

表現学習

高次元データの背後にある構造を取り出せるか？

森岡 博史

滋賀大学 データサイエンス学部

大阪大学 情報科学研究科

理化学研究所 革新知能統合研究センター(AIP)

自己紹介 森岡博史



Associate Prof., Faculty of Data Science, **Shiga University**



Associate Prof. (cross-appt.), Grad. Sch. of IST, **Osaka University**



(~2025) Research Scientist, **RIKEN-AIP**



Titech → **Kyoto-U. (PhD)** → **U. Helsinki** → **UCL** → **RIKEN-AIP** → **Shiga & Osaka U.**



Robotics



Neuroengineering: Brain activity analysis (fMRI, MEG, EEG, NIRS);

(control)

High-dimensionality, nonlinearity, nonstationarity, limited label, variability,...



Machine learning: Unsupervised nonlinear representation learning;

Nonlinear independent component analysis, causal representation,...

$$\mathbf{x} = \mathbf{f}(\mathbf{s})$$

今回の内容: 観測の背後にある表現を取り出せるか?

Observed data

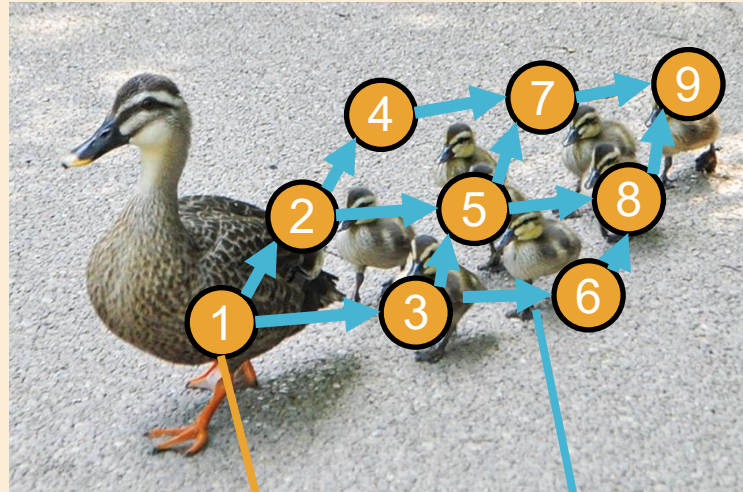


$\text{samples}(n) \times \text{pixels}(D)$

+ Explicit structural assumptions

Infer

Target: latent causal model



Latent variables

Latent dynamics, causality

Why learn this?

- 1 Intervention & decision making
- 2 Robust prediction / Generalization
- 3 Interpret hidden abstract mechanisms

- 問い: 観測データから, 背後にある表現を正しく(一意に)復元できるか?
- 今回のトーク: 本質的な難しさである識別性を説明
- 識別可能な例として非線形独立成分分析と因果表現学習を紹介

Outline

1. 表現学習とその難しさ
2. 非線形独立成分分析(NICA)
3. 因果表現学習への発展
4. まとめ

表現学習の必要性

高次元・冗長・観測条件の影響を含むデータ(x)

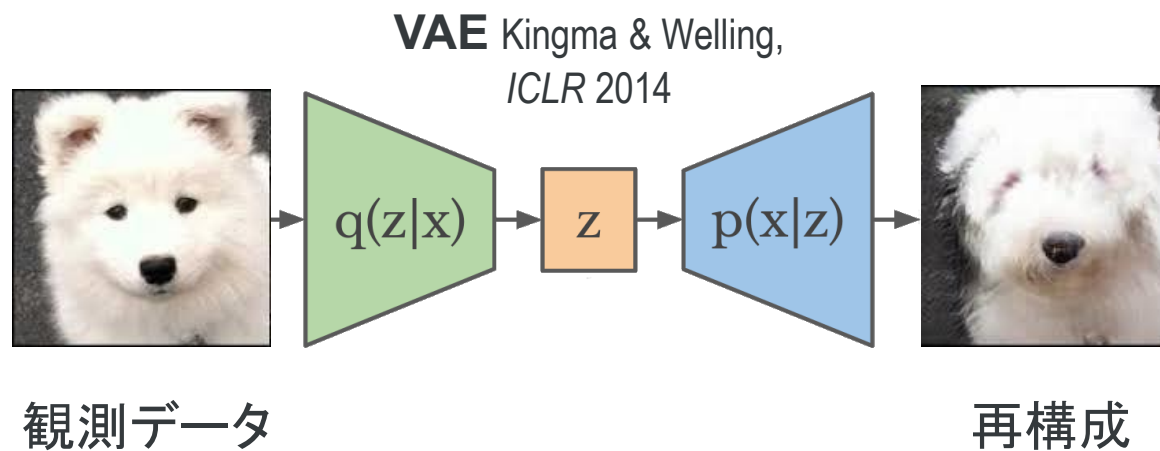


抽象的な表現(z)

$$\begin{bmatrix} 10 \\ -0.1 \\ \vdots \\ 2 \end{bmatrix} \begin{array}{l} \text{顔の特徴} \\ \text{向き} \\ \text{表情} \\ \text{背景} \\ \vdots \end{array}$$

- **表現学習**: 観測をよく表す, 抽象的で(しばしば低次元な)特徴をデータから学習
- (例)顔画像の場合: 個人特性, 表情, 姿勢, 照明, 背景など
- しかし, ラベルなしで学習した表現の「軸」には, どのような意味があるのか?

代表的な学習原理：再構成

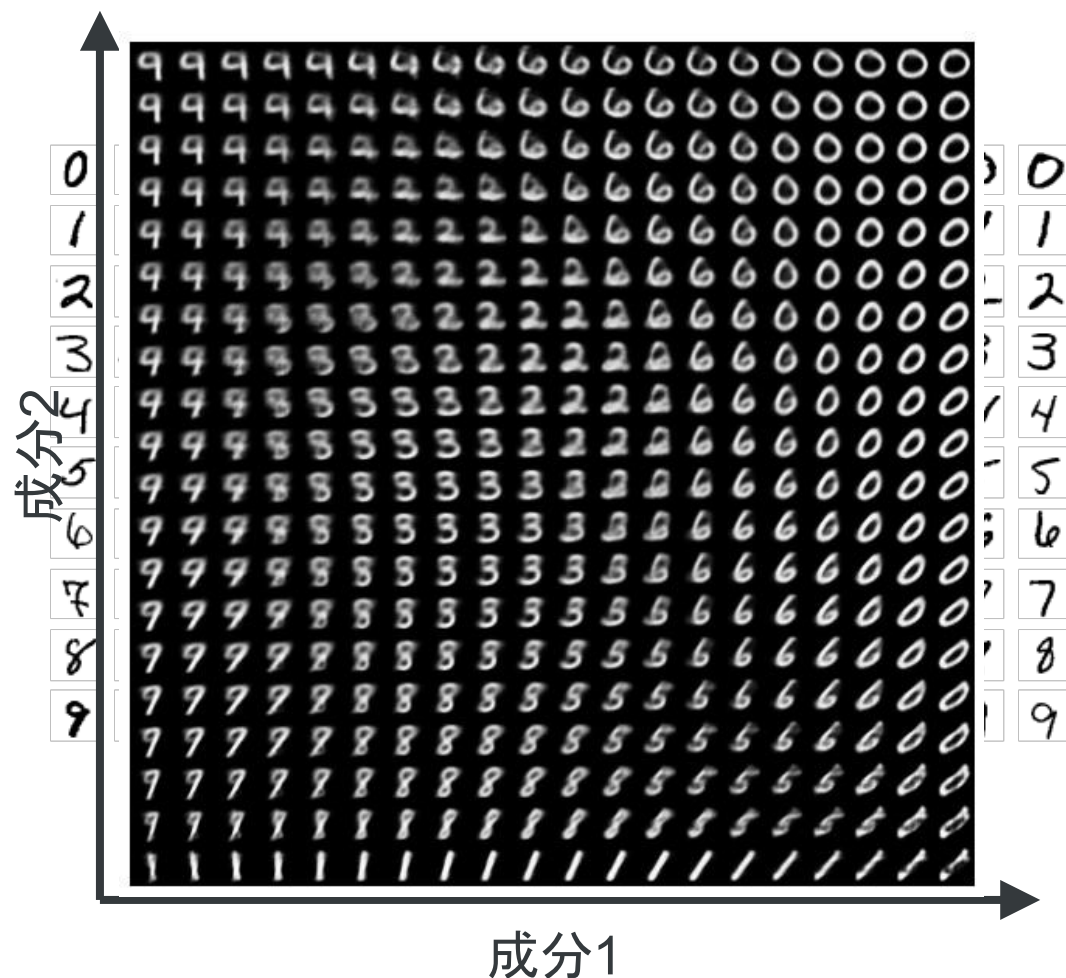


再構成原理： 観測を再構成できるように潜在表現（ z ）を学習する

しかし、再構成できれば表現の「軸」も一意に決まるのか？「よい」表現か？

表現学習の例

数字画像データ



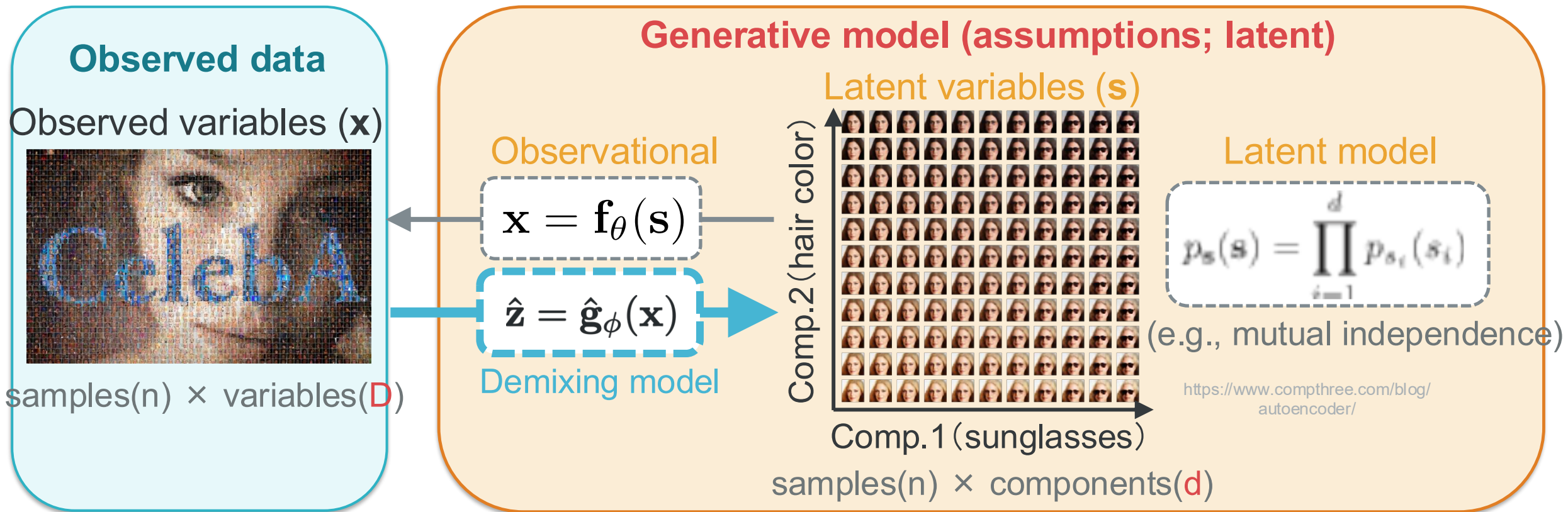
顔画像データ



表現学習の例



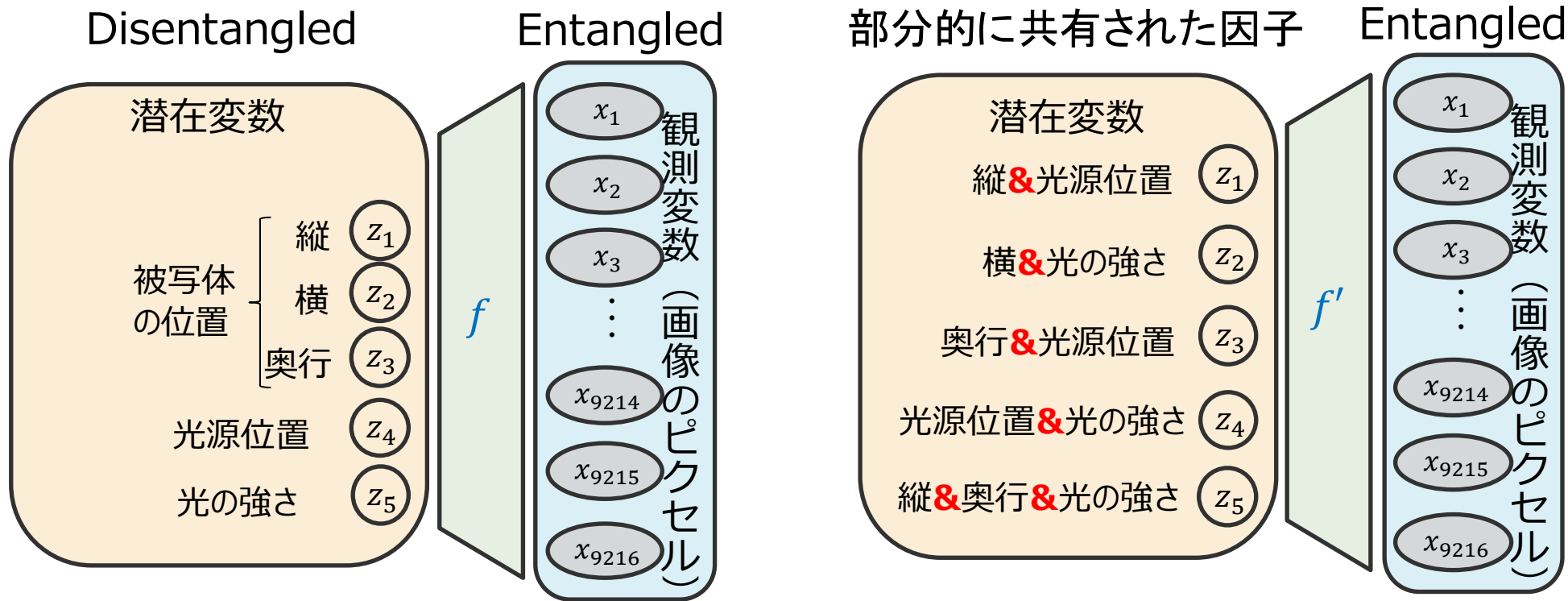
(教師なし)表現学習の問題設定



Goal: Identify and estimate the latent variables underlying the observations

But can we recover the true latent variables uniquely – and under what conditions?

