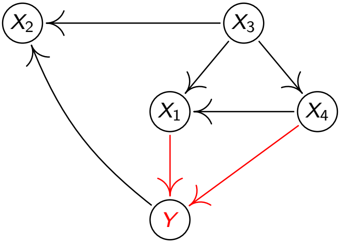
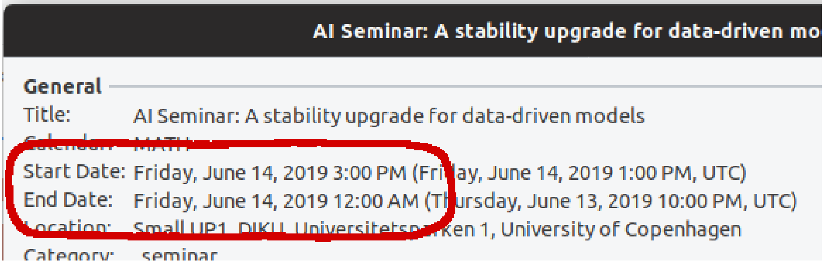
A stability upgrade for data-driven models



Jonas Peters
University of Copenhagen
AI Seminar, 14th June 2019


AI Seminar: A stability upgrade for data-driven mo
General
Title: AI Seminar: A stability upgrade for data-driven models
Calendar: MATH
Start Date: Friday, June 14, 2019 3:00 PM (Friday, June 14, 2019 1:00 PM, UTC)
End Date: Friday, June 14, 2019 12:00 AM (Thursday, June 13, 2019 10:00 PM, UTC)
Location: Small UP1, DIKU, Universitetsparken 1, University of Copenhagen
Category: seminar

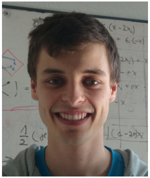
Stefan Bauer

Peter Bühlmann

Rune Christiansen

Martin Jakobsen

Nicolai Meinshausen



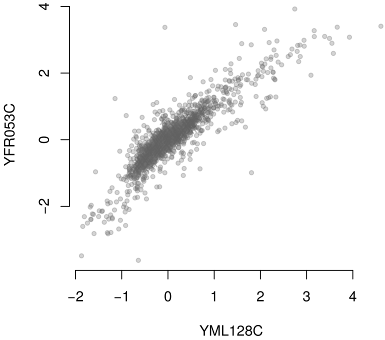
Niklas Pfister

Dominik Rothenhäusler

Mads Thøisen

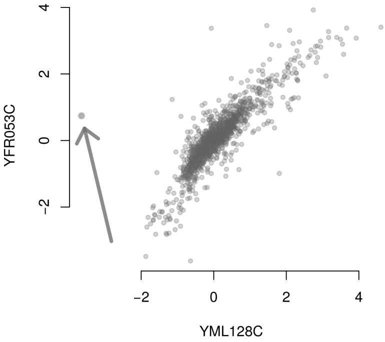
Given yeast data...

Given yeast data... this is the predictor with the strongest correlation:



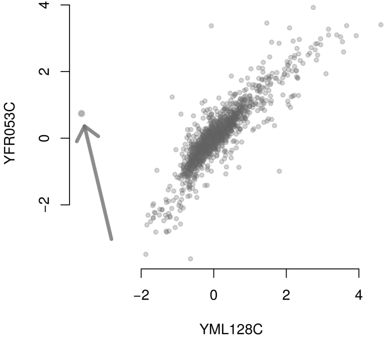
data from: Kemmeren et al. 2014

After an intervention on that predictor:



data from: Kemmeren et al. 2014

After an intervention on that predictor: (it's a non-causal predictor!)

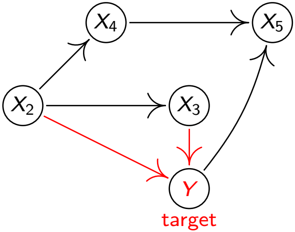


data from: Kemmeren et al. 2014

The Problem of Causal Discovery:

observed data $\xrightarrow{?}$ causal model

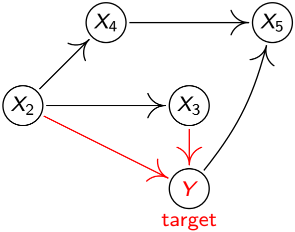
Y	X_2	X_3	X_4	X_5
3.4	−0.3	5.8	−2.1	2.2
1.7	−0.2	7.0	−1.2	0.4
−2.4	−0.1	4.3	−0.7	3.5
2.3	−0.3	5.5	−1.1	−4.4
3.5	−0.2	3.9	−0.9	−3.9
⋮	⋮	⋮	⋮	⋮



The Problem of Causal Discovery:

observed data $\xrightarrow{?}$ causal model

Y	X_2	X_3	X_4	X_5
3.4	−0.3	5.8	−2.1	2.2
1.7	−0.2	7.0	−1.2	0.4
−2.4	−0.1	4.3	−0.7	3.5
2.3	−0.3	5.5	−1.1	−4.4
3.5	−0.2	3.9	−0.9	−3.9
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

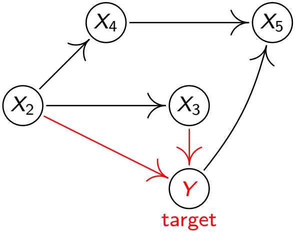


Here: Find the direct causes of Y!

