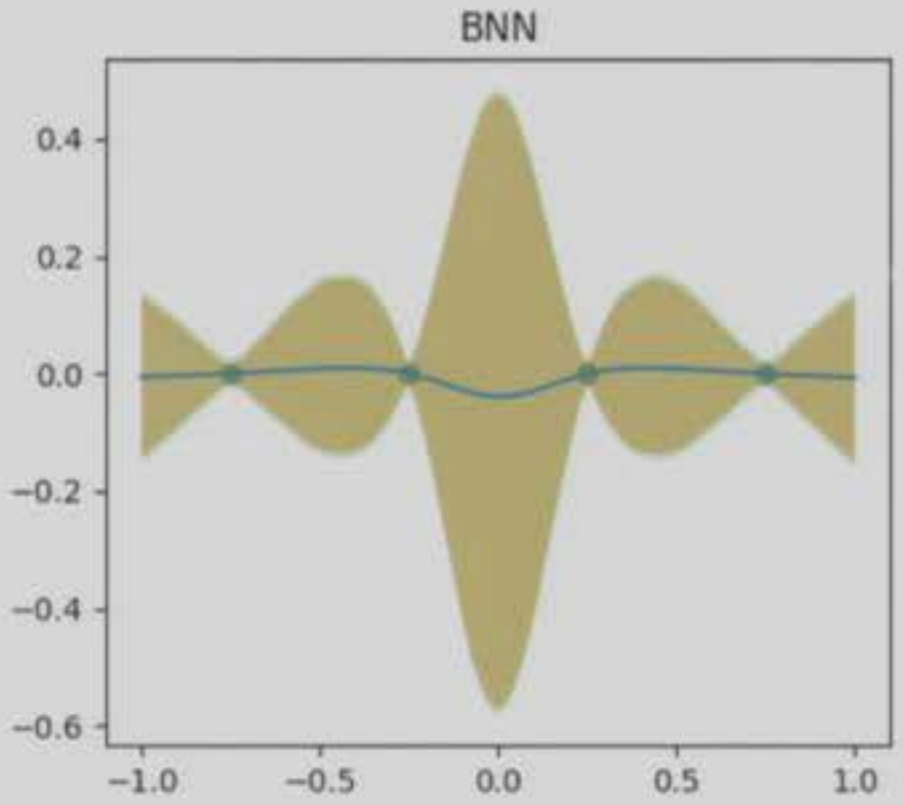
$$\begin{aligned} \mathbf{y} &= f_{\boldsymbol{\theta}}(\mathbf{x}) + \boldsymbol{\epsilon} \\ \epsilon &\sim \mathcal{N}(0, \sigma_y^2 I) \\ \theta_i &\sim \mathcal{N}(0, \sigma_\theta^2 I) \quad \forall i \end{aligned}$$

- $p(\boldsymbol{\theta}) \implies p(f)$



Not only hard to encode functional properties with BNNs; some properties are impossible to get

- For example, a standard BNN (with RBF activations) is nonstationary in amplitude variance (Williams, 1997)

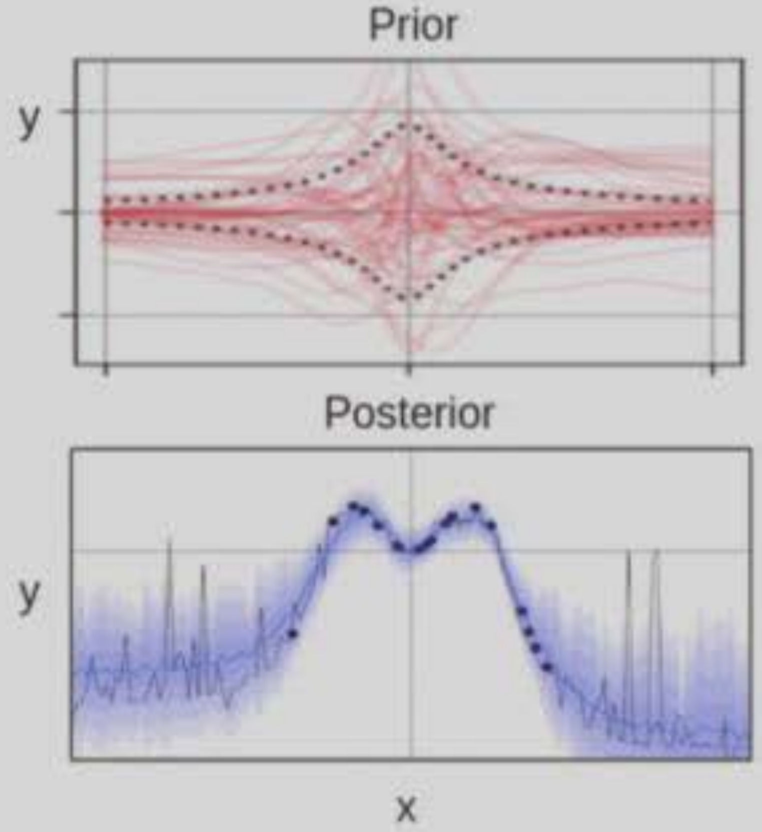


Background: Bayesian Neural Networks

- Assume prior on network parameters
- Most common, i.i.d Gaussians

$$
\begin{aligned}
\mathbf{y} &= f_{\boldsymbol{\theta}}(\mathbf{x}) + \boldsymbol{\epsilon} \\
\epsilon &\sim \mathcal{N}(0, \sigma_y^2 I) \\
\theta_i &\sim \mathcal{N}(0, \sigma_\theta^2 I) \quad \forall i
\end{aligned}
$$