正しい・意味のある表現とは？



- 各潜在要因が意味的に解かれている (disentangled) 表現
- どのような「意味」を仮定するかは問題依存. 独立に操作できる変数? 因果的に介入が加えられる変数? ... そのモデル化 (仮定, inductive bias) と識別性の保証が重要

識別性

識別性: 生成モデルに対する仮定を立てたとき, 観測分布からそのモデルが
(許される等価変換を除いて)唯一に絞り込める性質. 科学的解釈や下流課題
への汎化を正当化するために重要

生成モデル(仮定): 観測 $\mathbf{x} = \mathbf{f}_\theta(\mathbf{z})$, $\mathbf{z} \sim p_\theta(\mathbf{z})$
観測混合 (未知) 潜在成分 (未知)

識別性: $\forall(\theta, \theta'), \quad p(\mathbf{x}, \theta) = p(\mathbf{x}; \theta') \implies \theta \sim \theta'$

等価変換(例): 成分毎の(非線形)スケーリング, 並び替え. 仮定上本質的に不可避

注意点: 識別性は「学習の結果」でなく, 「モデルと仮定の性質」

識別性を保証しない仮定:再構成原理

(仮定)潜在成分は観測を再構成する → (推定法)再構成誤差最小化: AEなど

$$\text{目的関数: } \min_{\theta, \phi} \mathbb{E}_{p(\mathbf{x})} \left[\left\| \mathbf{x} - \hat{\mathbf{f}}_{\phi}(\mathbf{h}_{\theta}(\mathbf{x})) \right\|^2 \right]$$

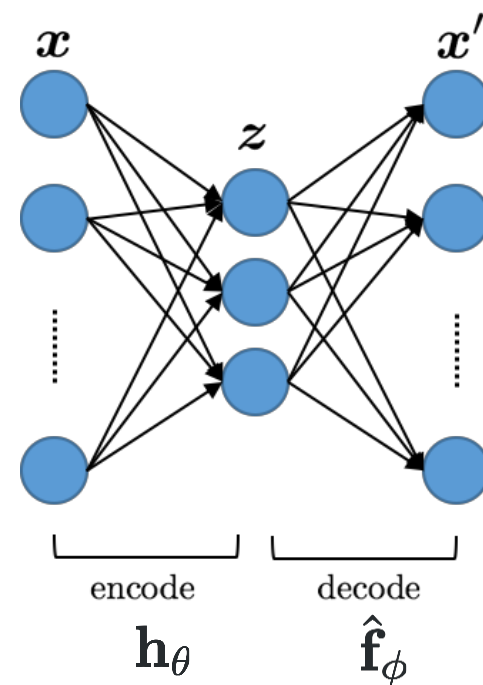
潜在成分(推定)

しかし, 任意の可逆変換Tに対して

$$(\hat{\mathbf{f}}_{\phi}, \mathbf{h}_{\theta}) \sim (\hat{\mathbf{f}}_{\phi} \circ \mathbf{T}^{-1}, \mathbf{T} \circ \mathbf{h}_{\theta})$$

(ケース1) (ケース2)

→ 再構成誤差は不変



観測分布・再構成の一意性 ≠ 表現の一意性

識別性の「不十分」な仮定: 相互独立性

(仮定) 潜在成分は互いに統計的独立 → (推定法) 相互情報量最小化: ICAなど

$$\text{目的関数: } \min_{\theta} I(\mathbf{h}_{\theta}(\mathbf{x})), \quad I(\mathbf{y}) = \text{KL} \left(p(\mathbf{y}) \parallel \prod_i p_i(y_i) \right)$$

潜在成分(推定)

しかし, 十分正則な連続確率変数の組 (y_1, y_2) に対して, 条件付きCDF $u = F_{y_2|y_1}(y_2|y_1)$ を新たな変数としてとることで $u \sim \text{Uniform}(0, 1)$, $u \perp\!\!\!\perp y_1$ Darmois construction, 1952

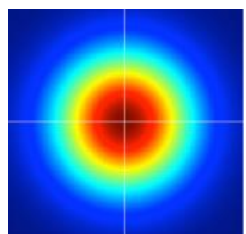
→ 相互独立性はいくらでも「作れる」(非線形変換で依存性を吸収できる)

相互独立性 ≠ 表現の一意性(非線形の場合)

識別性の「不十分」な仮定: 相互独立性 (図例)

線形モデル

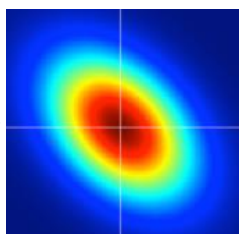
潜在成分(真)



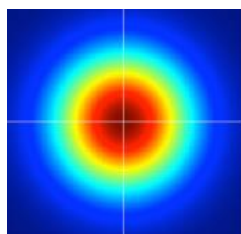
$$\mathbf{s}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$$

観測



潜在成分(偽)

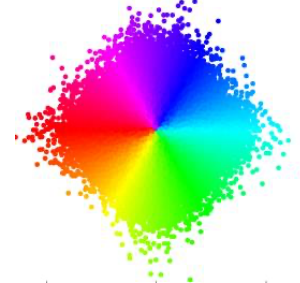


$$\mathbf{R}\mathbf{s}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\mathbf{x}_t = \mathbf{A}'\mathbf{s}'_t = \mathbf{A}\mathbf{R}^{-1}\mathbf{R}\mathbf{s}_t$$

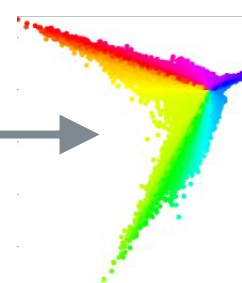
非線形モデル

潜在成分(真)

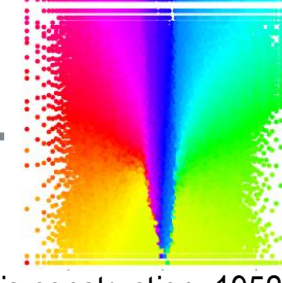


$$\mathbf{x}_t = \mathbf{f}_\theta(\mathbf{s}_t)$$

観測



潜在成分(偽)



Dargois construction, 1952

$$\mathbf{x}_t = \mathbf{f}_{\theta'}(\mathbf{s}'_t)$$

○ **線形モデル**: 非ガウス性を追加で仮定すればよい → 独立成分分析ICA Comon1994

× **非線形モデル**: 確率分布の仮定では不十分 → ICA分野で長い間未解決問題 Hyvärinen & Pajunen, 1999

ここまでのまとめ：表現学習の不定性とその原因

不定性の原因：表現学習問題の対称性の強さ(データ不足の問題ではない)

- 再構成性を保つ変換は極めて多く、独立性を保つ非自明な非線形変換も存在
- 異なる潜在表現が同じ観測分布や目的関数値を与えるため、これらの基準だけでは区別できない

→ 学習した各軸を、真の潜在要因と対応付ける根拠がない...

Question:

- 識別性が保証できる仮定とは？
- 仮定を満たすモデルはどのような目的関数を使えば学習可能？

Outline

1. 表現学習とその難しさ
2. 非線形独立成分分析(NICA)
3. 因果表現学習への発展
4. まとめ

すべての問題に対して最適なモデルは存在するか？

- 「すべてのモデルは間違っている. だが中には役立つものもある」 Box87
- ある分野でよく当てはまる仮定の集合 (inductive bias) が, 他の分野では機能しないことが多々ある

表現学習: 対象の特性をうまく表し, かつ観測の背後の潜在構造を一意に定める仮定は何か？

どうすれば対称性を破れるか：非線形独立成分分析(NICA)

非線形ICA問題とは...

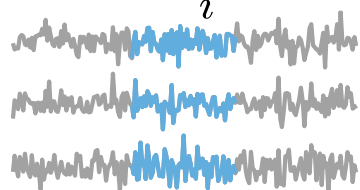
「独立性を保ちつつ、可逆変換の対称性を破る実用的かつ最小な構造」

を探す問題である

非線形ICA: 構造に対する追加条件 → 対称性を破る

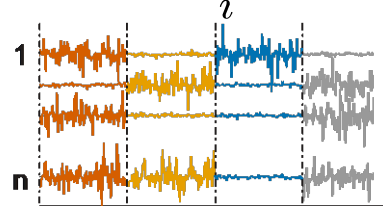
非線形独立成分分析: 独立性 + 時間・補助変数などの追加構造 → 識別性

$$p(\mathbf{s}) = \prod_i p_i(s_i) \longrightarrow p(\mathbf{s}|\mathbf{u}) = \prod_i p_i(s_i|\mathbf{u}) \left\{ \begin{array}{l} \text{条件付き独立} \\ \mathbf{u} \text{ に依存して分布が変化} \end{array} \right.$$



i.i.d

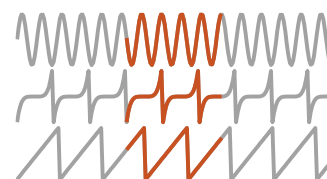
Hyvärinen & Pajunen, 1999



$\mathbf{u} = \text{segment}(t)$

非定常

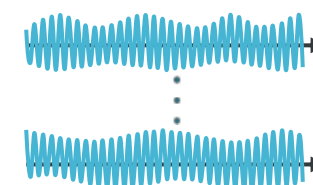
Hyvärinen and Morioka, *NIPS* 2016



$\mathbf{u} = \mathbf{s}_{t-1}$

自己回帰

Hyvärinen and Morioka, *AISTATS* 2017



$\mathbf{u} = t$

一般化

Hyvärinen et al., *AISTATS* 2019
Halva et al., *NeurIPS* 2021

なぜ構造が識別性を生む？

- 構造は「整合性条件」を増やす → 可逆変換の自由度が消える → 識別性が生まれる
- i.i.d. → 1つの分布だけ説明すればよい. 時間構造 → 複数の条件を同時に説明する必要

$\mathcal{G}$: すべての可逆非線形変換

$\mathcal{G}_{\text{structure}} \subset \mathcal{G}$: 条件と整合する変換

非線形独立成分分析(NICA)

では、モデルに識別性があったとして、どのような目的関数を用いれば学習できる？

データ構造を教師信号として利用する自己教師学習

- 構造の違いを識別する補助課題を解くことで、識別可能な潜在表現を間接的に学習

Hyvärinen and Morioka, *NIPS* 2016, *AISTATS* 2017

- 通常の分類問題として実装でき、比較的単純な学習手順で推定できる

非線形ICA 1: 時間区分定常モデル

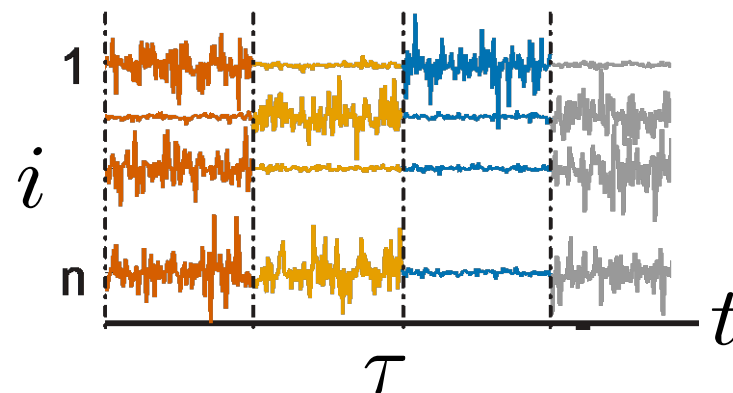
生成モデルの仮定: データ生成過程に対する我々の信念 (inductive bias)

- 観測モデル: $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, C^2 diffeomorphism

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t))$$

- 潜在モデル: Exponential family

$$\log p_{\tau}(s_i) = q_{i,0}(s_i) + \sum_{v=1}^V \underbrace{\lambda_{i,v}(\tau)}_{\text{パラメータ (非定常性)}} \underbrace{q_{i,v}(s_i)}_{\text{十分統計量}} - \log Z(\lambda_{i,1}(\tau), \dots, \lambda_{i,v}(\tau))$$



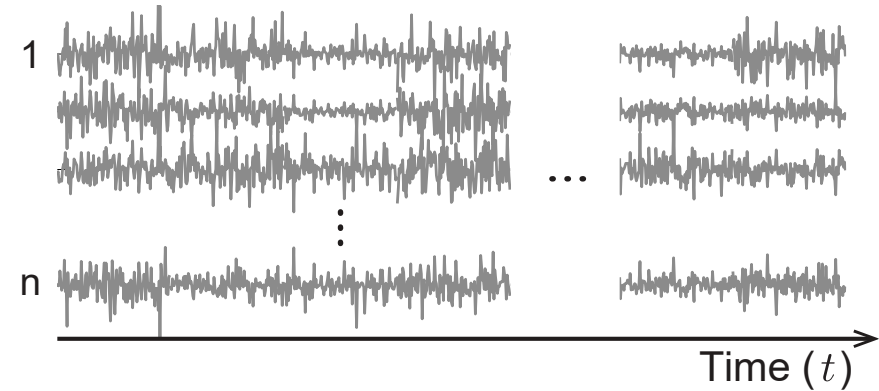
○ 例: ガウス分布で, 各時刻の平均と分散が λ_1, λ_2 で決まる: $q_{i1}(s_i) = s_i$, $q_{i2}(s_i) = s_i^2$

- 識別性: 確率分布が変数間・セグメント間で十分に变化すればOK(後述)
- 推定法: 非定常構造を利用した自己教師学習が可能

非線形ICA 1 Time-contrastive learning (TCL): 時間帯を予測する

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は非定常(時間区分定常)



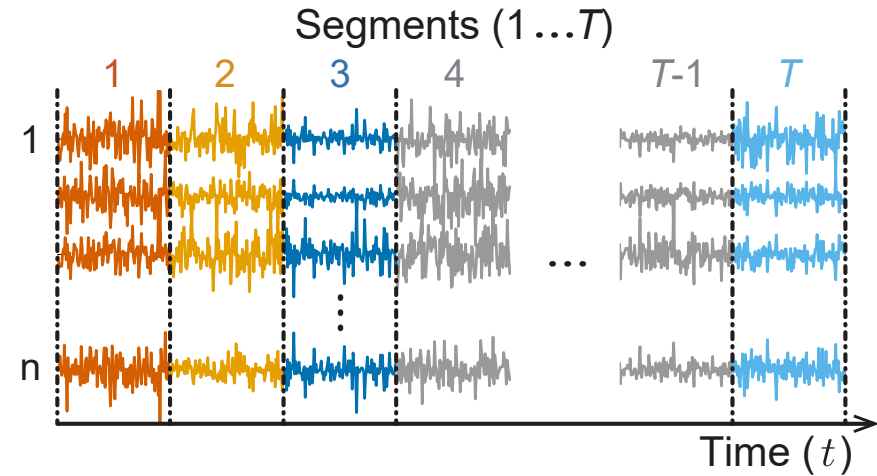
非線形ICA 1 Time-contrastive learning (TCL): 時間帯を予測する

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は非定常(時間区分定常)

Time-contrastive (TCL):

1. 定常な時間セグメントに分ける.