The Problem of Causal Discovery:

observed data $\xrightarrow{?}$ causal model

Y	X_2	X_3	X_4	X_5
3.4	−0.3	5.8	−2.1	2.2
1.7	−0.2	7.0	−1.2	0.4
−2.4	−0.1	4.3	−0.7	3.5
2.3	−0.3	5.5	−1.1	−4.4
3.5	−0.2	3.9	−0.9	−3.9
⋮	⋮	⋮	⋮	⋮



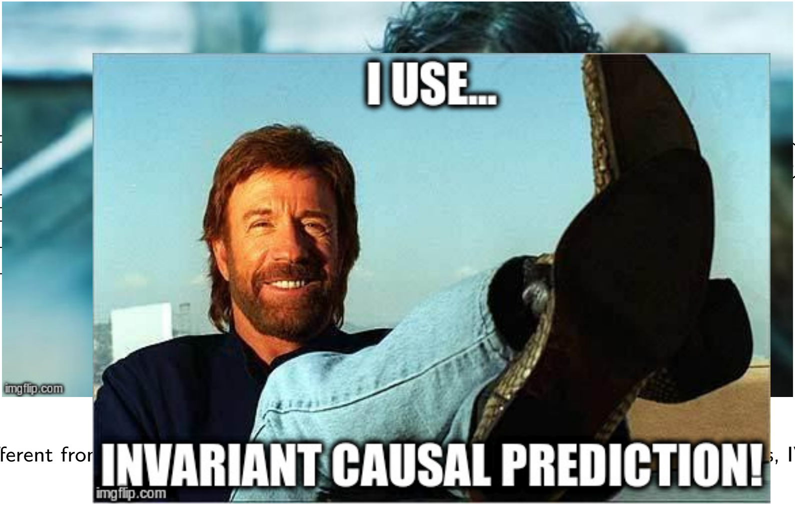
Here: Find the direct causes of Y!

(Different from: variable adjustment, propens. score matching, double robust methods, IVs, ...)

The Problem of Causal Discovery:

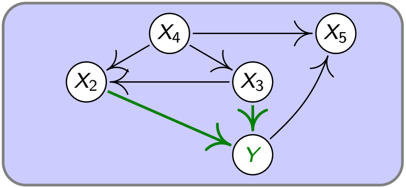
(Different from: variable adjustment, propens. score matching, double robust methods, IVs, ...)

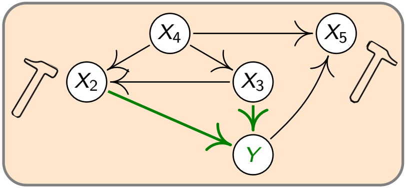
The Problem of Causal Discovery:



(Different fror … s, IVs, ...)

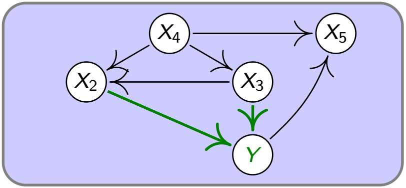
Fundamental assumption: $X_1, X_4 \to Y$ is invariant under interventions.

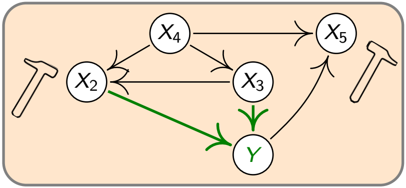




cf. modularity, autonomy, Haavelmo 1944, Aldrich 1989, Pearl 2009, ...

Fundamental assumption: $X_1, X_4 \to Y$ is invariant under interventions.





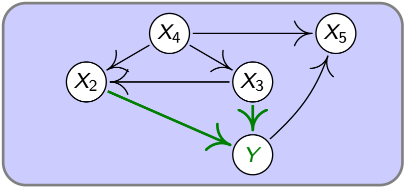
cf. modularity, autonomy, Haavelmo 1944, Aldrich 1989, Pearl 2009, ...

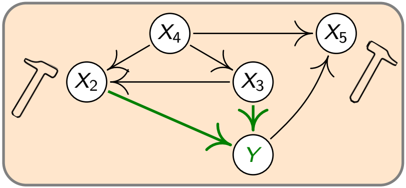
Key idea: Use and data and search for invariant models.

set	$\emptyset$	$\{2\}$	$\{3\}$	$\{4\}$	$\cdots$	$\{2,3\}$	$\{2,4\}$	$\cdots$	$\{2,3,4\}$
invariance	✗	✗	✗	✗	$\cdots$	✓	✗	$\cdots$	✓

$$\hat{S} := \bigcap_{S \text{ invariant}} S = \{1, 4\}$$

Fundamental assumption: $X_1, X_4 \to Y$ is invariant under interventions.





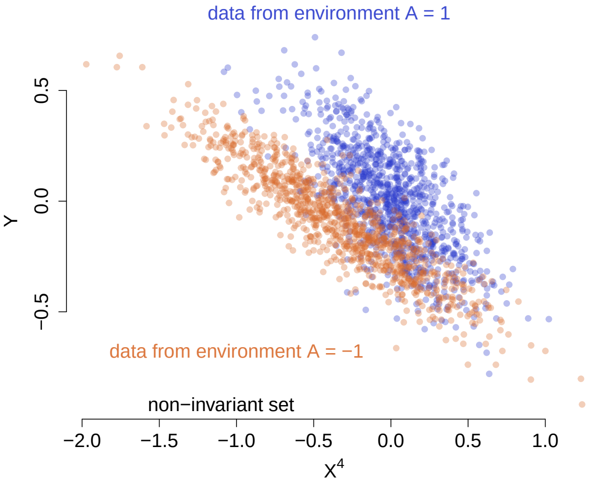
cf. modularity, autonomy, Haavelmo 1944, Aldrich 1989, Pearl 2009, ...

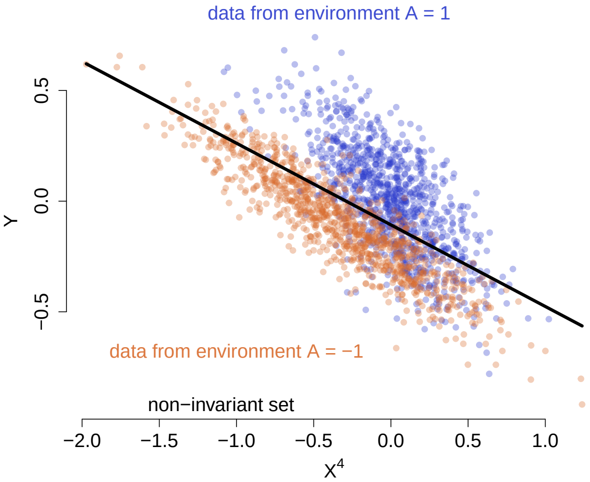
Key idea: Use and data and search for invariant models.

