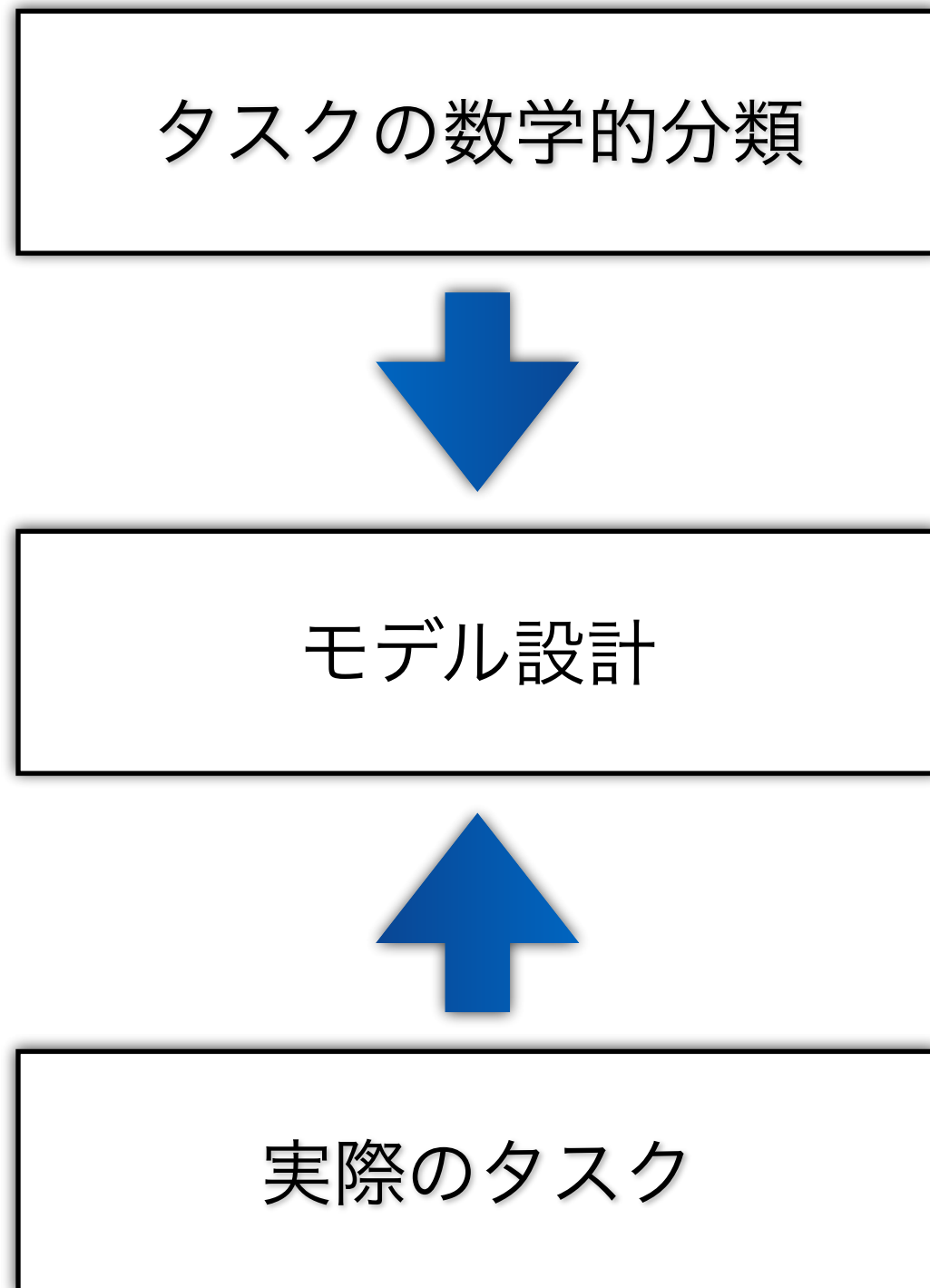


群対称性を用いた 深層学習

三内顕義 理研AIP/慶応大学

タスクにfitしたモデルの設計



対称性



CNN、RNN、GNN etc..