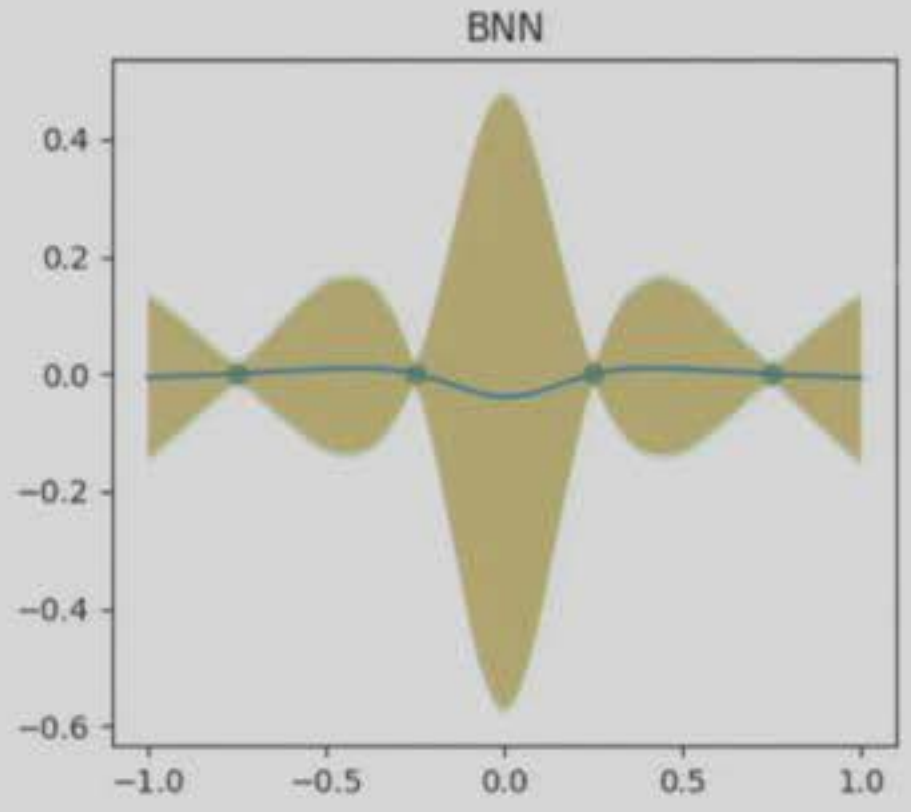
- $p(\boldsymbol{\theta}) \implies p(f)$



- But what does a prior over weights mean in function space?
  - Hard to know!