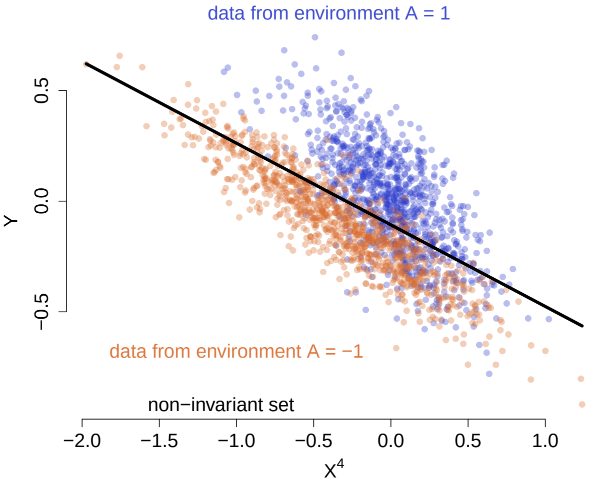
set	$\emptyset$	$\{2\}$	$\{3\}$	$\{4\}$	$\cdots$	$\{2,3\}$	$\{2,4\}$	$\cdots$	$\{2,3,4\}$
invariance	✗	✗	✗	✗	$\cdots$	✓	✗	$\cdots$	✓

$$\hat{S} := \bigcap_{S \text{ invariant}} S = \{1, 4\}$$

JP, Bühlmann, Meinshausen, JRSS-B 2016 (with discussion): $P(\hat{S} \subseteq S^*) \geq 1 - \alpha$.


data from environment A = 1
data from environment A = −1
non−invariant set
Y
X4
0.5
0.0
−0.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0


data from environment A = 1
data from environment A = −1
non−invariant set
Y
0.5
0.0
−0.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
X4

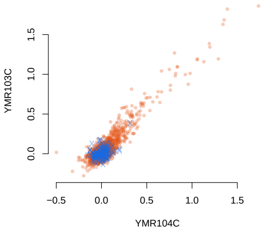


A stands for anchor.

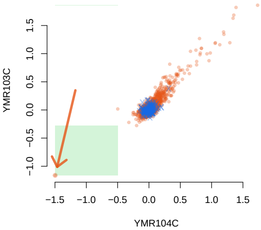
$$\mathbf{E}A(Y - X^4 b) \neq 0$$

Thus: X^4 is a non-invariant model.

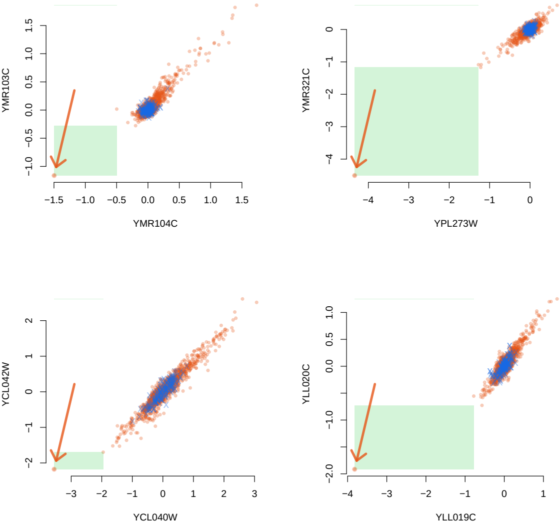
Predictors that are inferred to be causal by ICP...



...and the corresponding interventions on the predictor



…and the corresponding interventions on the predictor



Kemmeren et al. 2014

Discrete environments (gene data):

JP, Meinshausen, Bühlmann: *Causal inference using invariant prediction: identification and confidence intervals*, JRSSB 2016

Discrete environments (gene data):

JP, Meinshausen, Bühlmann: *Causal inference using invariant prediction: identification and confidence intervals*, JRSSB 2016

No environments (finance data): = time

Pfister, Bühlmann, JP: *Invariant causal pred. for seq. data*, JASA 2018

Discrete environments (gene data):

JP, Meinshausen, Bühlmann: *Causal inference using invariant prediction: identification and confidence intervals*, JRSSB 2016

No environments (finance data): ⚓ = time

Pfister, Bühlmann, JP: *Invariant causal pred. for seq. data*, JASA 2018

Nonlinear relations (fertility data):

Heinze-Deml, JP, Meinshausen: *Invariant Causal Prediction for Nonlinear Models*, Journal of Causal Inference 2018

Discrete environments (gene data):

JP, Meinshausen, Bühlmann: *Causal inference using invariant prediction: identification and confidence intervals*, JRSSB 2016

No environments (finance data): = time

Pfister, Bühlmann, JP: *Invariant causal pred. for seq. data*, JASA 2018

Nonlinear relations (fertility data):

Heinze-Deml, JP, Meinshausen: *Invariant Causal Prediction for Nonlinear Models*, Journal of Causal Inference 2018

Discrete hidden variables (Earth system data):

Christiansen, JP: *Invariance-based Causal Discovery in the Presence of Discrete Hidden Variables*, arXiv:1808.05541

Discrete environments (gene data):

JP, ... 6

No...

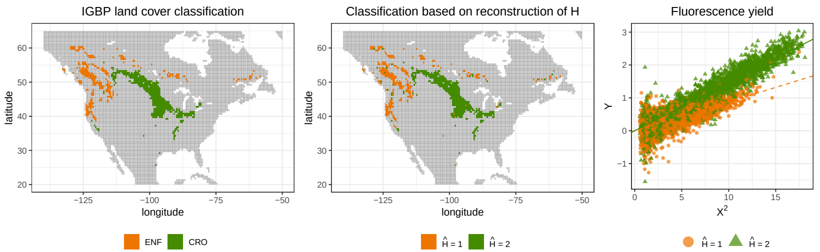
Pfis...

No...

Hei...

Di...

Chr...