画像、時系列データ、グラフ etc..

今回の研究対象

- ・ 対称性（群を元にした）を持ったタスクに対する深層ニューラルネットモデルの研究
- ・ タスクが対称性を持つという事前知識の下、特別な深層ニューラルネットを使うことで少数サンプルでもワークする(Deep sets, NeurIPS2018)
- ・ 基本的に考える群は対称群 (S_m) がベースとなる

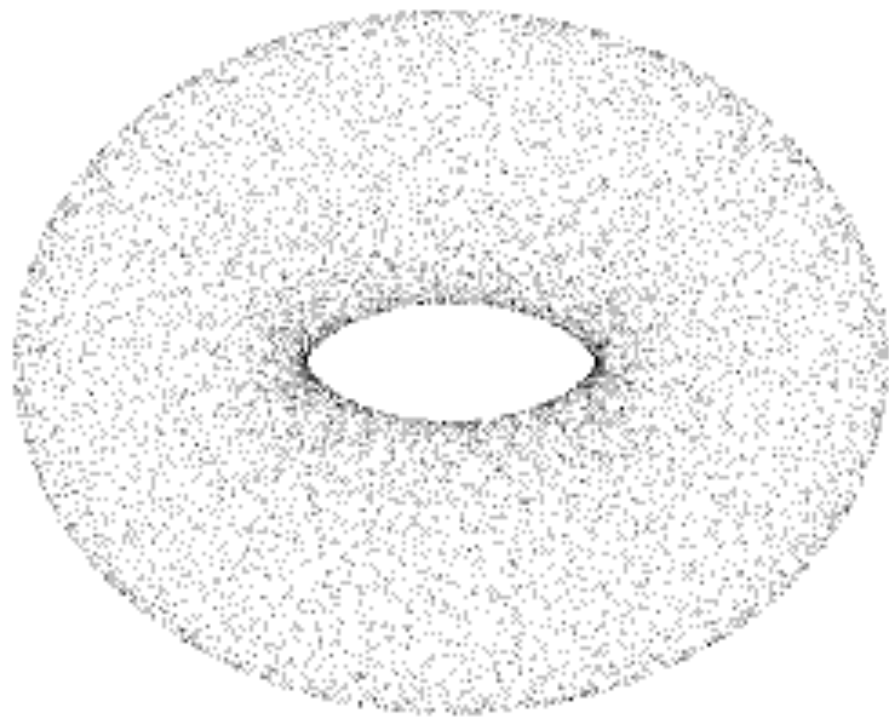
同変性のあるタスク

- 仲間外れ探し (Deep sets NeurlPS 2018)