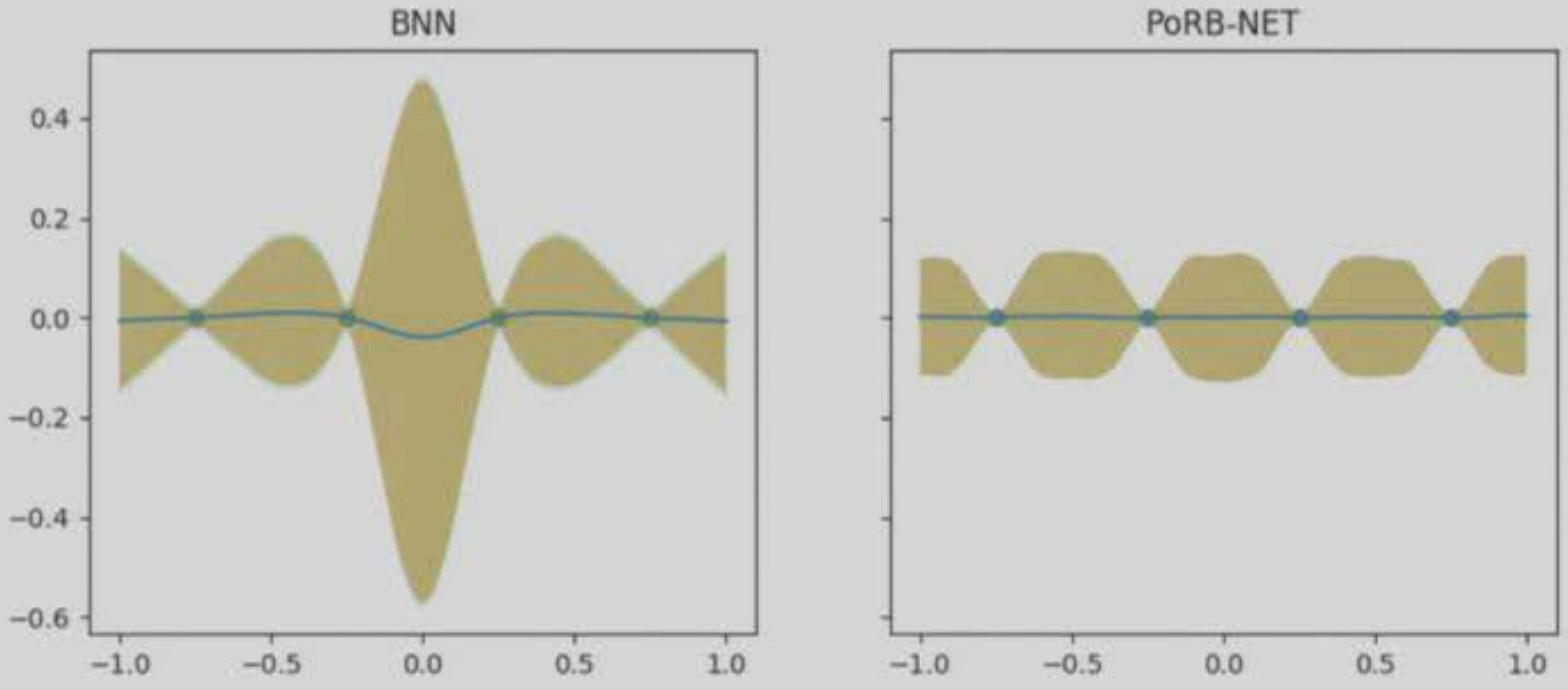
Not only hard to encode functional properties with BNNs; some properties are impossible to get

- For example, a standard BNN (with RBF activations) is nonstationary in amplitude variance (Williams, 1997)



Not only hard to encode functional properties with BNNs; some properties are impossible to get

- For example, a standard BNN (with RBF activations) is nonstationary in amplitude variance (Williams, 1997)



Question: can we design a Bayesian NN that exhibits stationarity? Yes!

Radial Basis Function Networks (RBFNs)

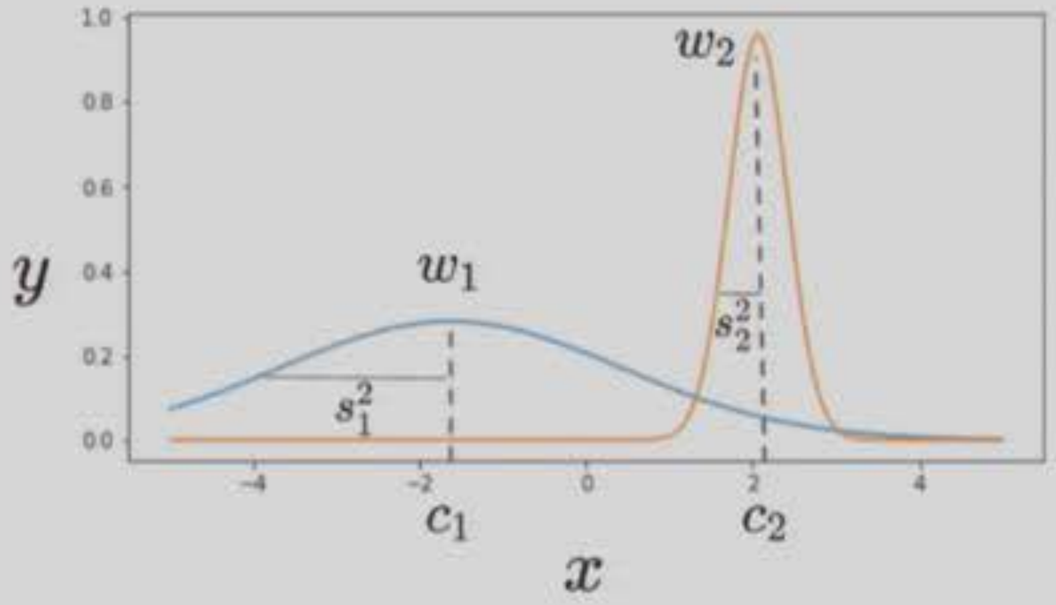
- Around since the 90s (Gyorfi et.al, 2002), recently renewed attention (Taghi et.al, 2004; Zadeh et.al, 2018)

Radial Basis Function Networks (RBFNs)

- Around since the 90s (Gyorfi et.al, 2002), recently renewed attention (Taghi et.al, 2004; Zadeh et.al, 2018)
- NN based on radial basis $\phi(\cdot)$, e.g., $\phi(x) = \exp(-x^2)$

$$f_{\boldsymbol{\theta}}(x) = b + \sum_{k=1}^{K} w_k \phi\big(s_k(x - c_k)\big),$$