非線形ICA 1 Time-contrastive learning (TCL): 時間帯を予測する

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は非定常(時間区分定常)

Time-contrastive (TCL):

1. 定常な時間セグメントに分ける.
2. 非線形特徴抽出器(MLP)と識別機(MLR)を, セグメントラベルを予測するように学習する.

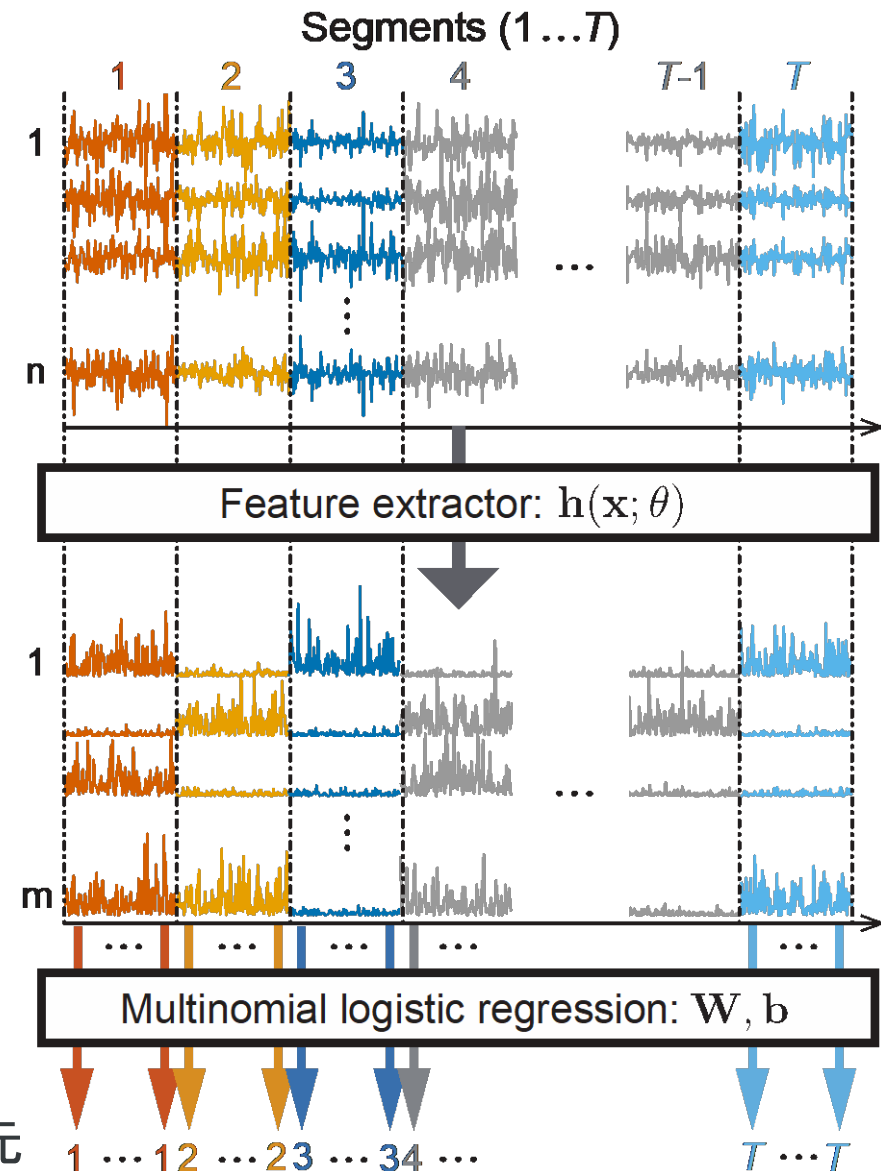
$$p(C = \tau | \mathbf{x}; \theta, \mathbf{W}, \mathbf{b}) = \frac{\exp(\mathbf{w}_\tau^\top \mathbf{h}_\theta(\mathbf{x}) + b_\tau)}{\sum_{j=1}^T \exp(\mathbf{w}_j^\top \mathbf{h}_\theta(\mathbf{x}) + b_j)}$$

識別性・推定法のアイデア:

最適なセグメント分類

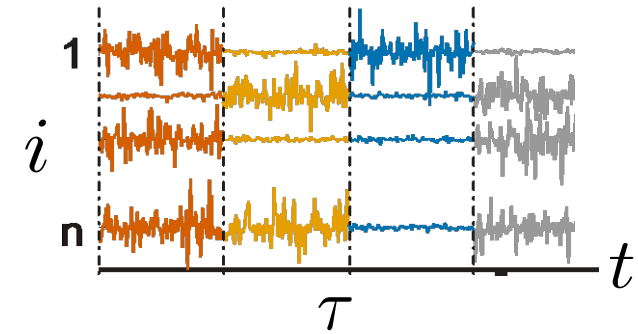
→ セグメント間の対数密度比を学習

→ 十分な分布変化で十分統計量を線形変換まで復元



仮定: TCLを以下の仮定を満たすデータに適用し $h(\mathbf{x})$ を学習

1. データが非線形ICAモデルで計測される; $\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t))$
2. $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is sufficiently smooth and invertible
3. $s_i(t)$ は互いに(条件付き)独立, 各セグメント τ における分布はexponential family ($V=1$)
$$\log p_\tau(s_i) = q_{i,0}(s_i) + \sum_{v=1}^V \lambda_{i,v}(\tau) q_{i,v}(s_i) - \log Z(\lambda_{i,1}(\tau), \dots, \lambda_{i,v}(\tau))$$
4. 分布が“十分に”非定常; $\mathbf{L} = [\lambda_{i,1}(\tau) - \lambda_{i,1}(1)]_{\tau,i}$ is full-column rank



Then: 学習特徴 $h(\mathbf{x})$ は、潜在成分の十分統計量 $q(\mathbf{s})$ を可逆線形変換まで表現:

$$q(\mathbf{s}_t) = \mathbf{A}h(\mathbf{x}_t; \boldsymbol{\theta}) + \mathbf{d} \quad \text{for some } \mathbf{A} \in \mathbb{R}^{n \times n}, \mathbf{d} \in \mathbb{R}^n$$

なぜcontrastをとると潜在変数が得られるのか？

観測モデル + 潜在モデル → 観測分布. モデル化の困難な項が混在

$$\begin{aligned}\log p_\tau(\mathbf{x}) &= \log p_\tau(\mathbf{f}^{-1}(\mathbf{x})) + \log |\det \mathbf{J}_{\mathbf{f}^{-1}}(\mathbf{x})| \\ &= \sum_{i=1}^n \left(\underbrace{q_{i,0}(s_i)}_{\text{不明}} + \underbrace{\sum_{v=1}^V \lambda_{i,v}(\tau) q_{i,v}(s_i)}_{\text{推定困難}} - \underbrace{\log Z_i(\tau)}_{\text{不明}} \right) + \log |\det \mathbf{J}_{\mathbf{f}^{-1}}(\mathbf{x})|\end{aligned}$$

最適な分類器は, この対数密度「比」を学習する → セグメント間で共通の項が消える

$$\begin{aligned}\log p_\tau(\mathbf{x}) - \log p_1(\mathbf{x}) &= \sum_{i=1}^n \sum_{v=1}^V \underbrace{(\lambda_{i,v}(\tau) - \lambda_{i,v}(1)) q_{i,v}(s_i)}_{\text{xに依存しない}} + \underbrace{c_\tau}_{\text{xに依存しない}} \\ &= \underbrace{(\mathbf{w}_\tau - \mathbf{w}_1)^\top \mathbf{h}(\mathbf{x}) + (b_\tau - b_1)}_{\text{推定モデル(線形識別機)}}\end{aligned}$$

分布変化が十分に多様(仮定4)なら, 十分統計量を線形に解き出せる

非線形ICA 2: 定常自己回帰モデル

生成モデルの仮定: データ生成過程に対する我々の信念 (inductive bias)

- 観測モデル: $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, C^2 diffeomorphism

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t))$$

- 潜在モデル: 成分ごとに独立な定常マルコフ過程 + 正則性条件

- 例1: 非ガウシアン $\log p(s(t)|s(t-1)) = \underline{G}(s(t) - \rho s(t-1))$

Non-Gaussian

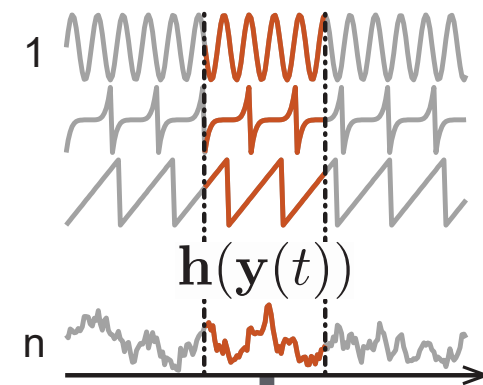
- 例2: 非線形ガウス自己回帰 $\log p(s(t)|s(t-1)) = -\lambda(s(t) - \underline{r}(s(t-1)))^2 + c$

Nonlinear

- 特別な例: 線形ガウス → 線形変換まで識別可能

- 識別性: 十分な隣接時点間依存性と非準ガウス性があればOK (後述)

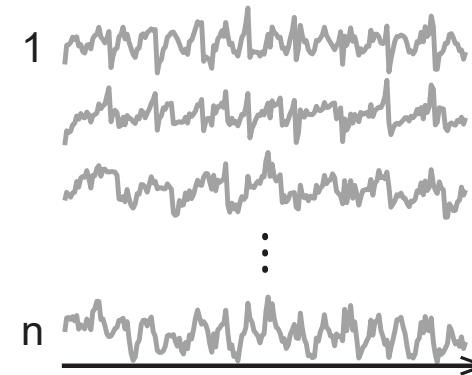
- 推定法: 隣接構造を利用した自己教師学習が可能



非線形ICA 2 Permutation-contrastive learning (PCL)

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は定常, 自己相関構造を持つ



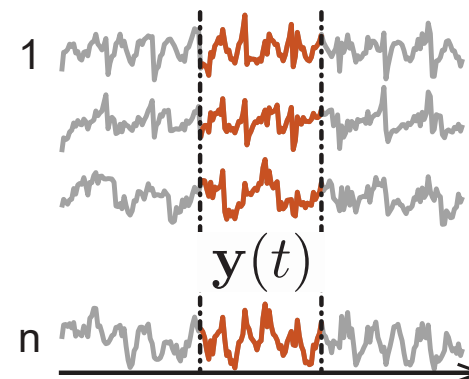
非線形ICA 2 PCL: 本物の時間的組とシャッフルした組を見分ける

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は定常, 自己相関構造を持つ

Permutation-contrastive (PCL):

1. 短い窓をとる; $y(t) = (x(t), x(t-1))$



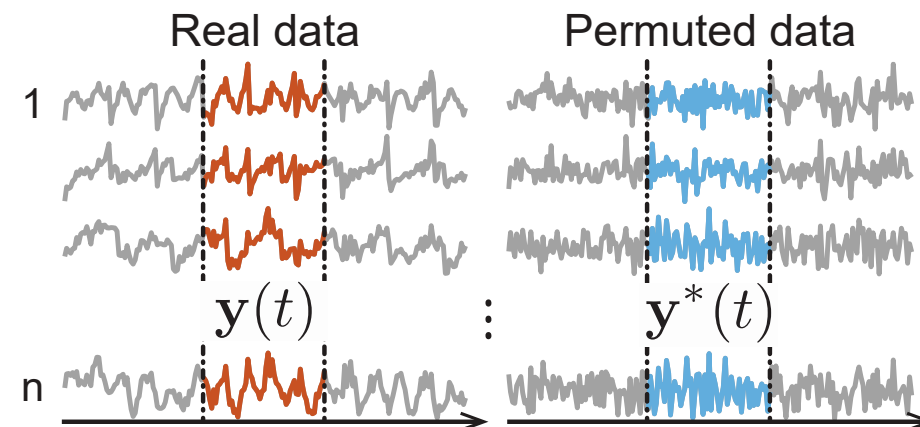
非線形ICA 2 PCL: 本物の時間的組とシャッフルした組を見分ける

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は定常, 自己相関構造を持つ

Permutation-contrastive (PCL):

1. 短い窓をとる; $y(t) = (x(t), x(t-1))$
2. 時間的にシャッフルする; $y^*(t) = (x(t), x^*)$



非線形ICA 2 PCL: 本物の時間的組とシャッフルした組を見分ける

生成モデルの仮定:

1. 観測信号はソース信号の非線形混合($f(s)$)
2. ソース信号は定常, 自己相関構造を持つ

Permutation-contrastive (PCL):