- 列ごとに一人だけ仲間外れがいる(他は二つの性質を共有)
- 写真の順を変えると答えもそれに応じて変わる

不変性のあるタスク



wikipedia[点群]より

- ・ 入力がポイントクラウドである分類問題
- ・ 点たちに順序づけをする必要があるが、その順番を変えても答えは変わらない

群論の復習

群とは

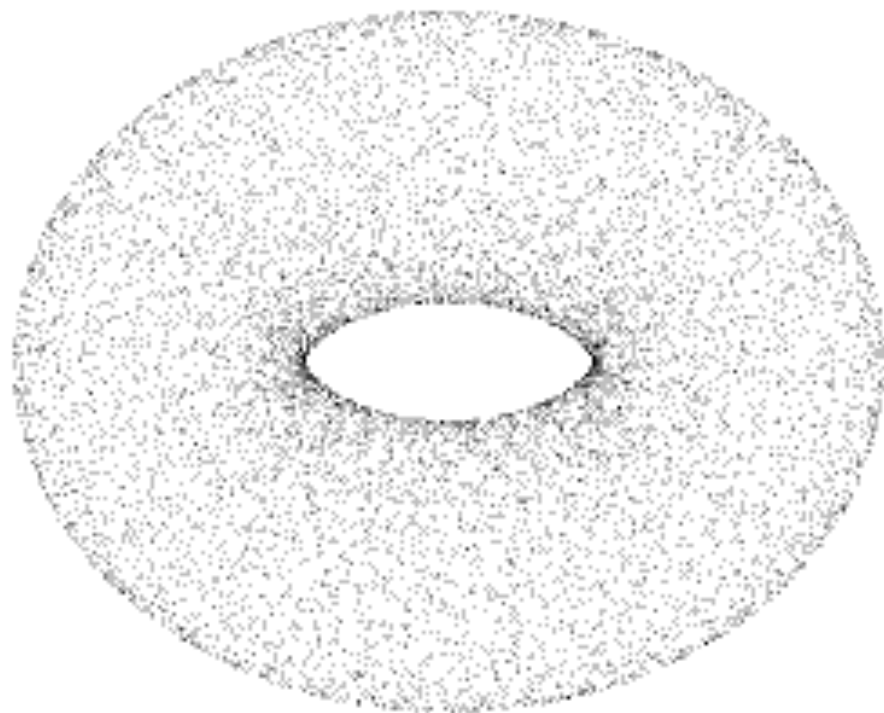
定義 集合 G は二項演算 $\cdot$ を持つ、つまり任意の $a, b \in G$ に対し $a \cdot b \in G$ が定まるとする。これが次の条件を満たす時、 $(G, \cdot)$ を群という。

(1) 任意の $a, b, c \in G$ に対して $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

(2) 単位元と呼ばれる元 $e \in G$ が存在し、任意の $a \in G$ に対して $a \cdot e = e \cdot a = a$.

(3) 任意の $a \in G$ に対して $a^{-1} \in G$ が存在し $a \cdot a^{-1} = a^{-1} \cdot a = e$.

群の例



- ・ 点の順序づけの置換群 S_M
- ・ 3次元空間の回転群 $SO(3)$
- ・ 正則行列のなす群 $GL(\mathbb{R})$

wikipedia[点群]より

対称群 S_m

- ・ S_M の元は M 点集合 $\{1, \dots, M\}$ の置換（全単射）を表す
- ・ $\cdot$ は写像の合成
- ・ S_M の元は対応表を使って表すことができる。例: $g_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$, $g_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$
- ・ g を S_M の元とする時、 $g(i)$ で i の行き先を表すことにする（上の例では $g_1(1) = 3$, $g_2(3) = 1$ など）
- ・ i と j を入れ替えるだけの置換（互換）を $(i\ j)$ と書くことにする