- s_k: scale
- c_k: center
- w_k: output weight
- b: output bias

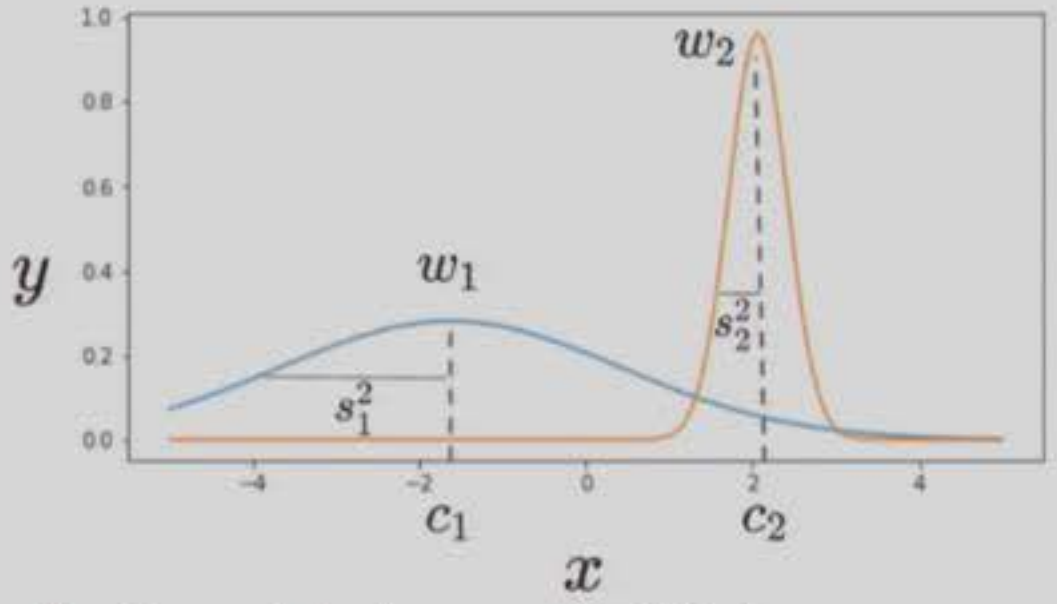


Radial Basis Function Networks (RBFNs)

- Around since the 90s (Gyorfi et.al, 2002), recently renewed attention (Taghi et.al, 2004; Zadeh et.al, 2018)
- NN based on radial basis $\phi(\cdot)$, e.g., $\phi(x) = \exp(-x^2)$

$$f_{\boldsymbol{\theta}}(x) = b + \sum_{k=1}^{K} w_k \phi\big(s_k(x - c_k)\big),$$

- s_k: scale
- c_k: center
- w_k: output weight
- b: output bias



- Equivalence to standard BNN exists, but not interpretable!

Our contribution: Poisson Process Radial Basis Function Network (PoRB-NET) [UAI, 2020]*

1. We propose an expressive prior for NNs

2. We show desirable properties:
   2.1 Stationarity
   2.2 Decoupling of lengthscale and amplitude variance
   2.3 Consistency

3. We demonstrate successful behavior empirically

(*) submitted

Our contribution: Poisson Process Radial Basis Function Network (PoRB-NET) [UAI, 2020]*

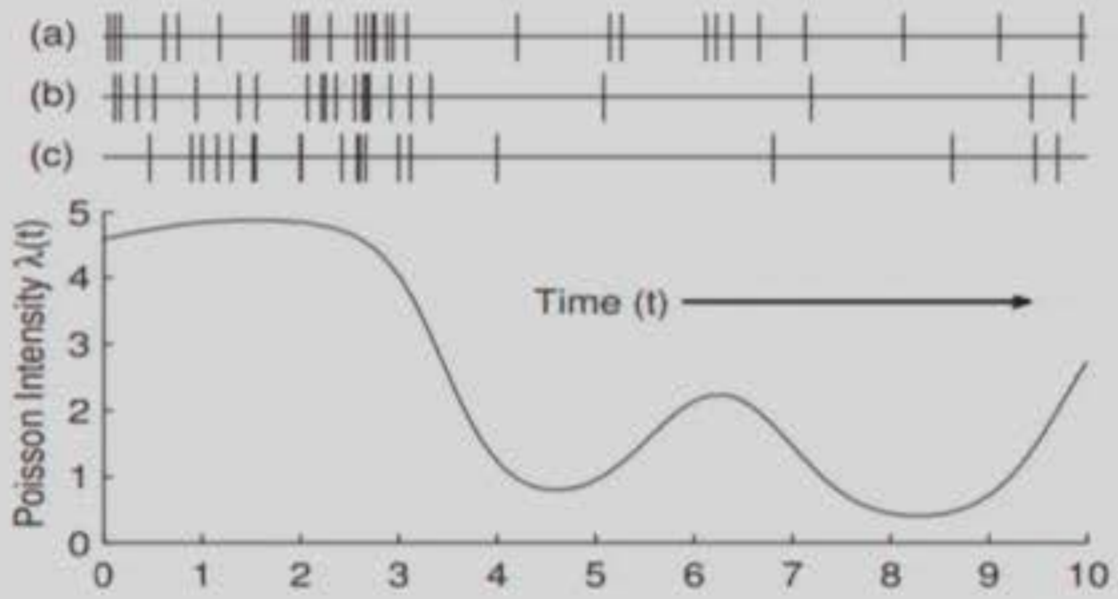
- **Key point**: Expressive prior over centers **c** and scales $\{s_k\}$.