Discrete environments (gene data):

JP, Meinshausen, Bühlmann: *Causal inference using invariant prediction: identification and confidence intervals*, JRSSB 2016

No environments (finance data): ⚓ = time

Pfister, Bühlmann, JP: *Invariant causal pred. for seq. data*, JASA 2018

Nonlinear relations (fertility data):

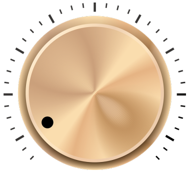
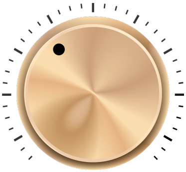
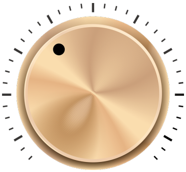
Heinze-Deml, JP, Meinshausen: *Invariant Causal Prediction for Nonlinear Models*, Journal of Causal Inference 2018

Discrete hidden variables (Earth system data):

Christiansen, JP: *Invariance-based Causal Discovery in the Presence of Discrete Hidden Variables*, arXiv:1808.05541

Survival data (registry data):

Laksafoss, JP: *Causal Methods for Survival Analysis*, work in progress

invariance  prediction invariance  prediction invariance  prediction invariance  prediction

Find a trade-off between

- invariance with respect to

AND

- predictive power

Idea 1: Anchor regression

$Y \in \mathbf{R}^1$: target
$X \in \mathbf{R}^{1\times d}$: predictors
$A \in \mathbf{R}^{1\times q}$: anchors, $\mathbf{E}A^t A = Id$

$$b^\gamma := \operatorname*{argmin}_b \underbrace{\mathbf{E}(Y - Xb)^2}_{\text{prediction}} + \gamma \underbrace{\|\mathbf{E}A^t(Y - Xb)\|_2^2}_{\text{invariance}}$$

Idea 1: Anchor regression

$Y \in \mathbf{R}^1$: target
$X \in \mathbf{R}^{1\times d}$: predictors
$A \in \mathbf{R}^{1\times q}$: anchors, $\mathbf{E}A^t A = Id$

$$b^\gamma := \operatorname*{argmin}_b \underbrace{\mathbf{E}(Y - Xb)^2}_{\text{prediction}} + \gamma \underbrace{\|\mathbf{E}A^t(Y - Xb)\|_2^2}_{\text{invariance}}$$

$\gamma \to 0$: OLS
$\gamma \to \infty$: IV solution (if identifiable)
$\gamma \to \infty$: best invariant predictor (if not identifiable)

Idea 1: Anchor regression

$Y \in \mathbf{R}^1$: target
$X \in \mathbf{R}^{1\times d}$: predictors
$A \in \mathbf{R}^{1\times q}$: anchors, $\mathbf{E}A^tA = Id$

$$b^\gamma := \underset{b}{\operatorname{argmin}} \; \underbrace{\mathbf{E}(Y - Xb)^2}_{\text{prediction}} + \gamma \, \underbrace{\|\mathbf{E}A^t(Y - Xb)\|_2^2}_{\text{invariance}}$$

$\gamma \to 0$:	OLS
$\gamma \to \infty$:	IV solution (if identifiable)
$\gamma \to \infty$:	best invariant predictor (if not identifiable)

- Anchor regression minimizes worst case prediction error under shift interventions.

Rothenhäusler, Bühlmann, Meinshausen, JP (arXiv:1801.06229)

Idea 1: Anchor regression

$Y \in \mathbf{R}^1$: target
$X \in \mathbf{R}^{1\times d}$: predictors
$A \in \mathbf{R}^{1\times q}$: anchors, $\mathbf{E}A^tA = Id$

$$b^\gamma := \underset{b}{\operatorname{argmin}} \ \underbrace{\mathbf{E}(Y - Xb)^2}_{\text{prediction}} + \gamma \underbrace{\|\mathbf{E}A^t(Y - Xb)\|_2^2}_{\text{invariance}}$$

$\gamma \to 0$:	OLS
$\gamma \to \infty$:	IV solution (if identifiable)
$\gamma \to \infty$:	best invariant predictor (if not identifiable)

- Anchor regression minimizes worst case prediction error under shift interventions. Rothenhäusler, Bühlmann, Meinshausen, JP (arXiv:1801.06229)
- The finite sample estimator is known as a k-class estimator for IV solution. Theil (1958), Nagar (1959) – joint work with Martin Emil Jakobsen

Idea 2:

N. Pfister, S. Bauer, JP:

Learning stable structures in kinetic systems: benefits of a causal approach

arXiv:1810.11776

Example: Maillard reaction

Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

Example: Maillard reaction Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$
$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \ldots$$

Example: Maillard reaction

Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

http://www.srfcdn.ch/radio/modules/dynimages/624/srf-1/2015/01/diverses/264377.150114_raclette_key.jpg

Example: Maillard reaction

Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$

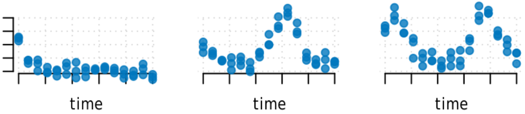
$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \dots$$

Example: Maillard reaction

Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$

$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \ldots$$



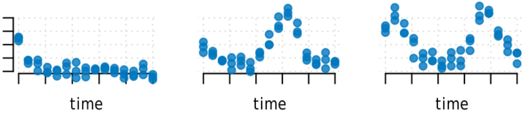
Example: Maillard reaction

Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$

$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \ldots$$

If the network structure is unknown, can we recover it from data?



Example: Maillard reaction

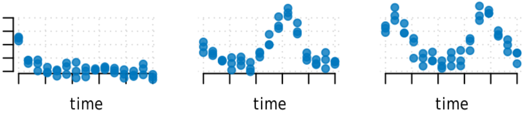
Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$

$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \ldots$$

If the network structure is unknown, can we recover it from data?

WANTED: system of ODEs: $\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X_0} = \mathbf{x_0}.$

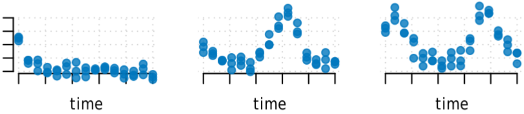


Example: Maillard reaction Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$
$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \ldots$$

If the network structure is unknown, can we recover it from data?