例: $(1\ 3) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$

作用とは

定義 G を群、 X を集合とする。 G が X に作用するとは、

(1) 任意の $a \in G$ と任意の $x \in X$ に対して $a(x) \in X$ が定まっていて、

(2) 単位元 e に対しては $e(x) = x$ であり、

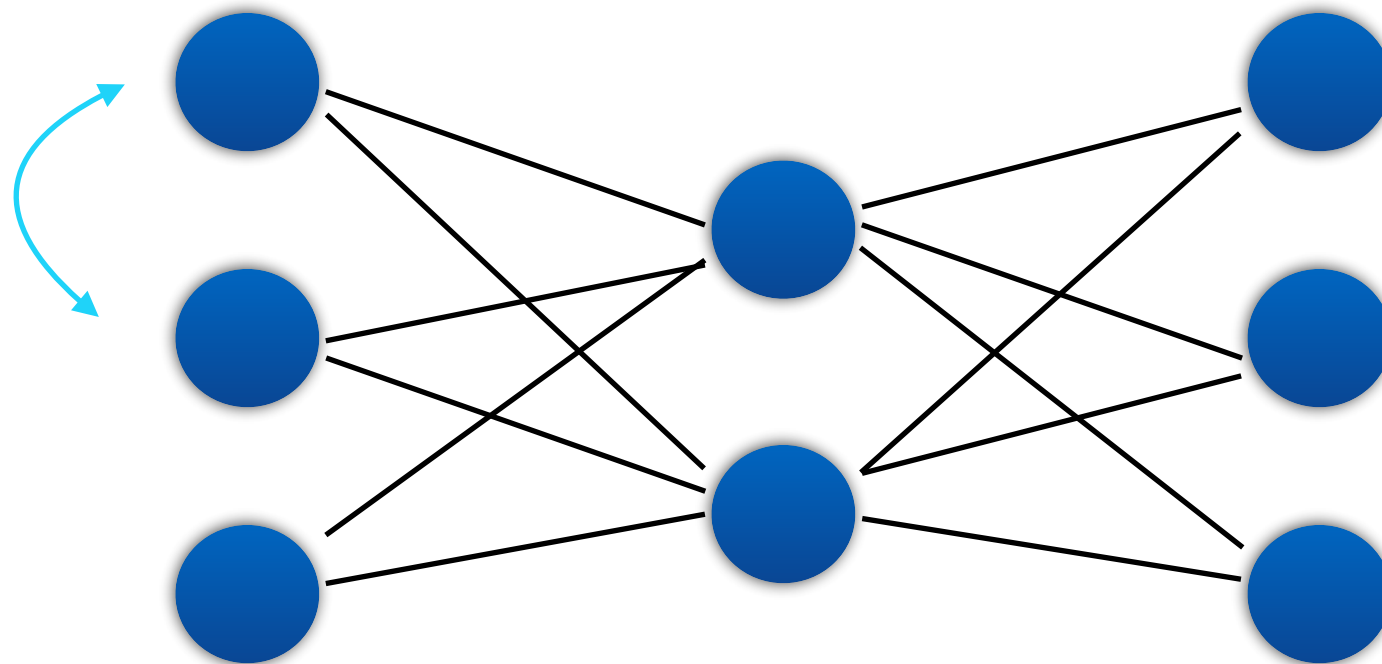
(3) 任意の $a, b \in G$ に対して $(a \cdot b)(x) = a(b(x))$ が成り立つことをいう。

・ 群 G が X に作用している時、 $G \curvearrowright X$ と書くことにする

ベクトル空間への自然な作用

ベクトル空間の標準基底 $e_1, \dots, e_m$ に対し、 $g \cdot e_i = e_{g(i)}$ とし、これを線形に拡張することで自然な作用が定義される。つまりベクトル $(x_1, \dots, x_M)$ と $g \in G$ に対し、 $g \cdot (x_1, \dots, x_M) = (x_{g^{-1}(1)}, \dots, x_{g^{-1}(M)})$ とする。

DNNの各層に対応するベクトル空間の場合、これはnodeの入れ替えに対応する



不変/同変な関数

定義 (1) 関数 $f: \mathbb{R}^M \rightarrow \mathbb{R}$ が置換不変とは

$f(g \cdot (x_1, \dots, x_M)) = f(x_1, \dots, x_M)$ が全ての置換 $g \in S_M$ について成り立つことを言う.

(2) 関数 $f = (f_1, \dots, f_M): \mathbb{R}^M \rightarrow \mathbb{R}^M$ が置換同変とは

$f(g \cdot (x_1, \dots, x_M)) = g \cdot [f_1(x_1, \dots, x_M), \dots, f_M(x_1, \dots, x_M)]^T$ が全ての置換 $g \in S_m$ について成り立つことを言う.