- Poisson process priors over **c**.

$$\mathbf{c} \,|\, \lambda \quad \sim \quad \text{Poisson Process}\,(\lambda)$$

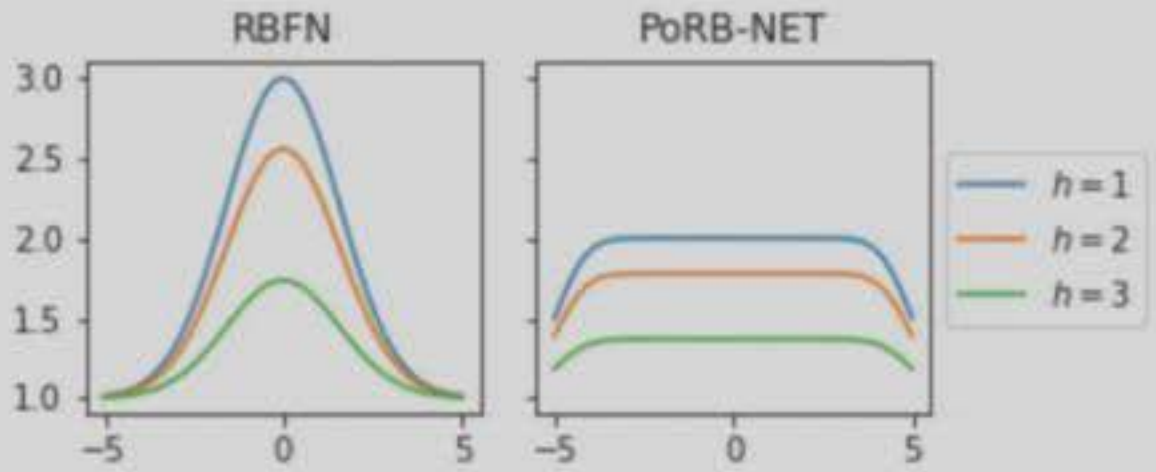
- Deterministic relationship between **c** and $\{s_k\}$.