1. 短い窓をとる; $y(t) = (x(t), x(t-1))$
2. 時間的にシャッフルする; $y^*(t) = (x(t), x^*)$
3. 非線形特徴抽出器 (MLP) と識別機 (LR) を, y と y^* を識別するように学習する (2値判別問題)

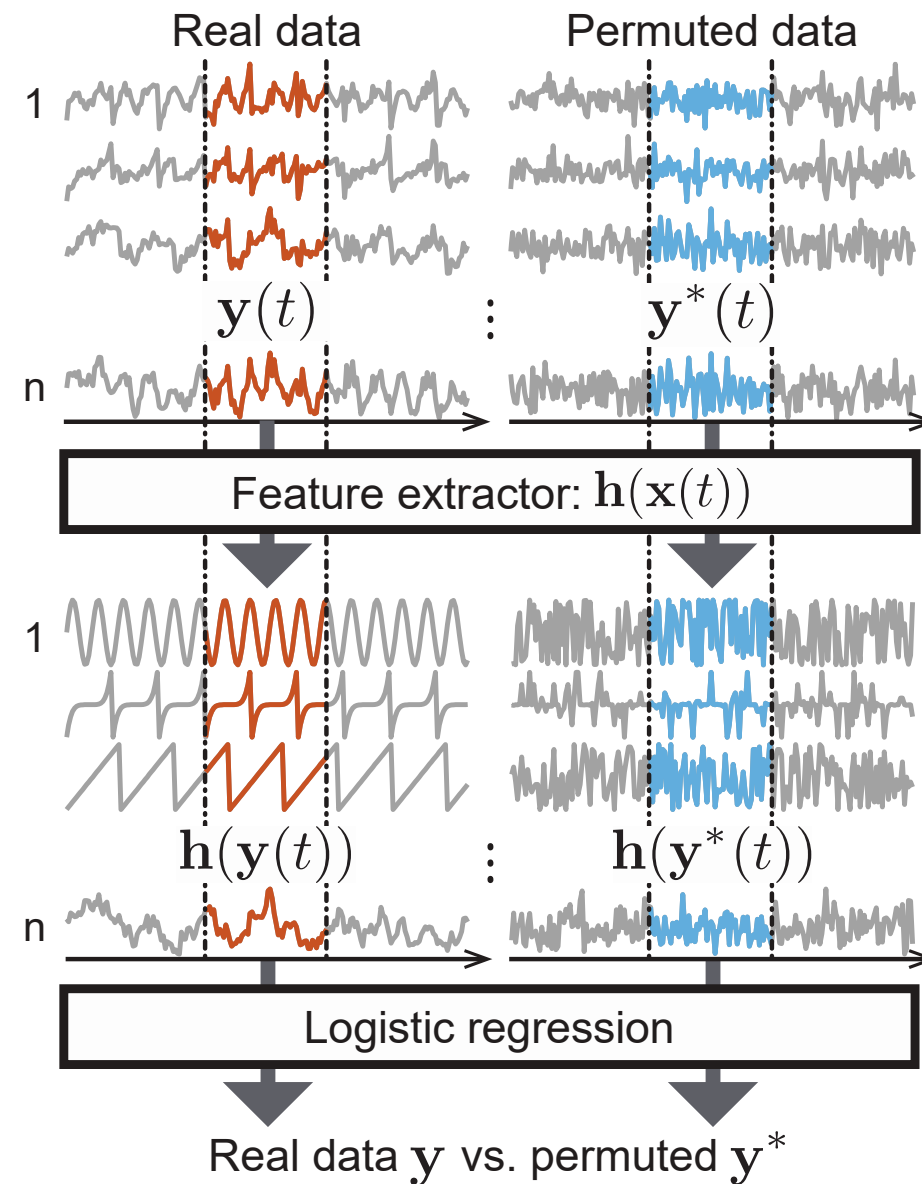
$$p(C = 1 | \mathbf{x}; \theta) = \sigma \left(\sum_{i=1}^n B_i (h_i(\mathbf{x}(t)), h_i(\mathbf{x}(t-1))) \right)$$

識別性のアイデア:

実データ／シャッフルデータの最適分類

→ 時間依存を表す対数密度比を学習

→ 仮定と分類器の構造のもとで潜在成分を復元



Permutation-contrastive learning (PCL)の定理

仮定:PCLを以下の仮定を満たすデータに適用し $h(\mathbf{x})$ を学習

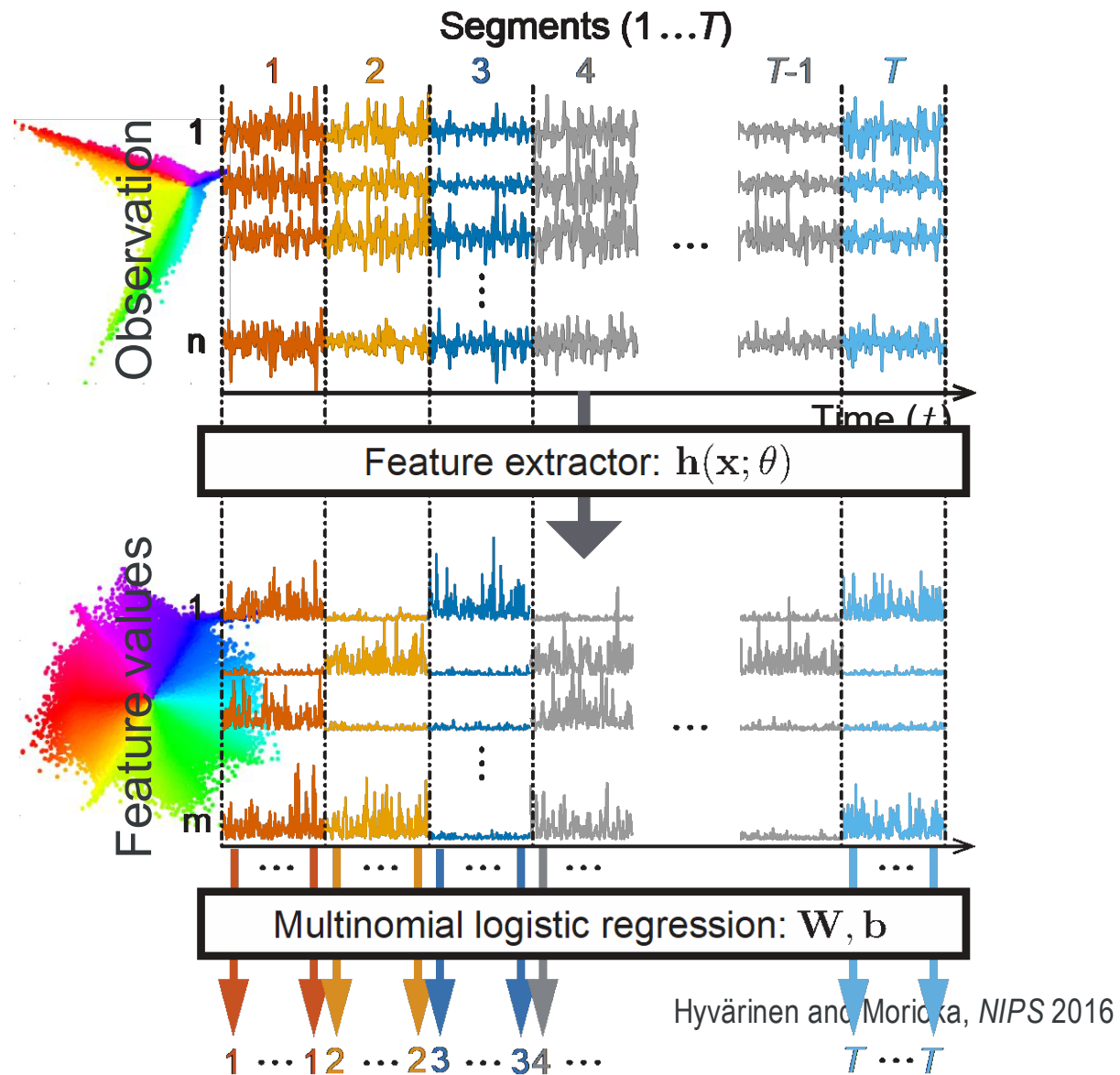
1. データが非線形ICAモデルで計測される; $\mathbf{x}(t) = \mathbf{f}(s(t))$
2. $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is sufficiently smooth and invertible $s_i(t)$ は互いに独立, 定常
3. **Uniform dependency:** $\frac{\partial^2 \log p(s(t), s(t-1))}{\partial s(t) \partial s(t-1)} \neq 0$ for all $(s(t), s(t-1))$ (stronger than dependence)
4. **Non-quasi-Gaussianity:** $\frac{\partial^2 \log p(s(t), s(t-1))}{\partial s(t) \partial s(t-1)} \neq c \alpha(s(t)) \alpha(s(t-1))$ (stronger than non-Gauss.)
5. Regression function: $p(C = 1 | \mathbf{x}; \theta) = \sigma \left(\sum_{i=1}^n B_i(h_i(\mathbf{x}(t)), h_i(\mathbf{x}(t-1))) \right)$

Then: $h(\mathbf{x})$ が潜在成分を表現. ただし成分ごとの可逆非線形変換と順序の不定性;

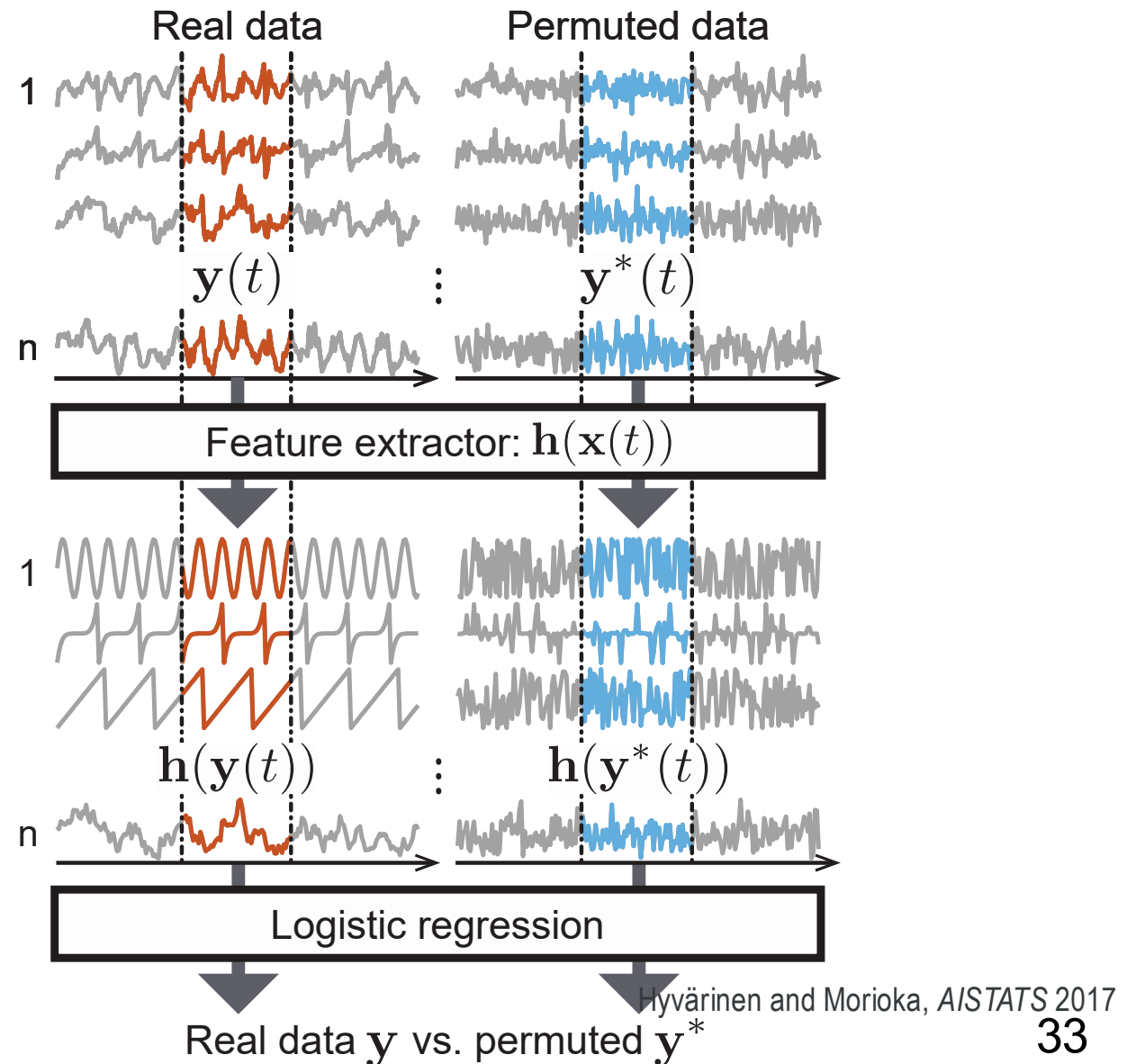
$$s_i(t) = k_i(h_{\pi(i)}(\mathbf{x}(t))) \text{ for some permutation } \pi \text{ and invertible scalar functions } k_i$$

TCL vs. PCL

Time-contrastive learning (TCL)

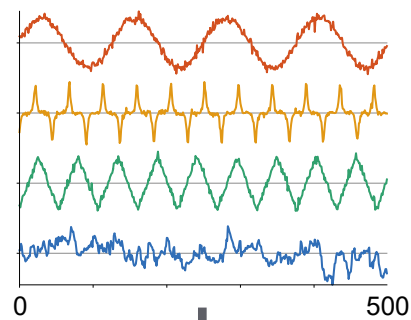


Permutation-contrastive learning (PCL)