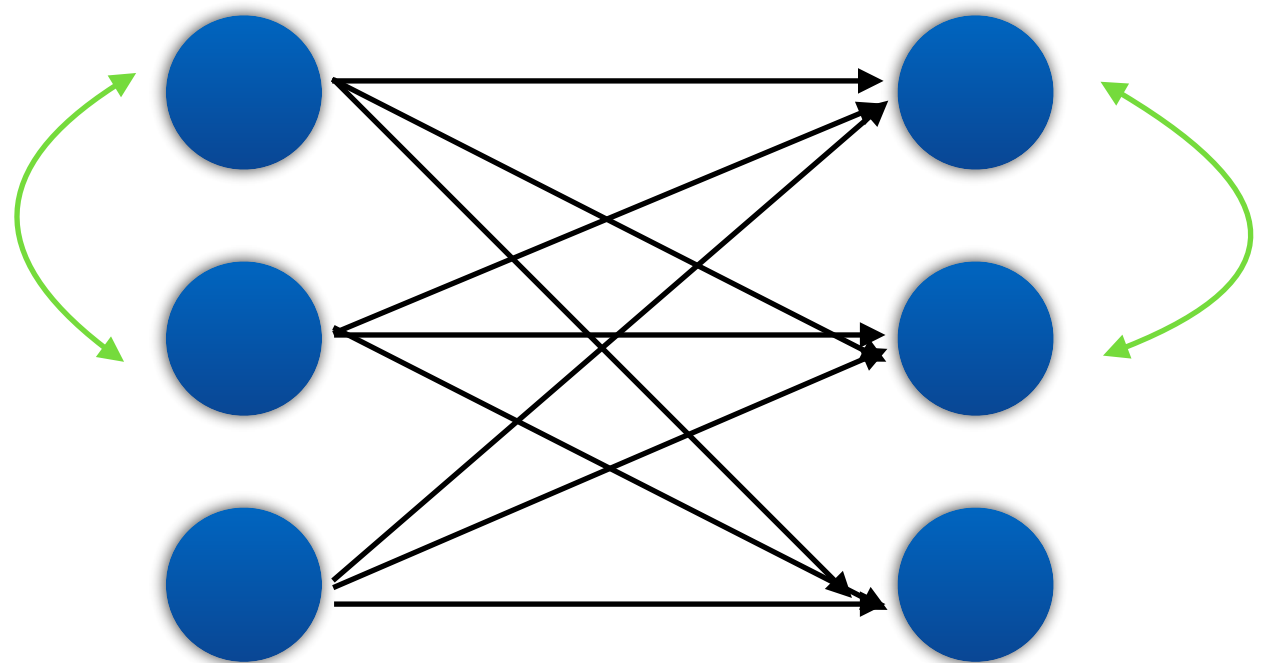
置換同変な ニューラルネット

置換同変な層

Lemma3(Deep sets): 関数 $f_{\Theta}(x) = \sigma(\Theta x)$ ($\Theta \in \mathbb{R}^{M \times M}$) が置換同変になるための必要十分条件は Θ が以下の形に書けることである

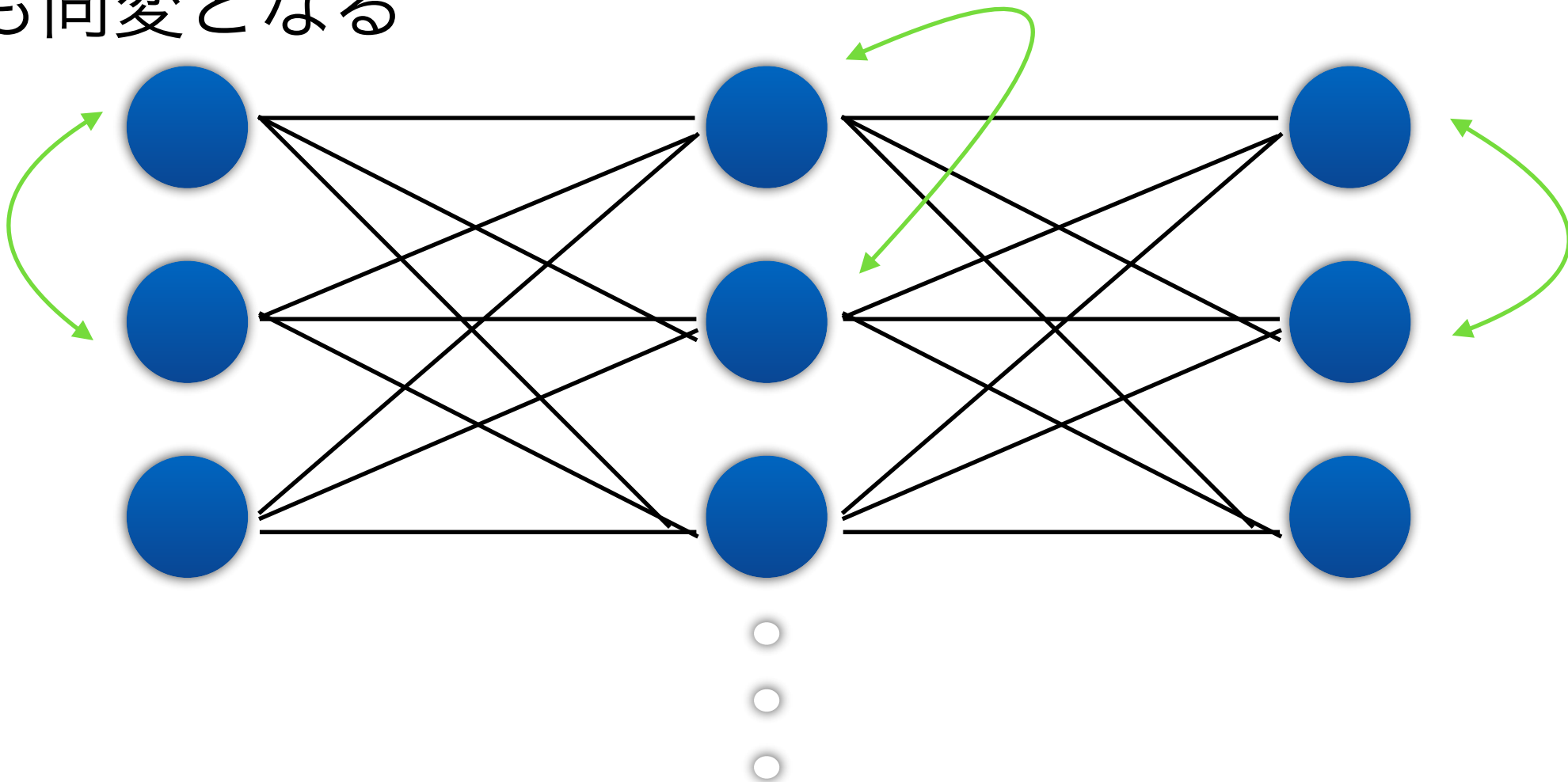
$$\Theta = \lambda I + \gamma(\mathbf{1}\mathbf{1}^T) \quad \lambda, \gamma \in \mathbb{R} \quad \mathbf{1} = [\mathbf{1}, \dots, \mathbf{1}]^T \in \mathbb{R}^M$$