WANTED: system of ODEs: $\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X_0} = \mathbf{x_0}$.
adding noise: $\tilde{\mathbf{X}}_t := \mathbf{X}_t + \varepsilon_t$.



Example: Maillard reaction Glu, Mel, C5, ForAc, Triose, Cn, AcAc, Amad, lysR, Fru, AMP

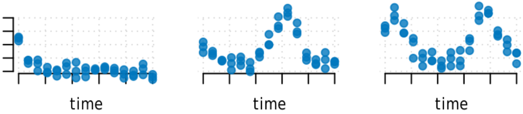
$$\frac{d}{dt}[\mathrm{Glu}]_t = -\theta_1[\mathrm{Glu}]_t + \theta_2[\mathrm{Fru}]_t - \theta_3[\mathrm{lysR}]_t[\mathrm{Glu}]_t$$

$$\frac{d}{dt}[\mathrm{Mel}]_t = \theta_4[\mathrm{AMP}]_t \qquad \ldots$$

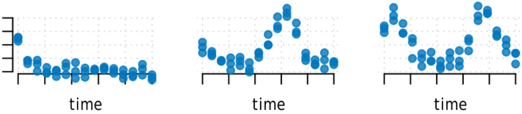
If the network structure is unknown, can we recover it from data?

WANTED: system of ODEs: $\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X_0} = \mathbf{x_0}$.
adding noise: $\tilde{\mathbf{X}}_t := \mathbf{X}_t + \varepsilon_t$.

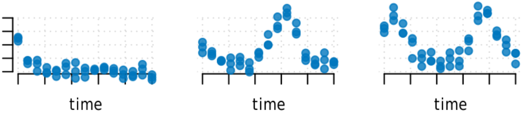
GIVEN: discrete time measurements (w/ repetitions): $\tilde{\mathbf{X}}_{t_1}, \ldots, \tilde{\mathbf{X}}_{t_m}$.



Brands et al. (2002): "Kinetic model. of reactions in heated monosaccharide-casein systems.", J. of agr. and food chemistry.



King et al, Nature 2004, Bongard et al, PNAS 2007, Calderhead et al, NIPS 2009, Schmidt et al, Science 2009, Oates, Mukherjee, AoAS 2012, Babtie et al, PNAS 2014, Hill et al, Nature Methods 2016, Chen et al, JASA 2016,
Rudy et al, Science Advances 2017, Brunten et al, Nature Comm 2017, Mikkelsen, Hansen, arXiv 1710.09308, many more...

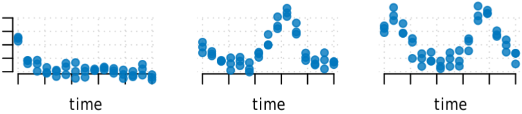


King et al, Nature 2004, Bongard et al, PNAS 2007, Calderhead et al, NIPS 2009, Schmidt et al, Science 2009, Oates, Mukherjee, AoAS 2012, Babtie et al, PNAS 2014, Hill et al, Nature Methods 2016, Chen et al, JASA 2016,
Rudy et al, Science Advances 2017, Brunten et al, Nature Comm 2017, Mikkelsen, Hansen, arXiv 1710.09308, many more...

nonlinear least squares:

$$\underset{\theta}{\operatorname{argmin}} \sum_{t \in \{t_1, \dots, t_m\}} \|\tilde{\mathbf{X}}_t - \mathbf{X}_t\|_2^2$$

$$\boxed{\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X_0} = \mathbf{x_0}}$$



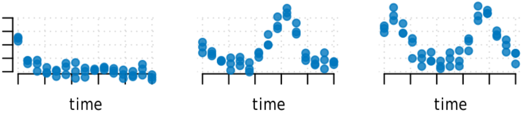
King et al, Nature 2004, Bongard et al, PNAS 2007, Calderhead et al, NIPS 2009, Schmidt et al, Science 2009, Oates, Mukherjee, AoAS 2012, Babtie et al, PNAS 2014, Hill et al, Nature Methods 2016, Chen et al, JASA 2016, Rudy et al, Science Advances 2017, Brunten et al, Nature Comm 2017, Mikkelsen, Hansen, arXiv 1710.09308, many more...

nonlinear least squares:

$$\underset{\theta}{\operatorname{argmin}} \sum_{t \in \{t_1, \ldots, t_m\}} \|\tilde{\mathbf{X}}_t - \mathbf{X}_t\|_2^2$$

$$\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X}_0 = \mathbf{x}_0$$

causality  prediction

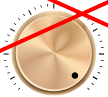


King et al, Nature 2004, Bongard et al, PNAS 2007, Calderhead et al, NIPS 2009, Schmidt et al, Science 2009, Oates, Mukherjee, AoAS 2012, Babtie et al, PNAS 2014, Hill et al, Nature Methods 2016, Chen et al, JASA 2016, Rudy et al, Science Advances 2017, Brunten et al, Nature Comm 2017, Mikkelsen, Hansen, arXiv 1710.09308, many more...

nonlinear least squares:

$$\underset{\theta}{\operatorname{argmin}} \sum_{t \in \{t_1, \ldots, t_m\}} \|\tilde{\mathbf{X}}_t - \mathbf{X}_t\|_2^2$$

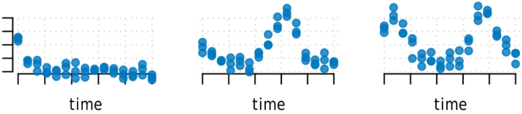
$$\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X}_0 = \mathbf{x}_0$$



causality prediction

$$\dot{Y}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X}_0 = \mathbf{x}_0$$

causality prediction



King et al, Nature 2004, Bongard et al, PNAS 2007, Calderhead et al, NIPS 2009, Schmidt et al, Science 2009, Oates, Mukherjee, AoAS 2012, Babtie et al, PNAS 2014, Hill et al, Nature Methods 2016, Chen et al, JASA 2016, Rudy et al, Science Advances 2017, Brunten et al, Nature Comm 2017, Mikkelsen, Hansen, arXiv 1710.09308, many more...