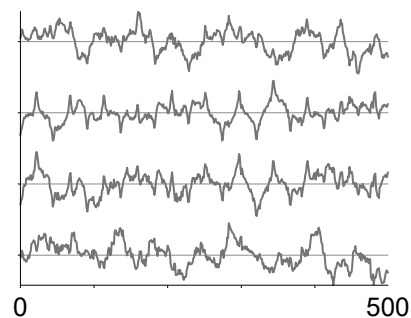


推定例: VAE vs NICA

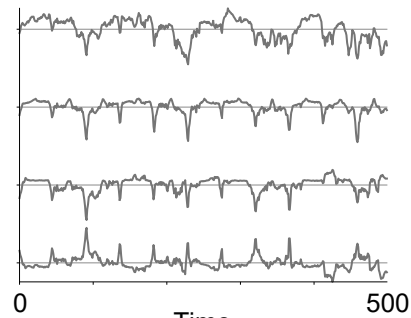
Independent components



Linear mixing



Nonlinear mixing

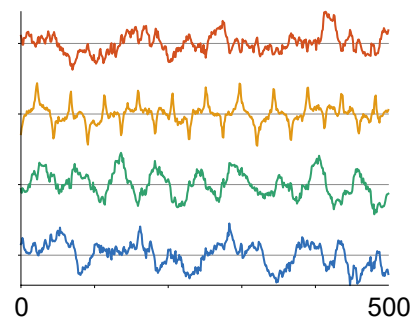


時間構造の仮定

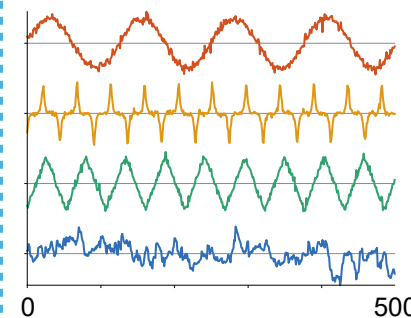
→ 推定可能

Estimated components

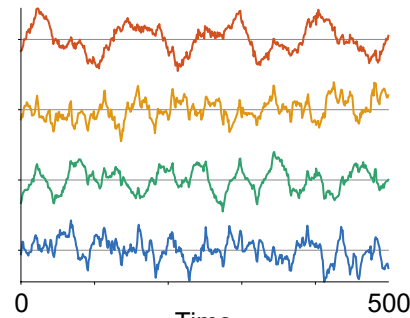
PCA



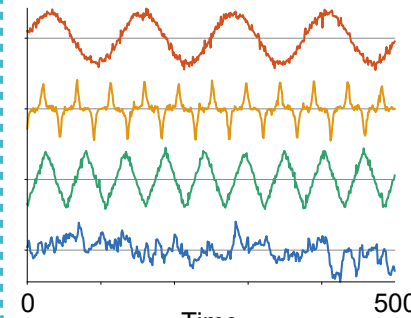
FastICA



VAE

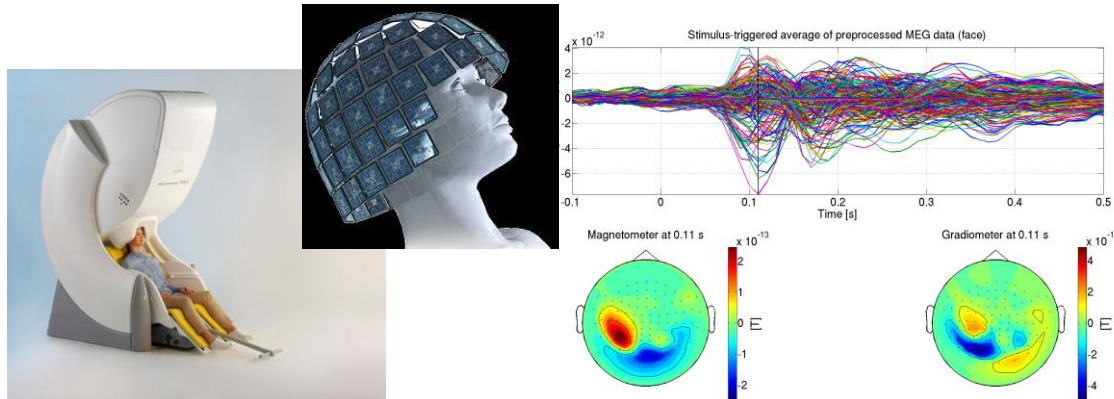


Nonlinear ICA (PCL)



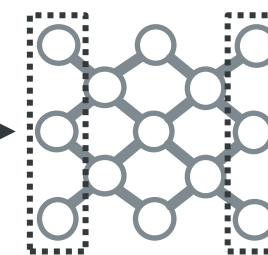
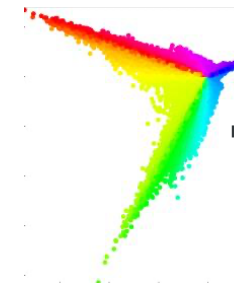
安静時脳活動計測データ(脳磁図)からの表現学習

データ: 9被験者, 安静時, MEG(30分, ラベルなし)

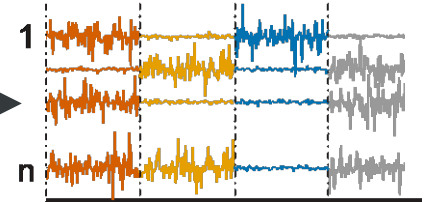


非線形ICA: Nonlinear ICA +
DNNMLPに基づく
Feature
extractor

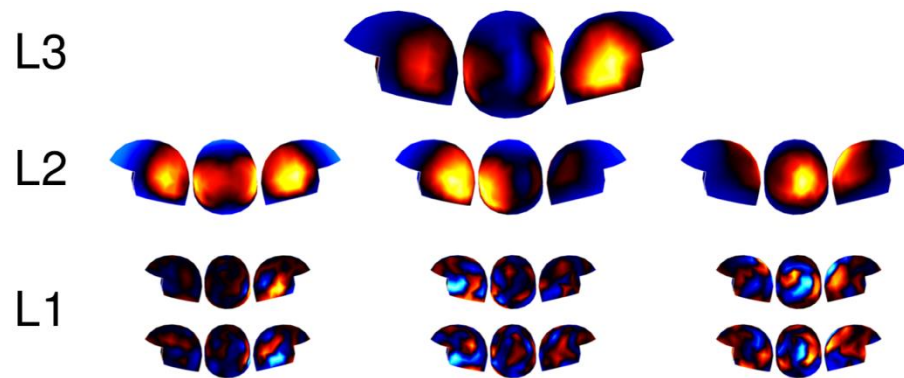
MEG 300dim.



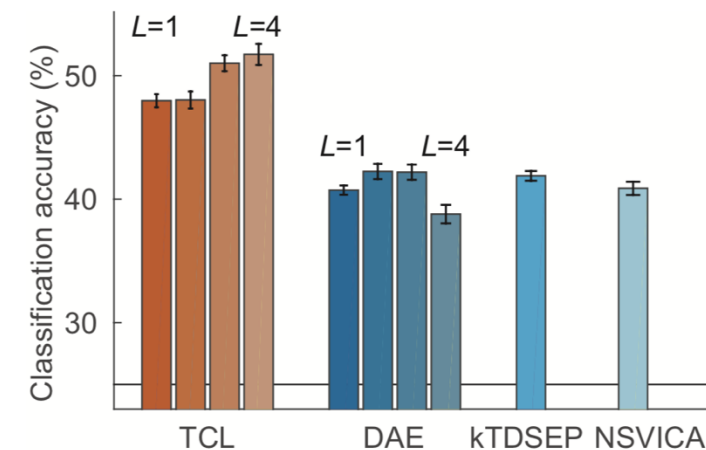
Nonstationary
features (10 dim)



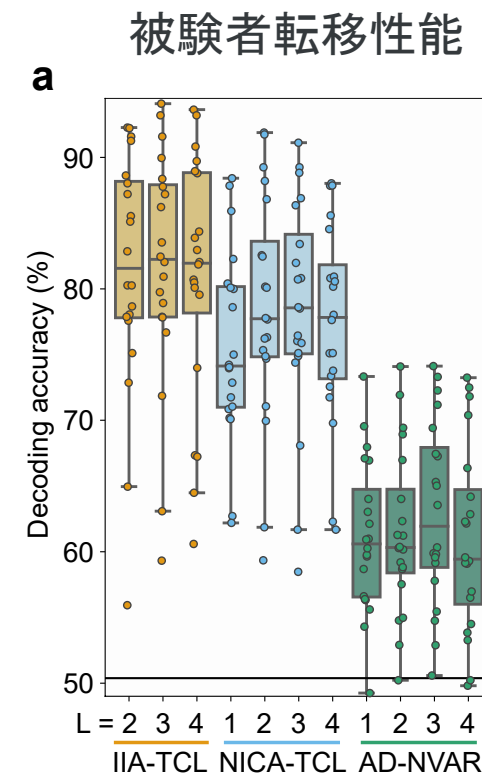
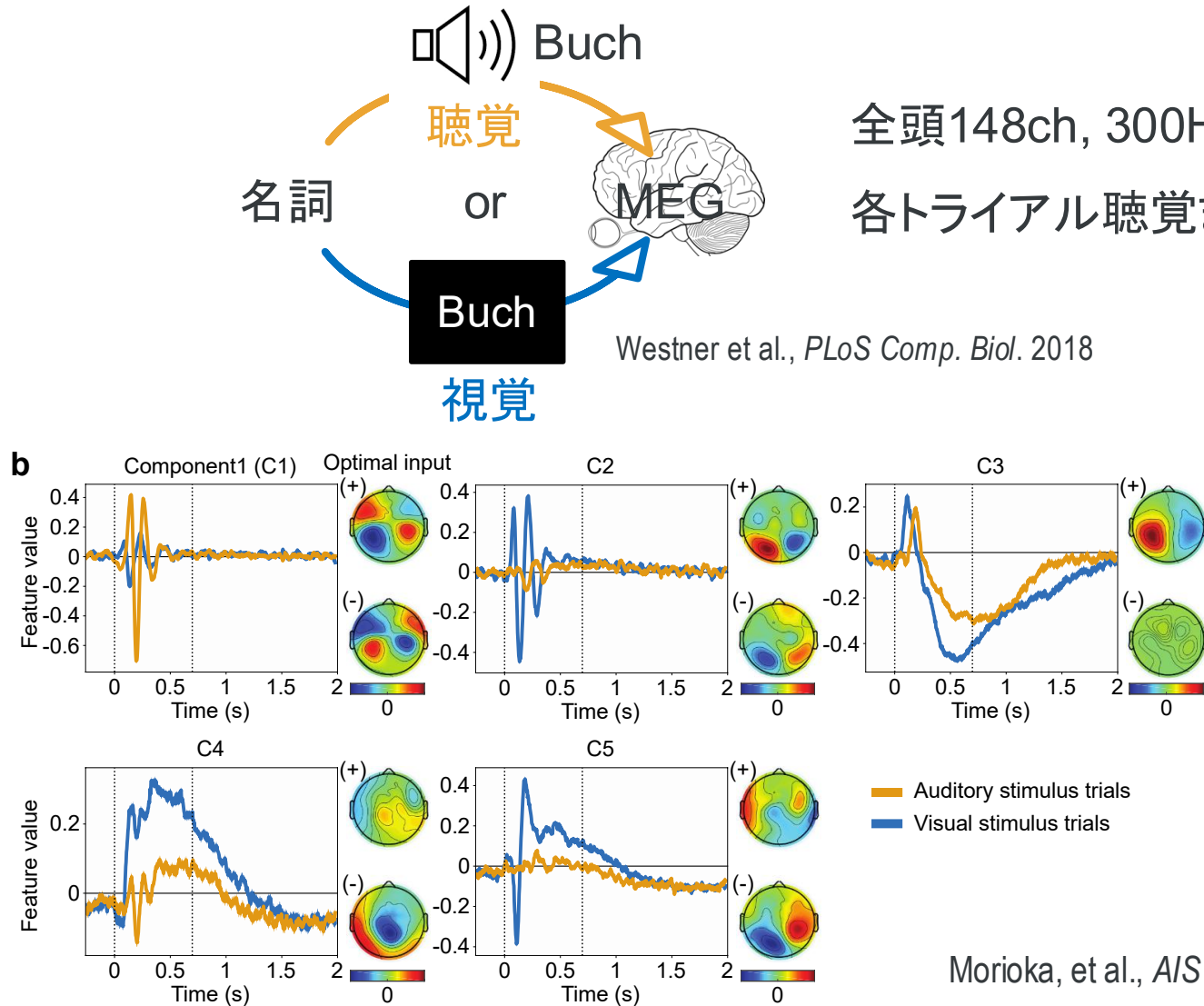
特徴量: 階層的空間パターン