- 各層の変換が同変を仮定すると行列が2-パラメータで具体的に書ける



同変深層ニューラルネット

- ・ 同変な層（のコピー）を積み重ねることで深層にする
- ・ 各層の変換が同変であることから深層が誘導する関数も同変となる



同変深層学習

- ・ 同変な深層NNはlem3より具体的なパラメータを持つため、このパラメータに対するback propagationを行うことで学習ができる
- ・ 結果として得られる関数は同変な関数となる
- ・ 同じ構造の場合、通常の深層ニューラルネットに比べて各層のコピー一つあたり $M^2 \rightarrow 2$ まで圧縮できている
- ・ 仲間外れ探し（前述）においてはtest accuracyが6.3%(CNN+fully conn.)から75%程度まで改善（CNN+同変）

置換不変な ニューラルネット

Kolmogorov-Arnoldの表現定理

定理：連続関数 $f : [0, 1]^M \rightarrow \mathbb{R}$ が置換不変関数になるための必要十分条件は次の形で表現できることである

$$f(x_1, \dots, x_M) = \rho\left(\sum_{m=1}^M \phi(x_m)\right)$$

ここで $\rho : \mathbb{R}^{M+1} \rightarrow \mathbb{R}$ と $\phi : \mathbb{R} \rightarrow \mathbb{R}^{M+1}$ は連続関数で ϕ は f によらない。

- ・ ϕ は具体的に $\phi(x) = [1, x, x^2, \dots, x^M]^\top$ の形で書ける
- ・ この定理から ϕ を近似するパートと ρ を近似するパートで分けて NN を構成するのが Deep sets のアイデア

$f(x_1, \dots, x_M) = \rho\left(\sum_{m=1}^M \phi(x_m)\right)$ の計算例

(1) $f(x_1, x_2) = x_1 x_2 (x_1 + x_2 + 3)$ とする.