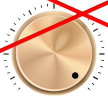
nonlinear least squares:

$$\operatorname*{argmin}_{\theta} \sum_{t \in \{t_1, \ldots, t_m\}} \|\tilde{\mathbf{X}}_t - \mathbf{X}_t\|_2^2$$

causal models are invariant

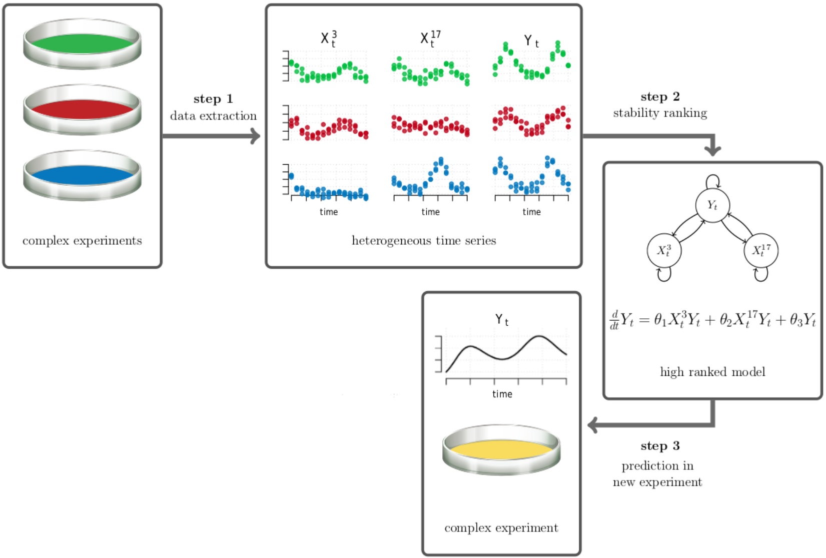
Haavelmo 1944, Aldrich 1989, Pearl 2009, ...

$$\dot{\mathbf{X}}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X_0} = \mathbf{x_0}$$

causality  prediction

$$\dot{Y}_t = g_\theta(\mathbf{X}_t, t), \quad \mathbf{X_0} = \mathbf{x_0}$$

causality  prediction



N. Pfister, S. Bauer, JP: *Identifying Causal Structure in Large-Scale Kinetic Systems*, arXiv:1810.11776

How invariant is

$$\dot{Y}_t = \theta_1 X_t^8 \text{ ?}$$

How invariant is

$$\dot{Y}_t = \theta_1 X_t^8 \text{ ?}$$

1. For each repetition i, smooth target trajectory. $\rightsquigarrow \hat{y}^{(i)}$

How invariant is

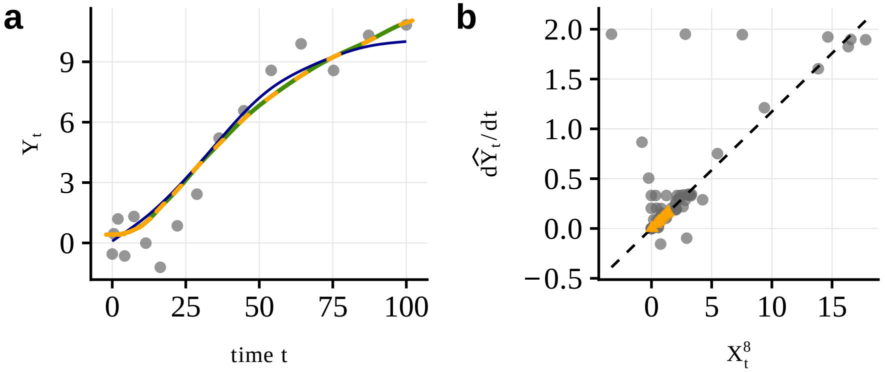
$$\dot{Y}_t = \theta_1 X_t^8 \text{ ?}$$

1. For each repetition i, smooth target trajectory. $\rightsquigarrow \hat{y}^{(i)}$
2. Obtain fitted values for target derivatives. $\rightsquigarrow \xi_{t_1}^{(i)}, \ldots, \xi_{t_m}^{(i)}$

How invariant is

$$\dot{Y}_t = \theta_1 X_t^8 \text{ ?}$$

1. For each repetition i, smooth target trajectory. $\rightsquigarrow \hat{y}^{(i)}$
2. Obtain fitted values for target derivatives. $\rightsquigarrow \xi_{t_1}^{(i)}, \ldots, \xi_{t_m}^{(i)}$
3. Smooth target trajectory, with constraints on derivatives. $\rightsquigarrow \hat{y}^{(i)}$



N. Pfister, S. Bauer, JP: *Identifying Causal Structure in Large-Scale Kinetic Systems*, arXiv:1810.11776

How invariant is

$$\dot{Y}_t = \theta_1 X_t^8 \ ?$$

1. For each repetition i, smooth target trajectory. $\rightsquigarrow \hat{y}^{(i)}$
2. Obtain fitted values for target derivatives. $\rightsquigarrow \xi_{t_1}^{(i)}, \ldots, \xi_{t_m}^{(i)}$
3. Smooth target trajectory, with constraints on derivatives. $\rightsquigarrow \hat{y}^{(i)}$

How invariant is

$$\dot{Y}_t = \theta_1 X_t^8 \text{ ?}$$

1. For each repetition i, smooth target trajectory. $\rightsquigarrow \hat{y}^{(i)}$
2. Obtain fitted values for target derivatives. $\rightsquigarrow \xi_{t_1}^{(i)}, \ldots, \xi_{t_m}^{(i)}$
3. Smooth target trajectory, with constraints on derivatives. $\rightsquigarrow \hat{y}^{(i)}$
4. Score for model ranking

$$\sum_{i=1}^{n} \left[\mathrm{RSS}^{(i)} - \mathrm{RSS}^{(i)}\right] / \left[\mathrm{RSS}^{(i)}\right],$$

where $\mathrm{RSS}^{(i)} := \sum_\ell \left(\hat{y}_{t_\ell}^{(i)} - Y_{t_\ell}^{(i)}\right)^2$.