下流タスク: 4種類の運動想像の判別



脳活動計測データ(MEG)から外部刺激により駆動される脳活動成分を推定



非線形ICAにより, 刺激モダリティに対応し, 被験者間で汎化する成分を教師なしで抽出

Outline

1. 表現学習とその難しさ
2. 非線形独立成分分析(NICA)
3. 因果表現学習への発展
4. まとめ

独立性から因果関係へ：因果表現学習が目指すもの

Observed data

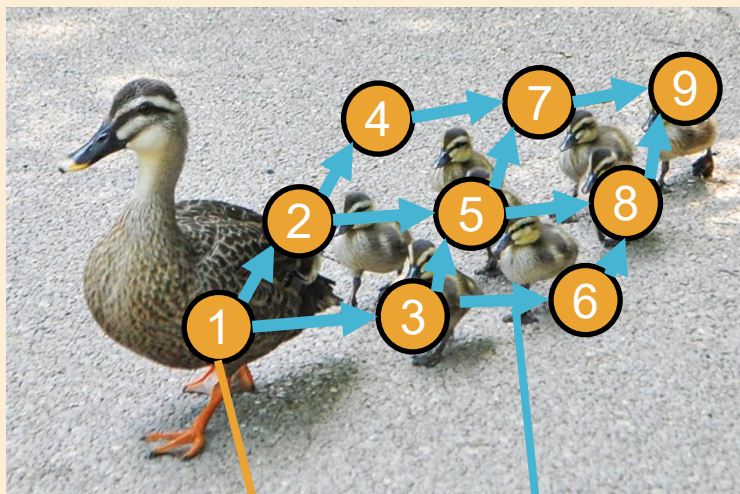


$\text{samples}(n) \times \text{pixels}(D)$

+ Explicit structural assumptions

Infer

Target: latent causal model



Causal variables

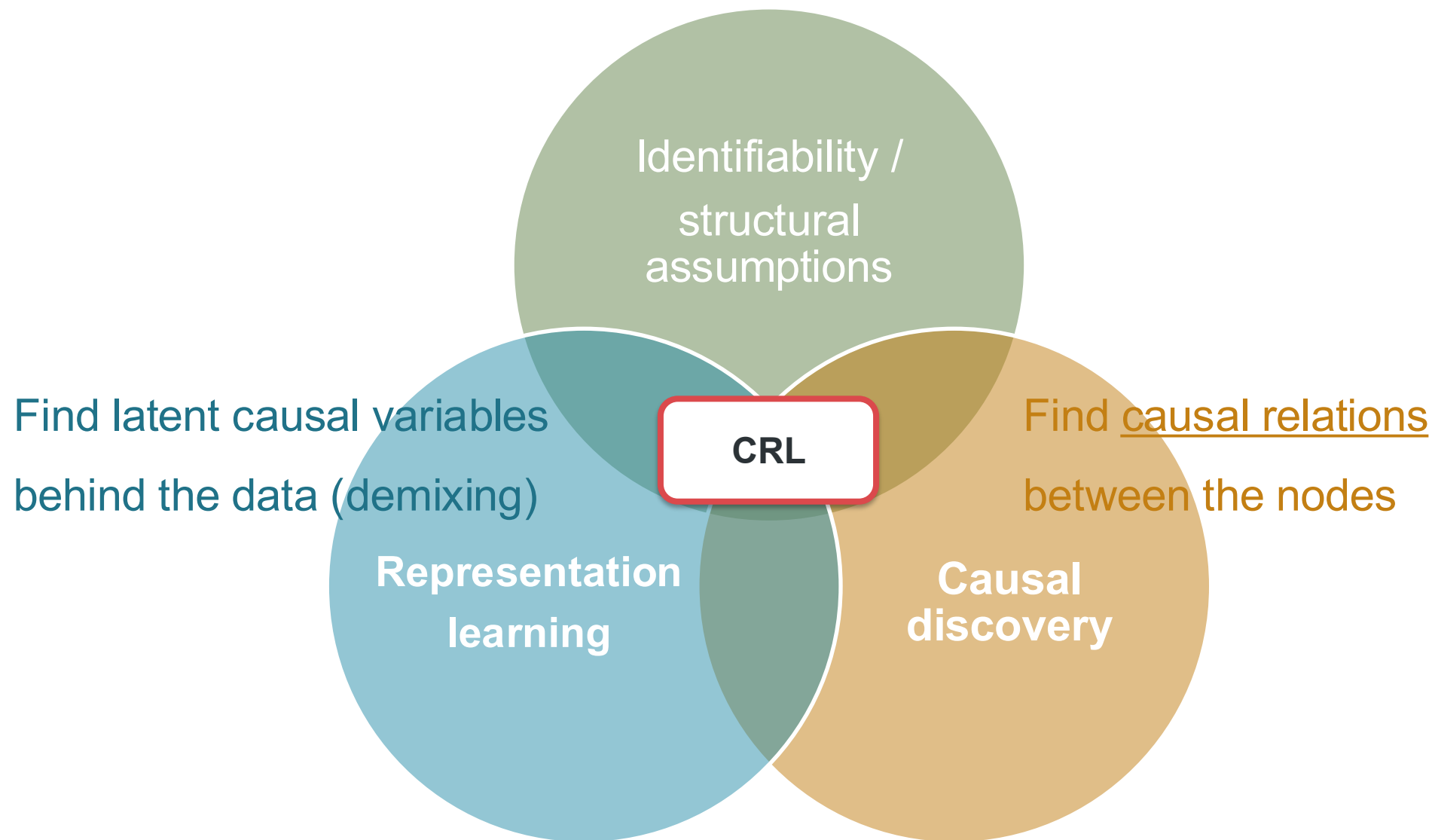
Causal relations

Why learn this?

- 1 Intervention & decision making
- 2 Robust prediction / Generalization
- 3 Interpret hidden abstract mechanisms

Goal: Recover the hidden causal variables and their causal relations behind entangled observations, for interpretation, intervention, generalization,...

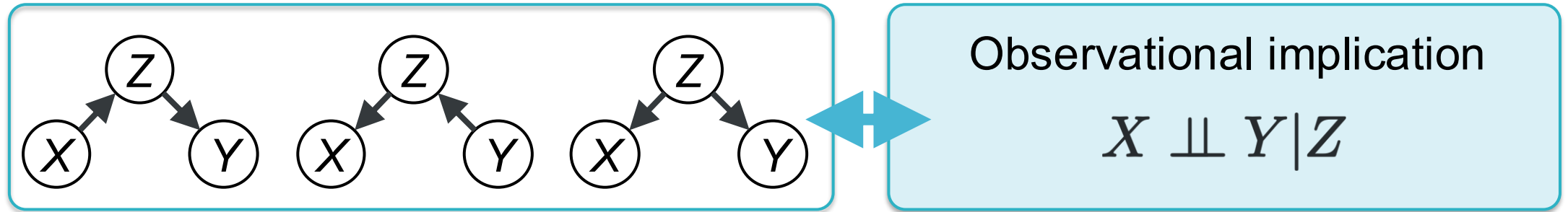
因果表現学習を構成する3つの要素



Key : structural assumptions make the joint problem identifiable

因果探索にも識別性の問題がある

Observational data may not identify a unique DAG



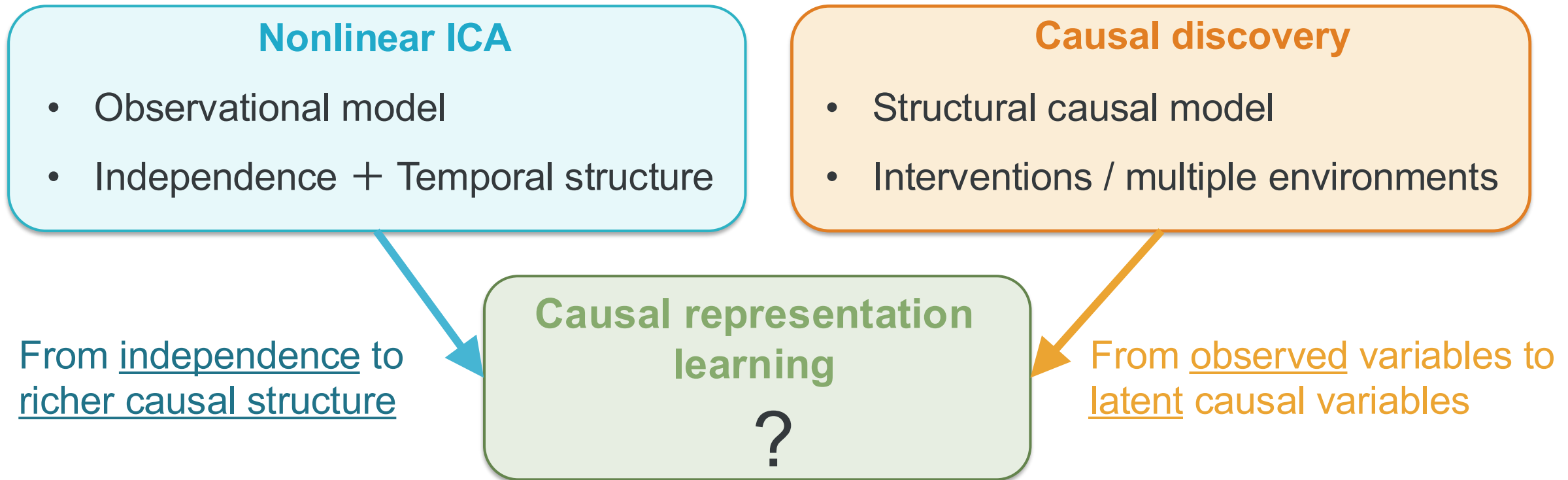
Same CI relation → Markov equivalence

How can the equivalence be broken? → **Additional assumptions**

- **Functional / noise asymmetry**; LiNGAM: linear + non-Gaussian noise;
ANM: additive noise + independence
- **Interventions / distribution changes**; Extra regimes can orient edges that are observationally indistinguishable.

Key point: Additional structure can break observational equivalence

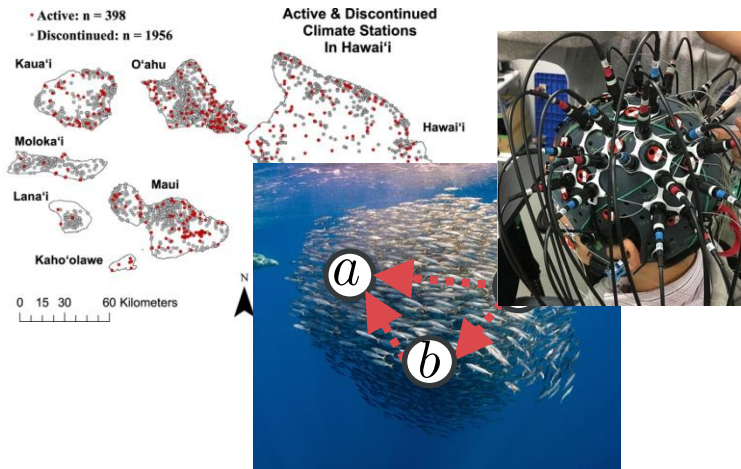
非線形ICAと因果探索を組み合わせればよいのか？



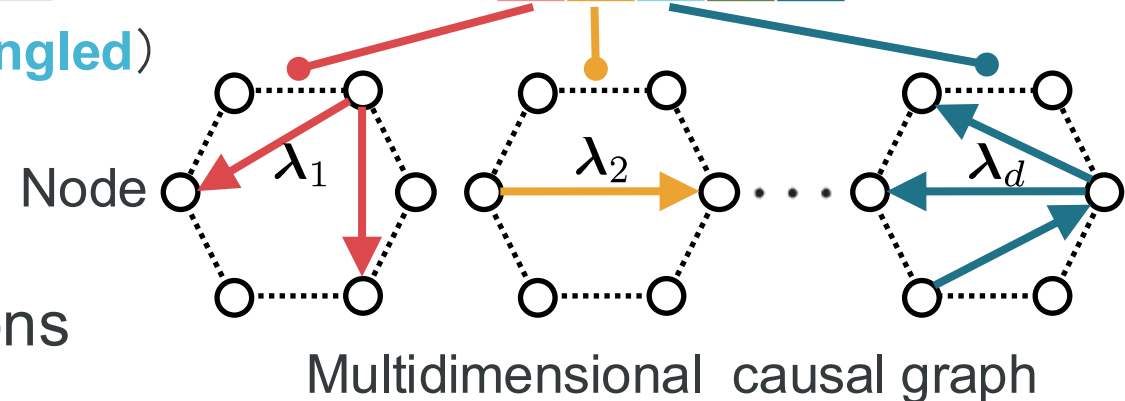
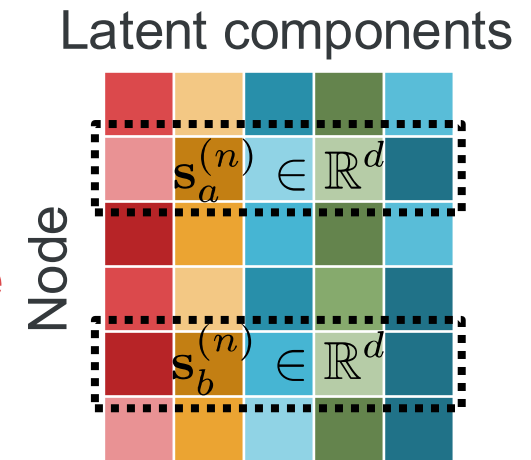
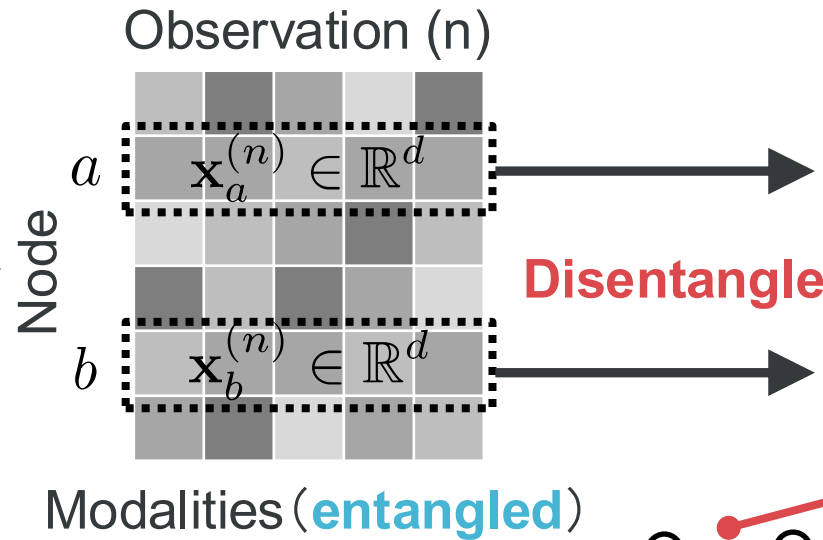
Key: CRL inherits the identifiability issues from both sides — and introduces additional ones. 独立成分を取り出したあとに因果探索すればよいわけではない。観測に既知のグループ構造がある場合の2つの識別可能な因果表現学習法を紹介