$$s_k^2 \quad = \quad \lambda^2(c_k)$$



Properties of PoRB-NET

1. **Stationarity**

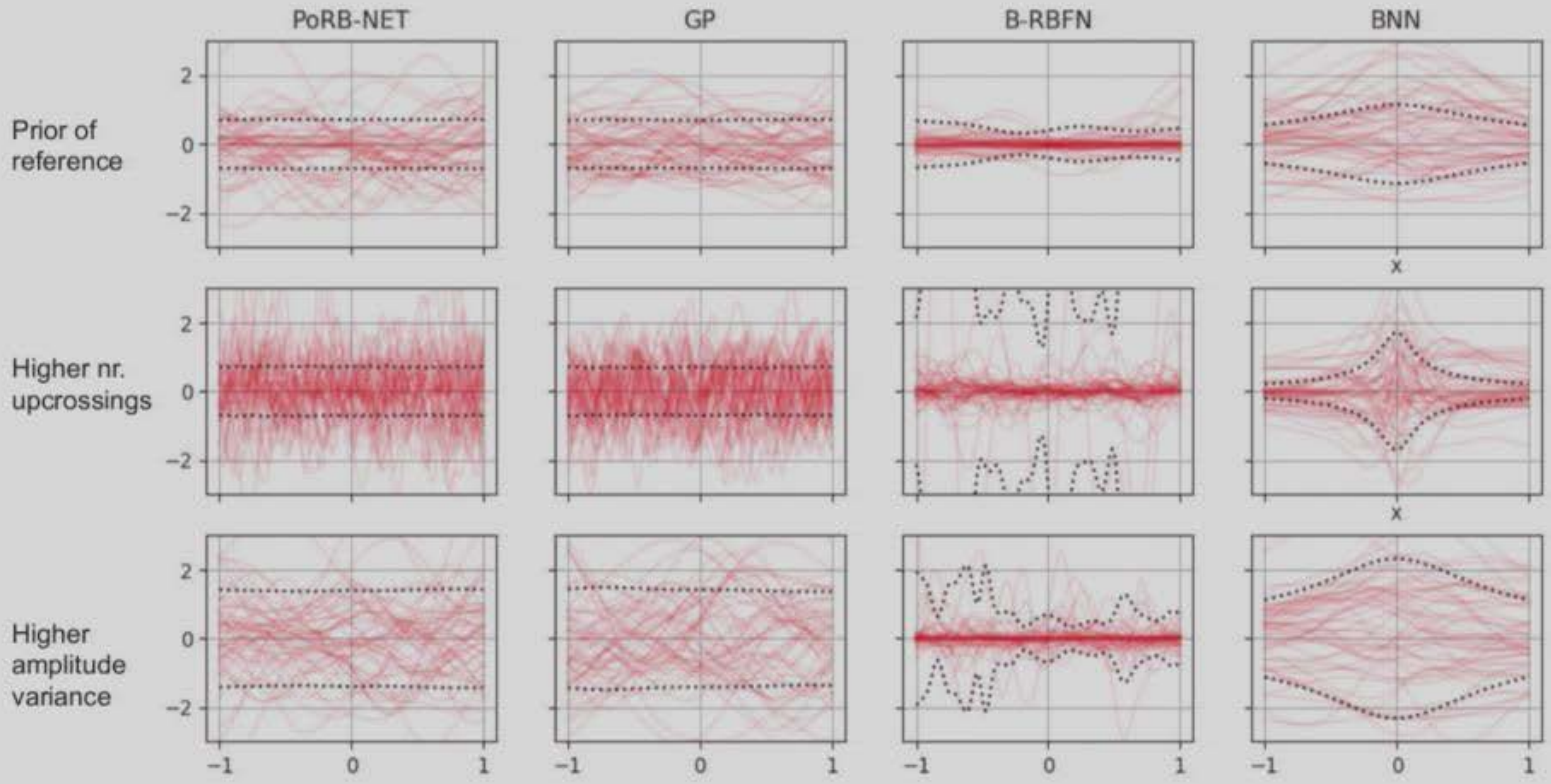


2. **Decoupled lengthscale and amplitude variance**

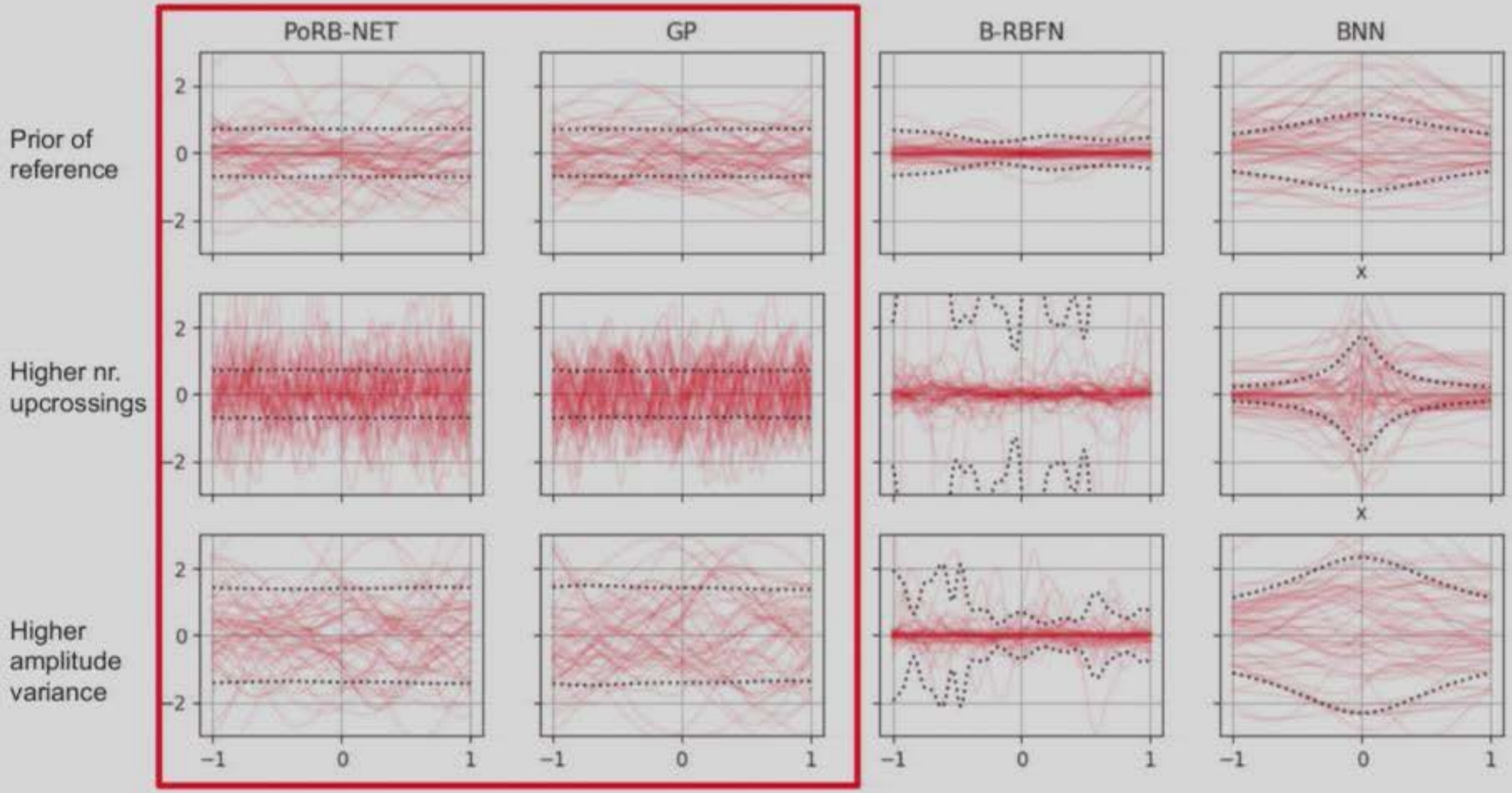
$$\mathrm{Cov}\left(f(x_1), f(x_2)\right) = \sigma_b^2 + \tilde{\sigma}_w^2 \exp\left\{-\lambda^2 \left(\frac{x_1 - x_2}{2}\right)^2\right\}$$