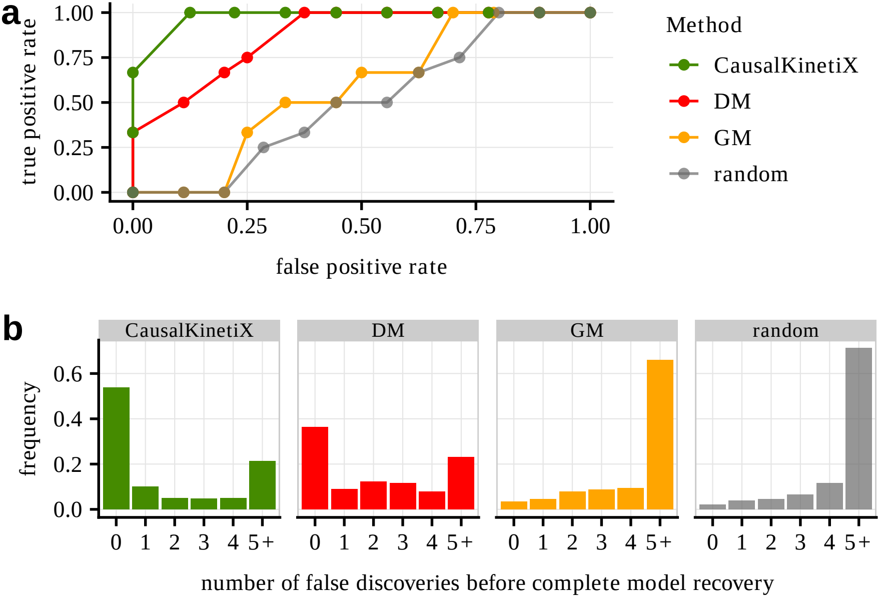
How invariant is

$$\dot{Y}_t = \theta_1 X_t^8 \ ?$$

1. For each repetition i, smooth target trajectory. $\rightsquigarrow \hat{y}^{(i)}$
2. Obtain fitted values for target derivatives. $\rightsquigarrow \xi_{t_1}^{(i)}, \ldots, \xi_{t_m}^{(i)}$
3. Smooth target trajectory, with constraints on derivatives. $\rightsquigarrow \hat{y}^{(i)}$
4. Score for model ranking

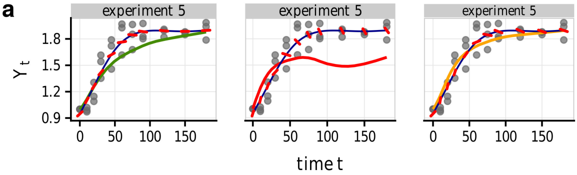
$$\sum_{i=1}^{n} \left[\mathrm{RSS}^{(i)} - \mathrm{RSS}^{(i)}\right] / \left[\mathrm{RSS}^{(i)}\right],$$

   where $\mathrm{RSS}^{(i)} := \sum_\ell \left(\hat{y}_{t_\ell}^{(i)} - Y_{t_\ell}^{(i)}\right)^2$.
5. Turn the score for models into a score for variables/complexes.



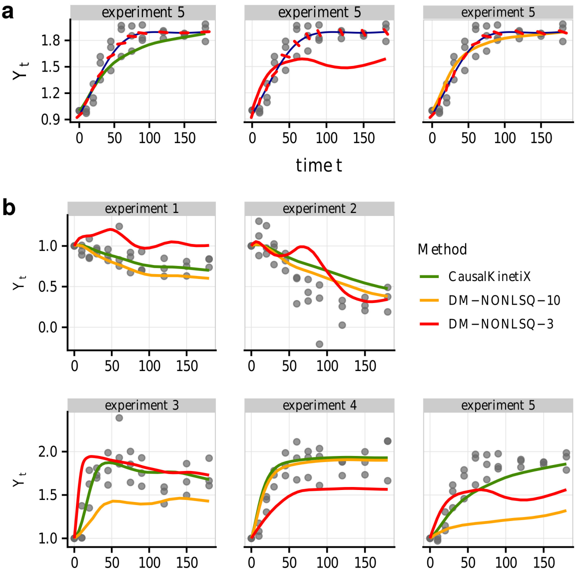
N. Pfister, S. Bauer, JP: *Identifying Causal Structure in Large-Scale Kinetic Systems*, arXiv:1810.11776

a) In-sample plot



N. Pfister, S. Bauer, JP: *Identifying Causal Structure in Large-Scale Kinetic Systems*, arXiv:1810.11776

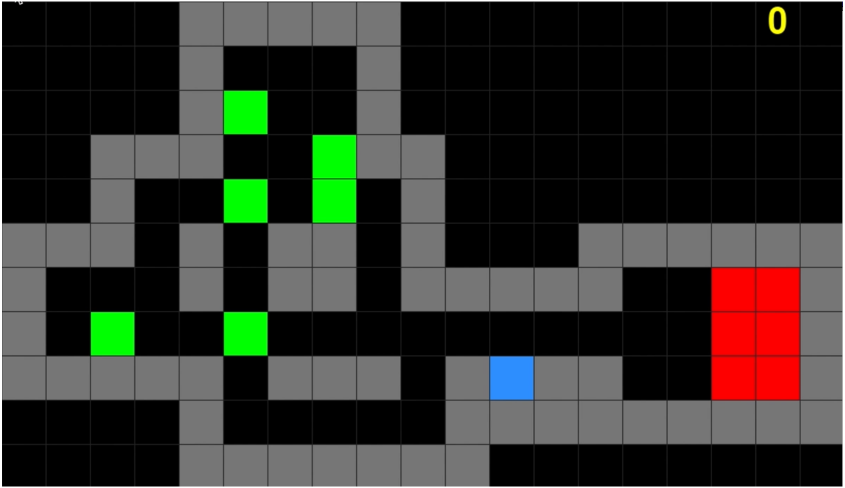
b) Out-of-sample plot



N. Pfister, S. Bauer, JP: *Identifying Causal Structure in Large-Scale Kinetic Systems*, arXiv:1810.11776

Idea 3: Exploiting Invariance in Reinforcement Learning:

joint work with Mads Thøisen

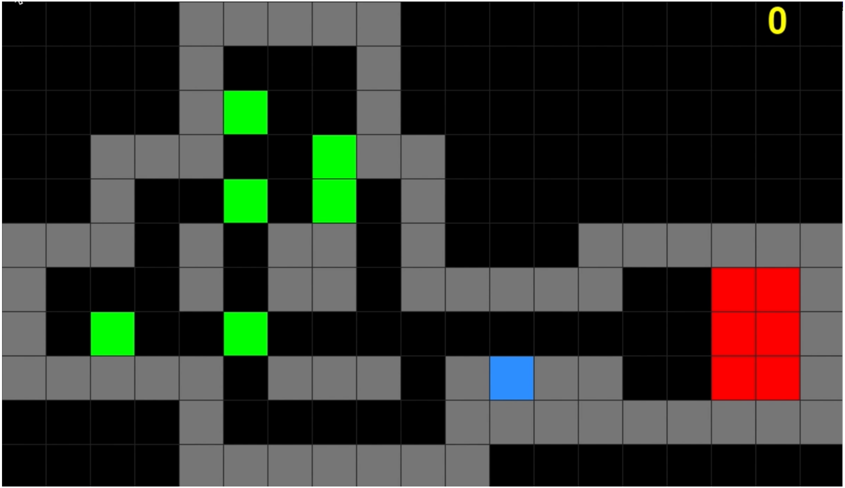


1. *Q*-learning: 1st epoch

Sutton and Barto (2015)

Idea 3: Exploiting Invariance in Reinforcement Learning:

joint work with Mads Thøisen

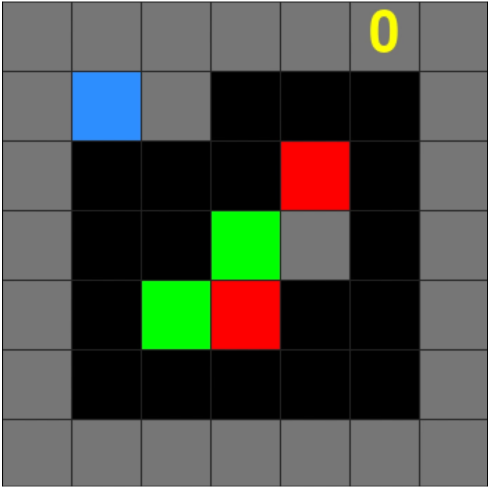


2. Q-learning: 20,000th epoch

Sutton and Barto (2015)

Idea 3: Exploiting Invariance in Reinforcement Learning:

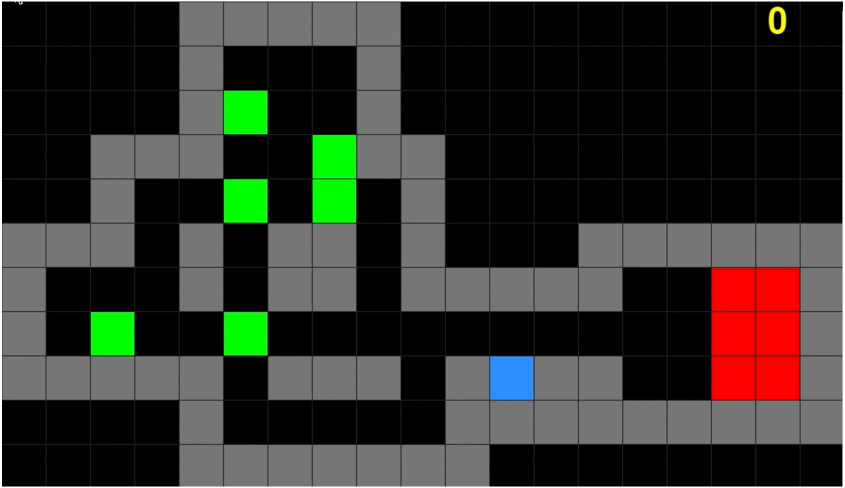
joint work with Mads Thøisen



3. Learning deadlocks from smaller levels.

Idea 3: Exploiting Invariance in Reinforcement Learning:

joint work with Mads Thøisen



4. Q-learning: 20,000th epoch – exploiting deadlocks

Conclusions

- Causal models $\neq$ predictive models.
- Causal models are invariant $\rightsquigarrow$ causal discovery.
- Combining invariance and predictability $\rightsquigarrow$ distributional robustness. (anchor regression, metabolic networks, reinforcement learning)

Conclusions

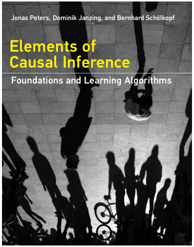
- Causal models $\neq$ predictive models.
- Causal models are invariant $\rightsquigarrow$ causal discovery.
- Combining invariance and predictability $\rightsquigarrow$ distributional robustness. (anchor regression, metabolic networks, reinforcement learning)

invariance  prediction

Conclusions

- Causal models $\neq$ predictive models.
- Causal models are invariant $\rightsquigarrow$ causal discovery.
- Combining invariance and predictability $\rightsquigarrow$ distributional robustness. (anchor regression, metabolic networks, reinforcement learning)

invariance  prediction



Tusind tak!

Book: JP, D. Janzing, B. Schölkopf: *Elements of Causal Inference: Foundations and Learning Algorithms*, MIT Press 2017

JP, P. Bühlmann, N. Meinshausen: *Causal inference using inv. pred.: identif. and conf. interv.* (w/ discussion), JRSSB 2016

R. Christiansen and JP: *Switching Regression Models and Causal Inference in the Presence of Latent Variables*, arXiv:1808.05541

N. Pfister, P. Bühlmann, JP: *Invariant causal prediction for sequential data*, JASA 2018

D. Rothenhäussler, P. Bühlmann, N. Meinshausen, JP: *Anchor regression: heterogeneous data meets causality*, arXiv:1801.06229

N. Pfister, S. Bauer, JP: *Identifying Causal Structure in Large-Scale Kinetic Systems*, arXiv:1810.11776