1: Entangled multimodal sensor networks, grouped as locations

Setting: Sensor modalities are entangled within each spatial location



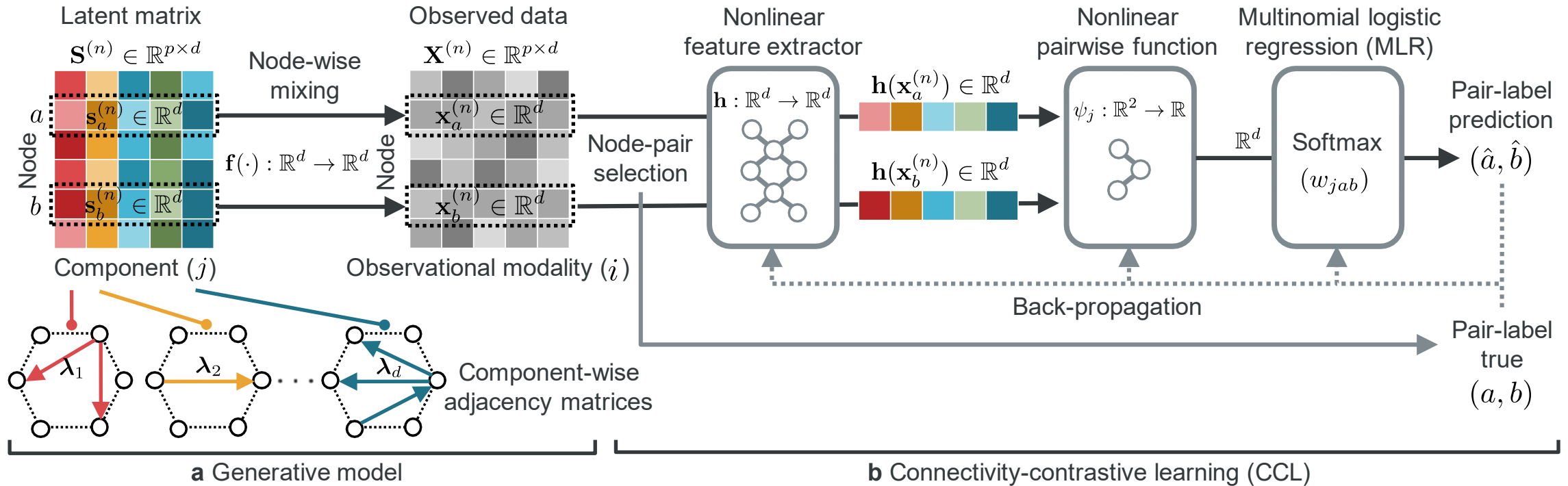
(node $\times$ feature) $\times$ sample (n)



Goal: Disentangle latent modalities and estimate their causal relations across locations

Identifiability assumptions: asymmetric interactions, component-specific causal graphs, sufficient heterogeneity across components

Self-supervised framework: Connectivity-contrastive learning (CCL)



$$p(\{s_{aj}\}_{a \in \mathcal{V}}) \propto \prod_{a \in \mathcal{V}} \bar{q}_{aj}(s_{aj}) \prod_{(a,b) \in \mathcal{E}} \exp(\lambda_{jab} \phi_j(s_{aj}, s_{bj}))$$

MI + Asymmetry

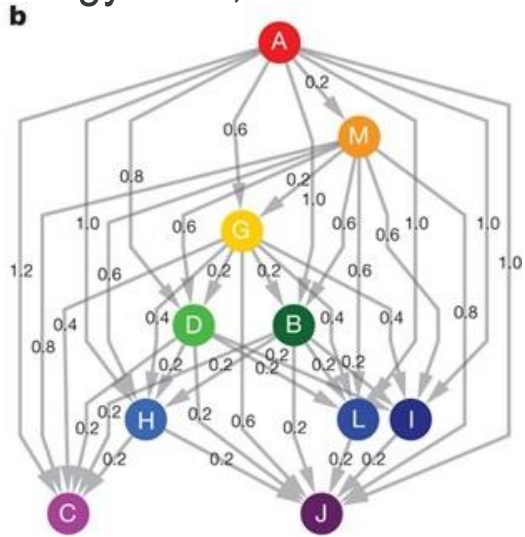
$$\text{Loss: } L = - \sum_n \sum_{(a,b) \in \mathcal{E}} \frac{\exp(\sum_{j=1}^d z_{jab}(\mathbf{x}_a^{(n)}, \mathbf{x}_b^{(n)}))}{\sum_{(l,m) \in \mathcal{E}} \exp(\sum_{j=1}^d z_{jlm}(\mathbf{x}_a^{(n)}, \mathbf{x}_b^{(n)}))}$$

Self-supervised estimation: predict edge-labels (a, b) of paired-observations $(\mathbf{x}_a^{(n)}, \mathbf{x}_b^{(n)})$

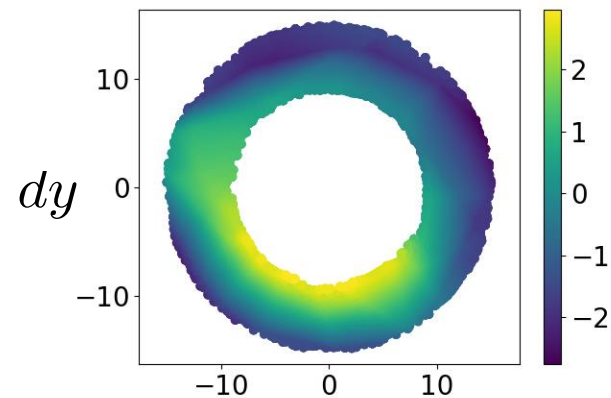
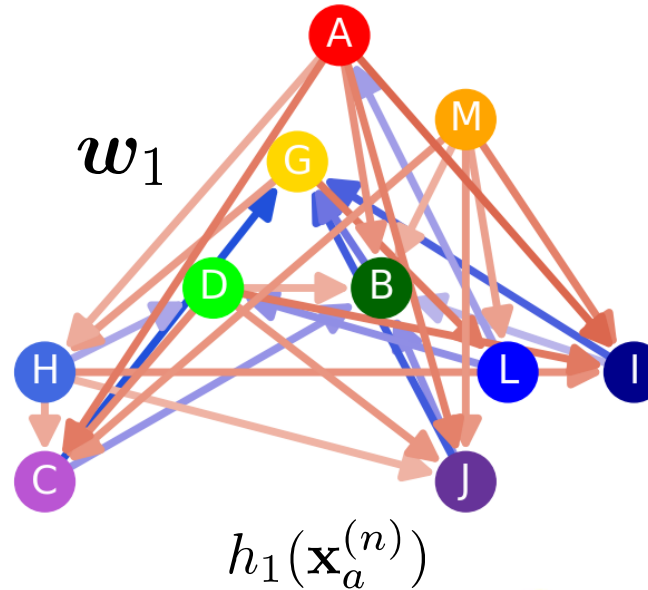
Theorem: $\mathbf{h}(\mathbf{x}_a^{(n)}) = \mathbf{s}_a^{(n)}$ and $\mathbf{w}_j = \lambda_j$ (up to some indeterminacy)

実データへの適用:ハトの群れ

Nagy et al., *Nature* 2010

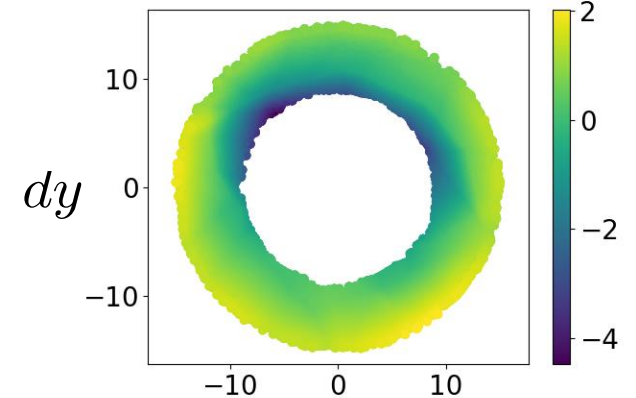
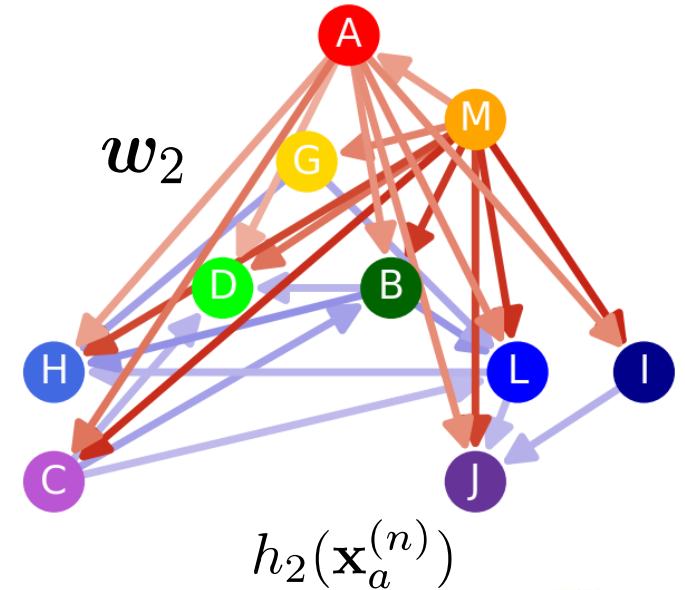


Component 1



向きの変化に関する
相互作用

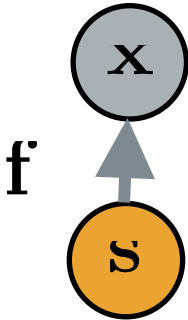
Component 2



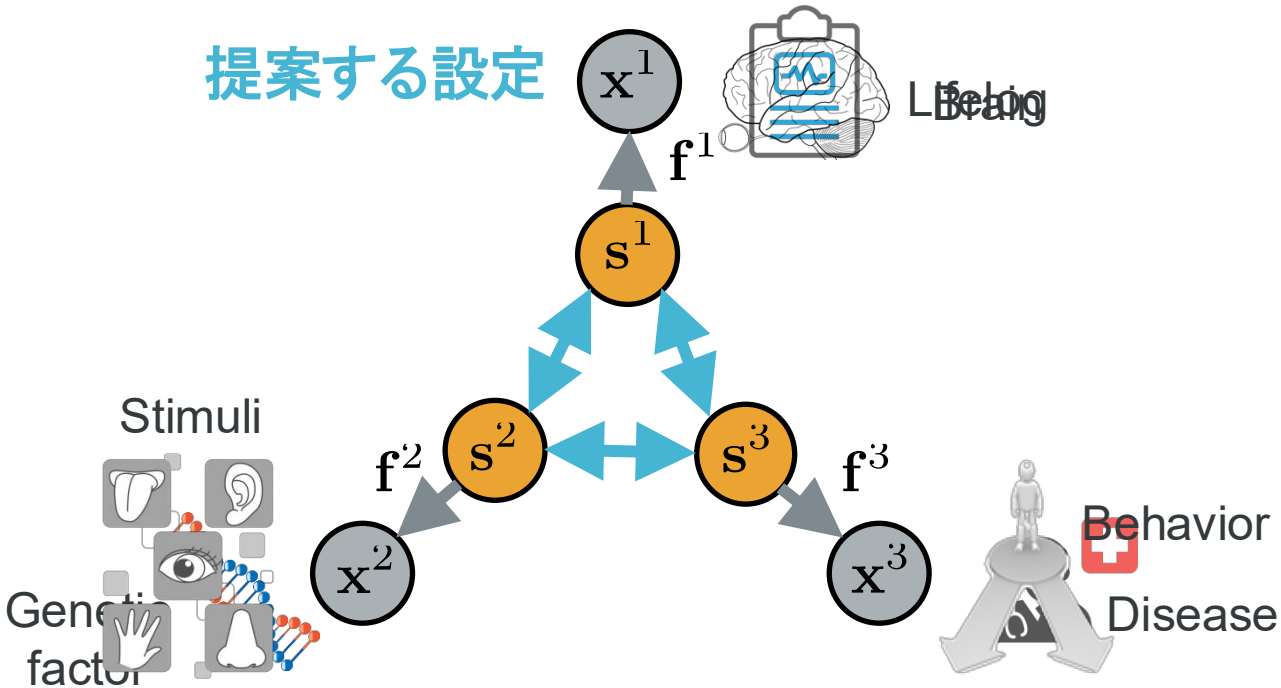
速度の変化に関する
相互作用

2: 観測変数のグループ構造を利用した因果表現学習

一般的な設定



提案する設定



観測

$$\mathbf{x} = \mathbf{f}(\mathbf{s})$$

$$\mathbf{x} = [\mathbf{x}^1, \dots, \mathbf{x}^M] = [\mathbf{f}^1(\mathbf{s}^1), \dots, \mathbf{f}^M(\mathbf{s}^M)]$$

因果構造

$$p(\mathbf{s}) = \prod_{a \in \mathcal{V}} p_a(s_a | \text{pa}(s_a)) \quad p(\mathbf{s}) \propto \underbrace{\left[\prod_{m \in \mathcal{M}} \exp(\bar{\phi}^m(\mathbf{s}^m)) \right]}_{\text{グループ内}} \underbrace{\prod_{m \neq m'} \prod_{(a,b) \in \mathcal{V}_s^m \times \mathcal{V}_s^{m'}} \exp(\lambda_{ab}^{mm'} \phi(s_a^m, s_b^{m'}))}_{\text{グループ間}}$$

識別性

仮定が弱すぎるため識別不能 グループ間の因果モデルと非対称性を仮定→識別可

推定の枠組み: グループ間の依存を利用する

1. グループごとにサンプルをシャッフルし, グループ間の依存を壊した人工データを作る

$$\underline{\mathbf{x} = [\mathbf{x}^1, \dots, \mathbf{x}^M]} \quad \text{vs.} \quad \underline{\mathbf{x}^* = [\mathbf{x}^{1*}, \dots, \mathbf{x}^{M*}]}$$

$y = 1$: 元のデータ

$y = 0$: シャッフルデータ(グループ間の依存を破壊)