3. **Consistency Theorem:**
A PoRB-NET with uniform intensity function is Hellinger consistent as the number of observations goes to infinity.

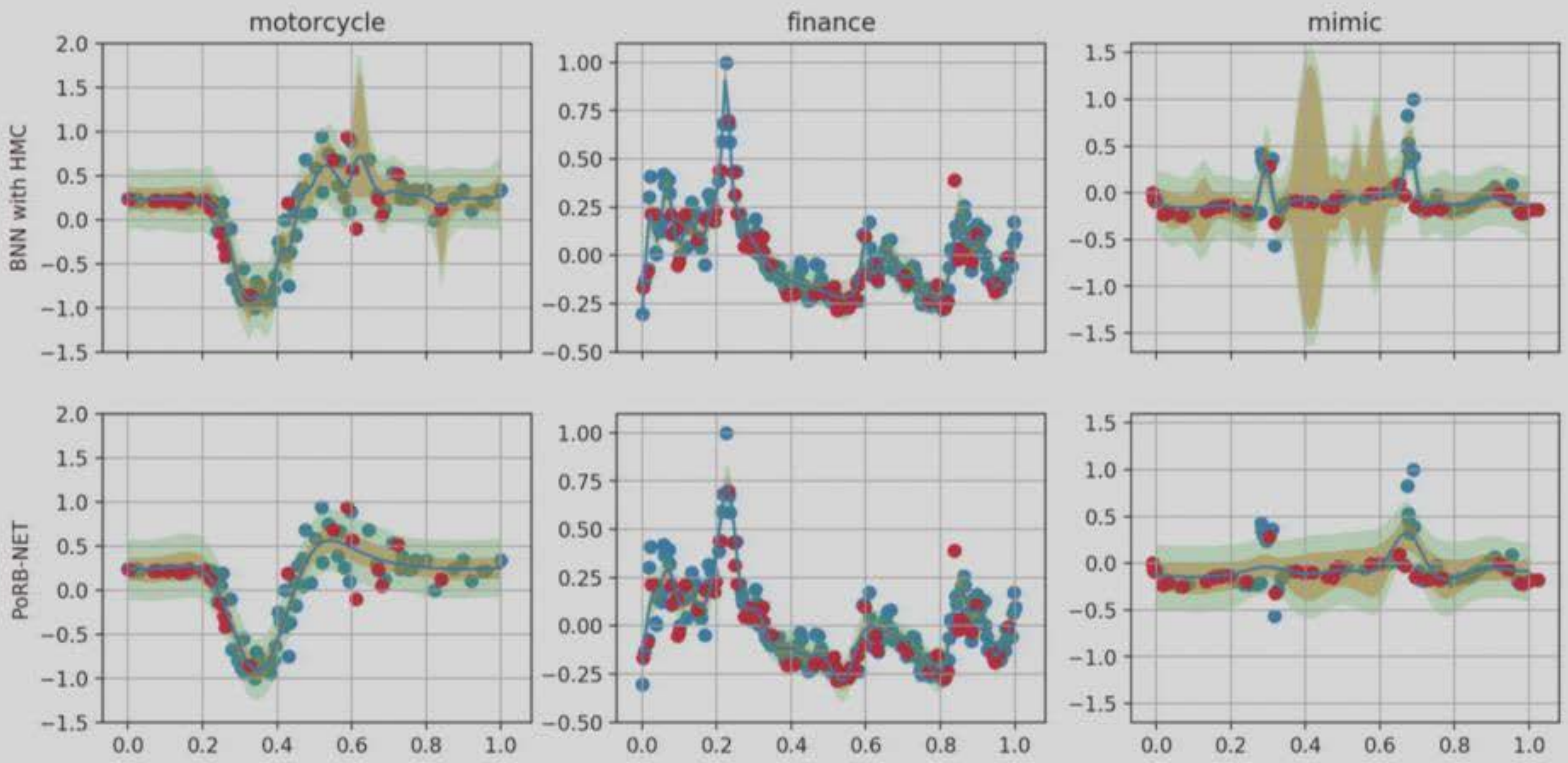
PoRB-NET allows for easy specification of lengthscale and signal variance like a GP