2. データのラベル y を予測するロジスティック回帰分類器を学習する

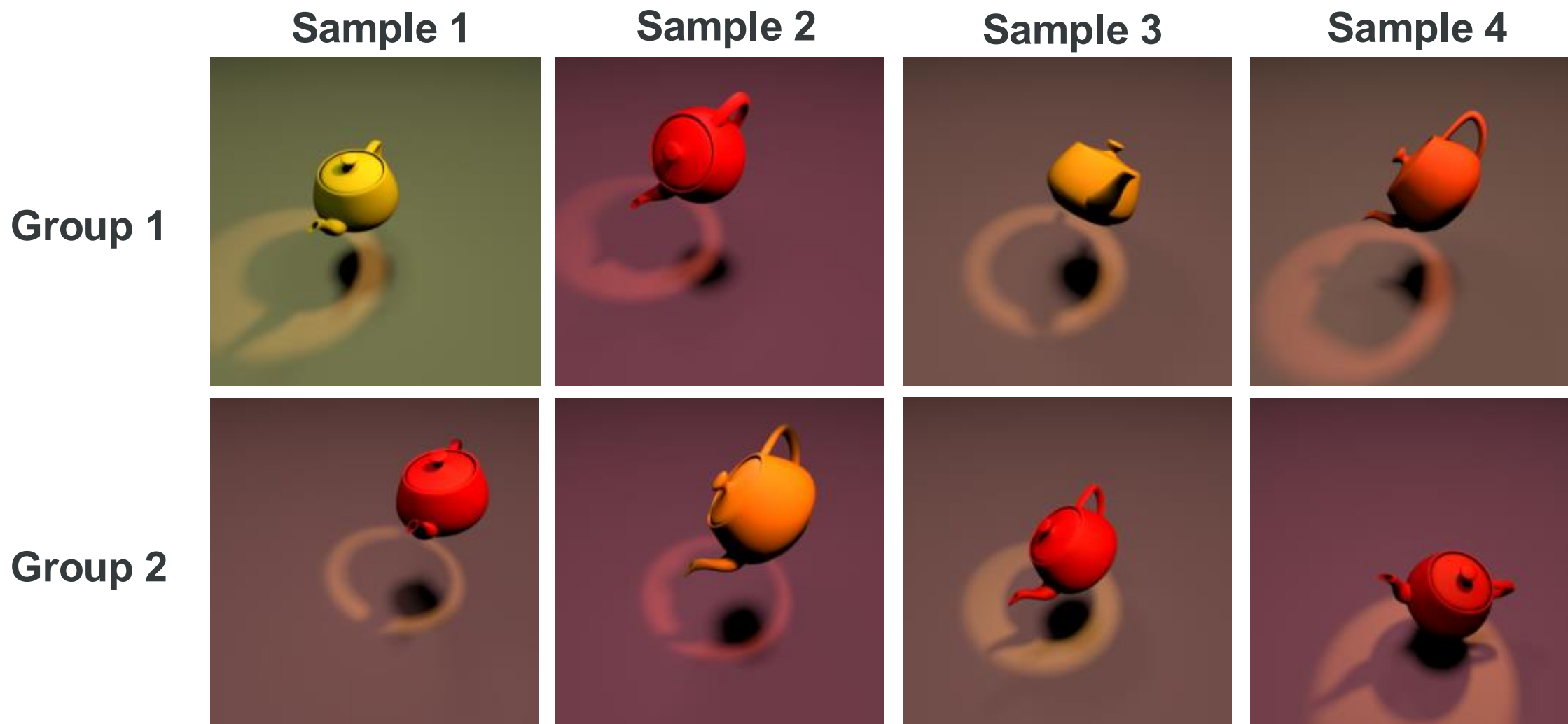
$$\text{Cross-entropy loss: } L = - \sum_n y_n \log(\hat{y}_n) + (1 - y_n) \log(1 - \hat{y}_n)$$

$$r(\mathbf{x}) = c + \sum_{m \in \mathcal{M}} \bar{\psi}^m(\mathbf{h}^m(\mathbf{x}^m)) + \sum_{m \neq m'} \sum_{(a,b) \in \mathcal{V}_S^m \times \mathcal{V}_S^{m'}} w_{ab}^{mm'} \psi(h_a^m(\mathbf{x}^m), h_b^{m'}(\mathbf{x}^{m'}))$$

理論的保証: 一定の条件のもとで, 各グループの潜在変数 $\mathbf{s}^m = \mathbf{h}^m(\mathbf{x}^m)$ と

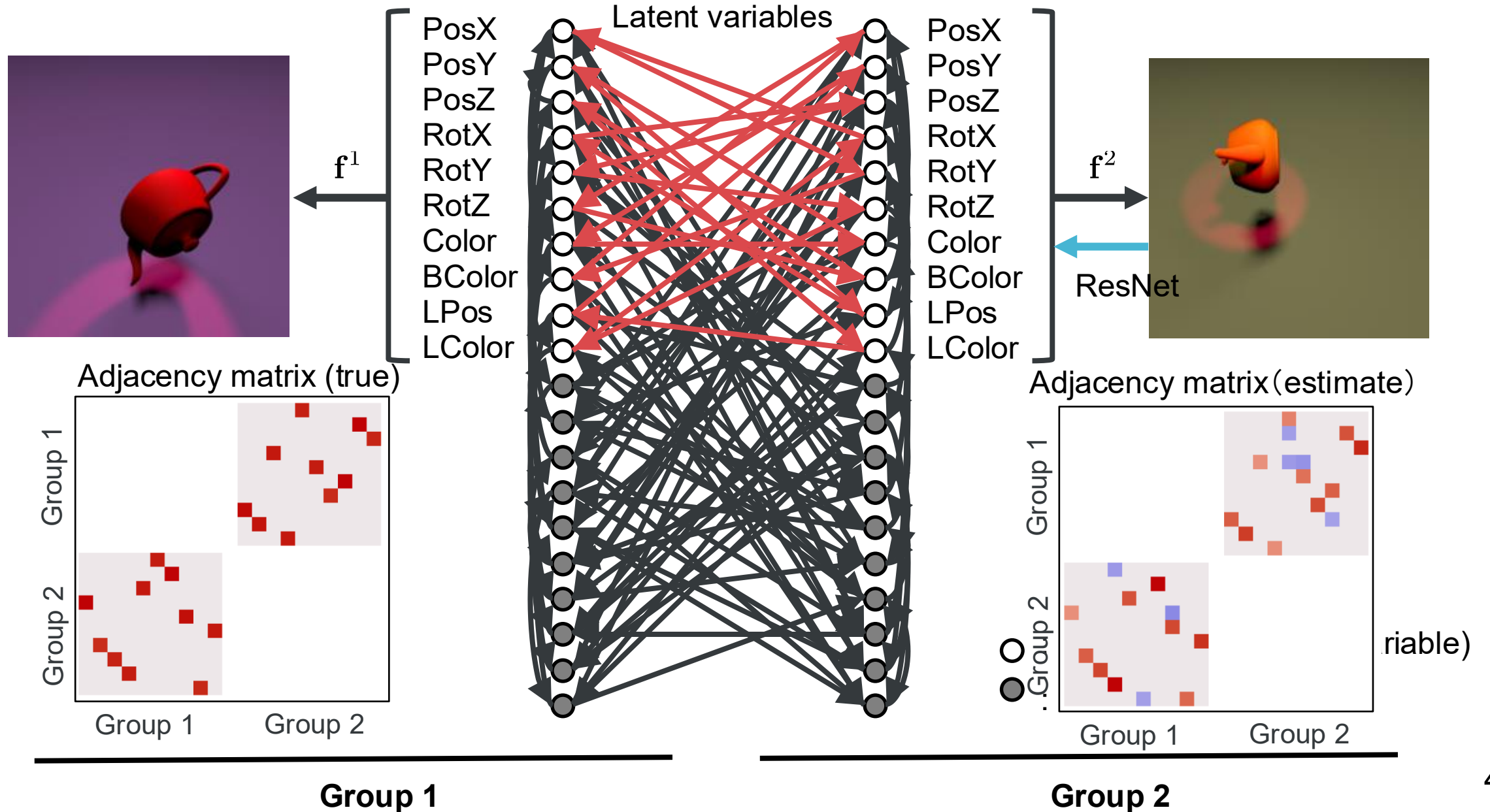
グループ間の因果関係 $(\lambda_{ab}^{mm'}) = (w_{ab}^{mm'})$ を, 許容される不定性を除いて識別できる

高次元画像データへの適用



目的: 画像を生成する(各グループ)10個の潜在因子と, それらの因果関係を推定

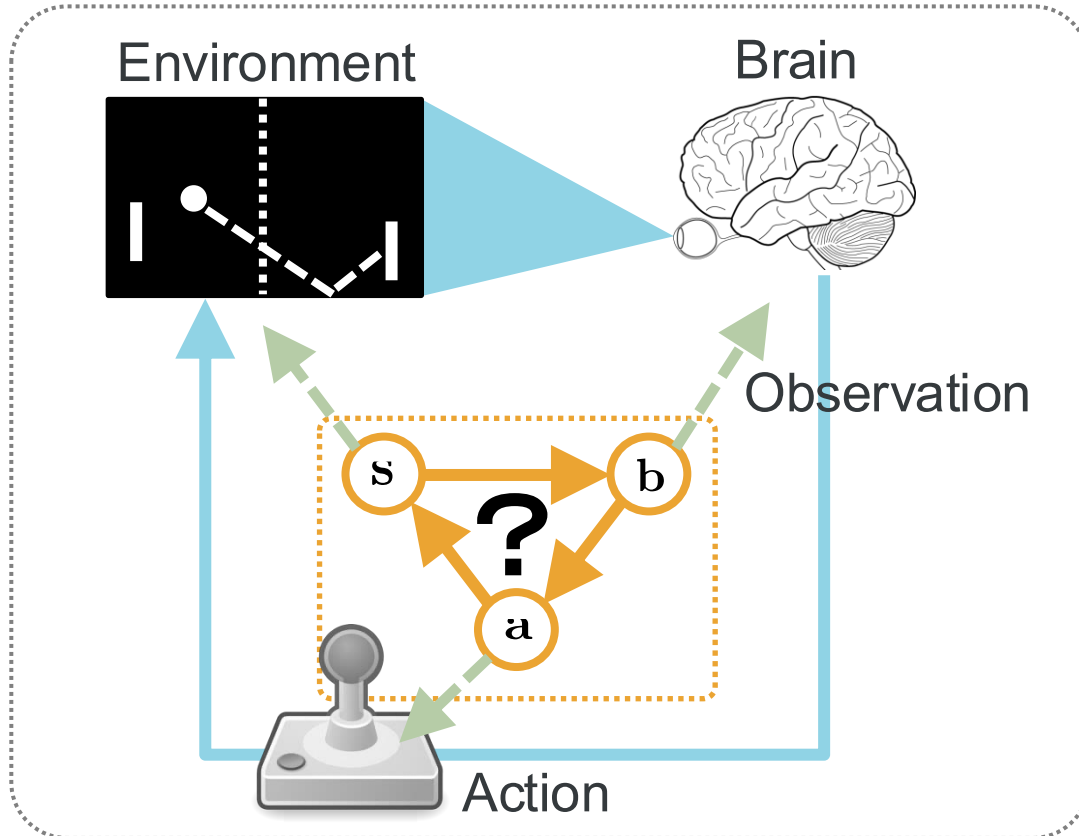
潜在因果モデル: 正解と推定結果



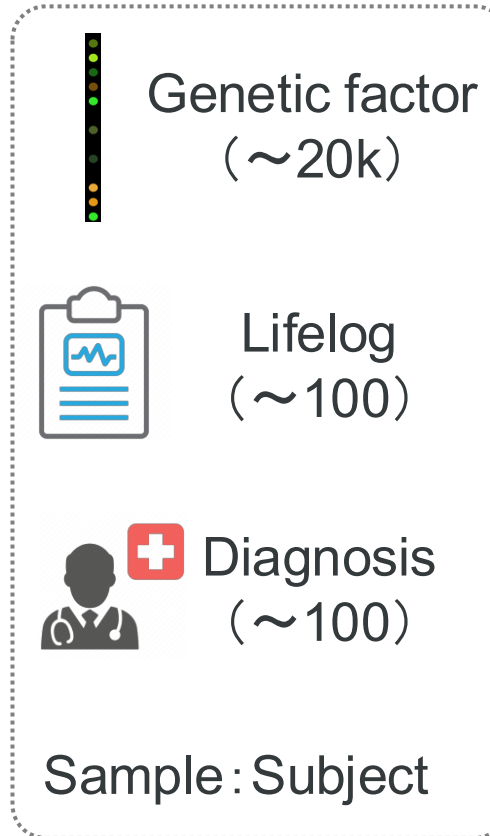
マルチモーダルデータへの応用可能性

$$[\mathbf{x}^1, \dots, \mathbf{x}^M] = [\mathbf{f}^1(\mathbf{s}^1), \dots, \mathbf{f}^M(\mathbf{s}^M)]$$

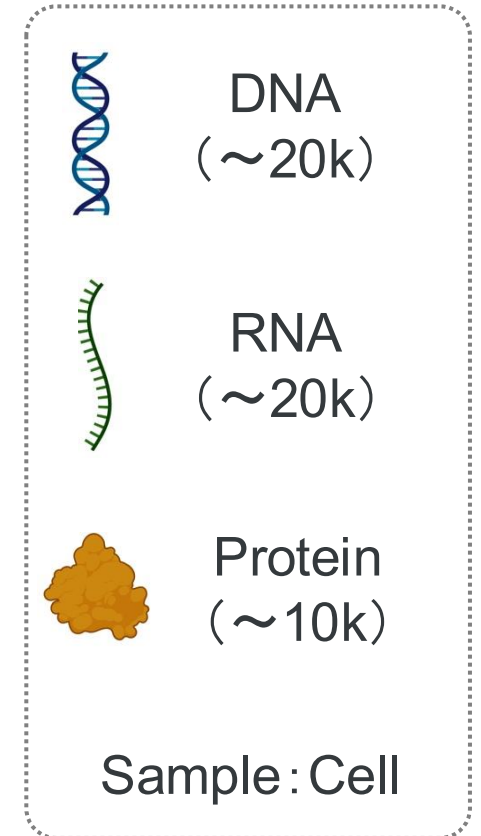
Multimodal Brain imaging



Medical



Multomics



グループ間に潜む因果関係を推定

表現学習に残された課題

より豊かな潜在モデル: 巡回, 潜在交絡, 時空間的因果, 高次相互作用

より弱い構造情報: 構造の種類や境界が未知・不完全な場合, 介入・レジーム・補助ラベルの欠損やノイズ

統計的推論: 不確実性の定量化と有限標本理論

モデル検証: 仮定違反の評価, 汎化可能性, 実データ評価, 介入・予測への応用, 解釈・説明可能性

まとめ

1

表現を学習できることと、真の潜在変数を復元できることは異なる

2

識別性には、データ生成過程に関する追加構造が必要である

3

対照学習は、その構造を学習信号に変換する方法である

表現学習 = 仮定 + 識別性 + 推定法