PoRB-NET allows for easy specification of lengthscale and signal variance like a GP



PoRB-NET is able to capture non-stationary patterns in real scenarios, adapting the lengthscale locally



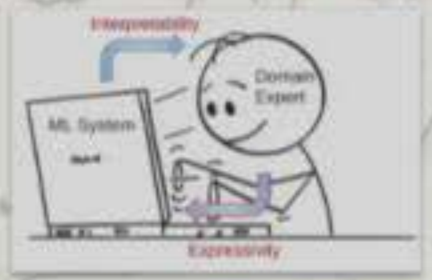
My contributions

Machine Learning
[UAI, 2020]* [JMLR, 2020]
[DBL@NeurIPS, 2019]
[BNP@NeurIPS, 2018]
[AABI, 2019]

Information Theory
[Entropy, 2016]

Sport Sciences
[PlosOne, 2016]

Economics
[PlosOne, 2018]
[ICASSP, 2018]



Cancer Research
[BMC Cancer, 2019]
[Nature Com, 2019]*

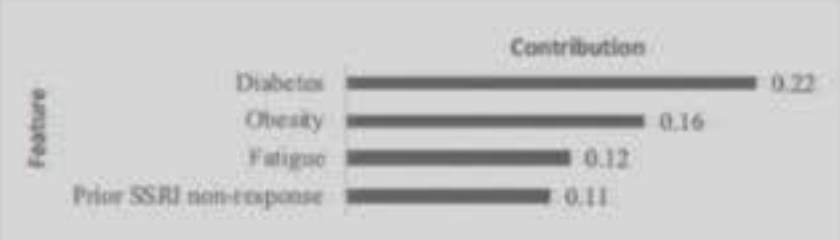
Healthcare
[JAMA Psych, 2020]
[Nature Trans. Psych, 2020]
[AJP, 2019]*

From the lab to the clinic

- Ongoing user study at MGH, Boston
  - Impact of explanations
  - Usefulness, trust...

Why are these therapies being recommended?

The following **patient features** had the highest contributions to system.13's predictions:



Which antidepressant medication would you be most likely to prescribe in this situation?

HARVARD UNIVERSITY

Patient Details:
Jessica is a 37 year old woman who is married and works part time. She presents with 9 months of depressed mood and lack of appetite. She has a seizure disorder, and current medications include Omeprazole and Celecoxib. Prior treatment with Citalopram had no effect on depressed mood.

System.15 Recommendation: **FLUOXETINE**

Top 5 therapies with highest probability for stability:

Therapy	Predicted Stability*	Predicted Dropout Risk**
Fluoxetine	.76	.05
Sertraline	.67	.05
Paroxetine	.64	.10
Venlafaxine	.60	.14
Vortioxetine	.55	.15

*Stability: continued use of the same medication for at least 3 months
**Dropout: early treatment discontinuation following prescription

Why are these therapies being recommended?

The following **rules** had the highest contributions to system.15's predictions:

1. *If underweight or lack of appetite, favor weight gain, favor Mirtazapine*
2. *If underweight or lack of appetite, avoid appetite suppressants, avoid nausea-inducing, avoid SNRI's, avoid Sertraline*
3. *If lack of response to Paroxetine, avoid SSRI's*

Current and future research agenda

Contact: *melanie@seas.harvard.edu*
https://melaniefp.github.io/

Impactful real-world problems:

- Personalize prescription of antidepressants
- Prognosticate outcomes for in-vitro fertilization
- ...

Useful machine learning methodology:

- How to better quantify model uncertainty?
- How to combine expert knowledge with data-driven evidence?
- How to learn task-meaningful representations?

Acknowledgements

Special thanks to:

- Beau Coker
- Finale Doshi-Velez
- All members of DTAK!

- Oscar Puig
- Francesca Milletti
- Fernando Perez-Cruz

- Isabel Valera
- Maria Lomeli
- Zoubin Ghahramani

CRCS Center for Research on Computation and Society
at Harvard John A. Paulson School of Engineering and Applied Sciences

HDSI | Harvard Data Science Initiative

Universidad Carlos III de Madrid

Roche

DTAK

Thank you for listening!

My contributions

Information Theory
[Entropy, 2016]

Sport Sciences
[PlosOne, 2016]

Economics
[PlosOne, 2018]
[ICASSP, 2018]

Machine Learning
[UAI, 2020]*
[JMLR, 2020]
[DBL@NeurIPS, 2019]
[BNP@NeurIPS, 2018]
[AABI, 2019]

Cancer Research
[BMC Cancer, 2019]
[Nature Com, 2019]*

Healthcare
[JAMA Psych, 2020]
[Nature Trans. Psych, 2020]
[AJP, 2019]*