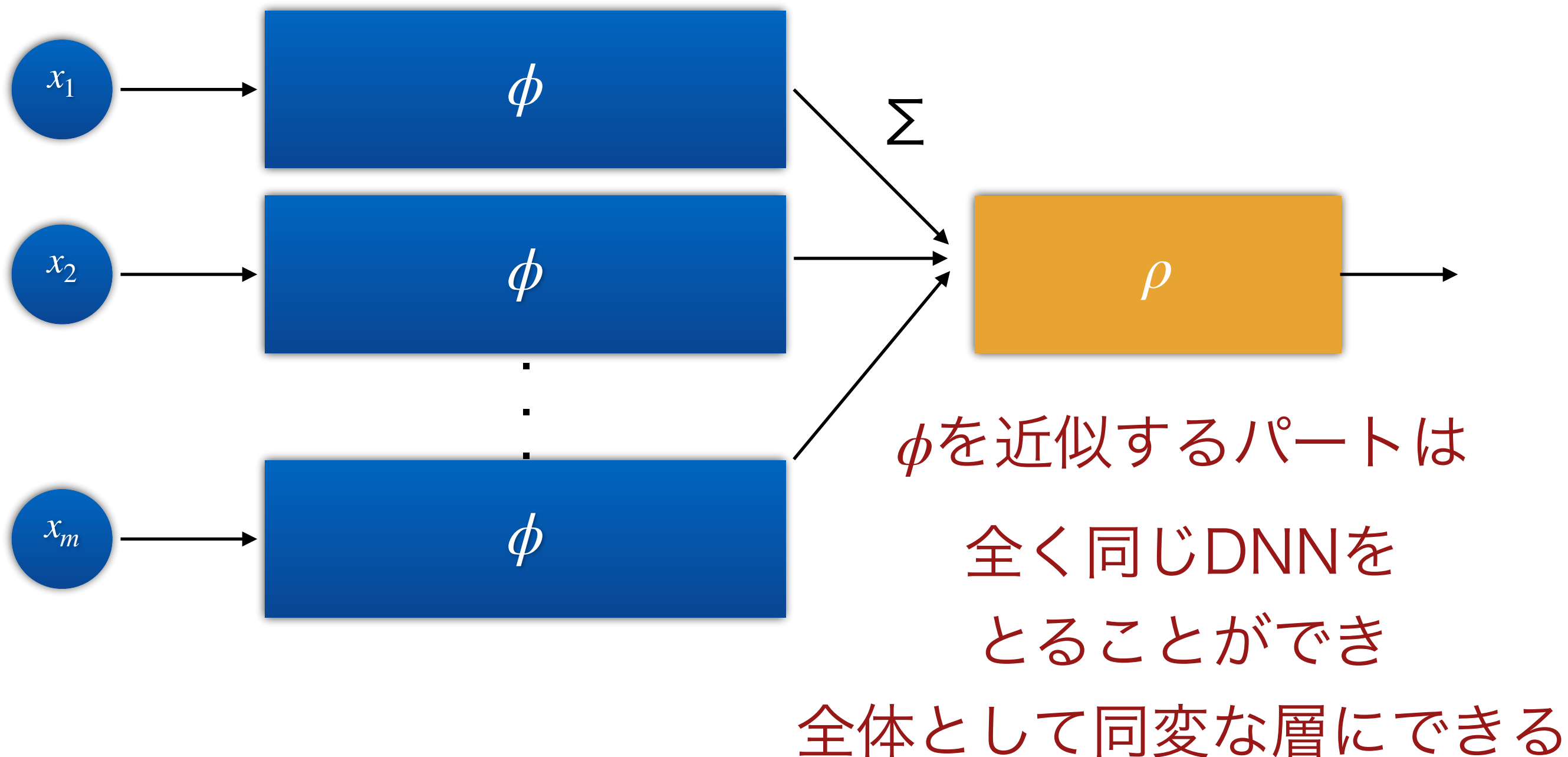
$\phi(x) = [x, x^2, x^3]$, $\rho([u, v, w]) = uv - w + 3(u^2 - v)/2$ と置
くと $\rho(\phi(x_1) + \phi(x_2)) = f(x_1, x_2)$ が成り立つ.

(2) $f(x_1, x_2, x_3) = x_1 x_2 x_3 + x_1 + x_2 + x_3$ とする。

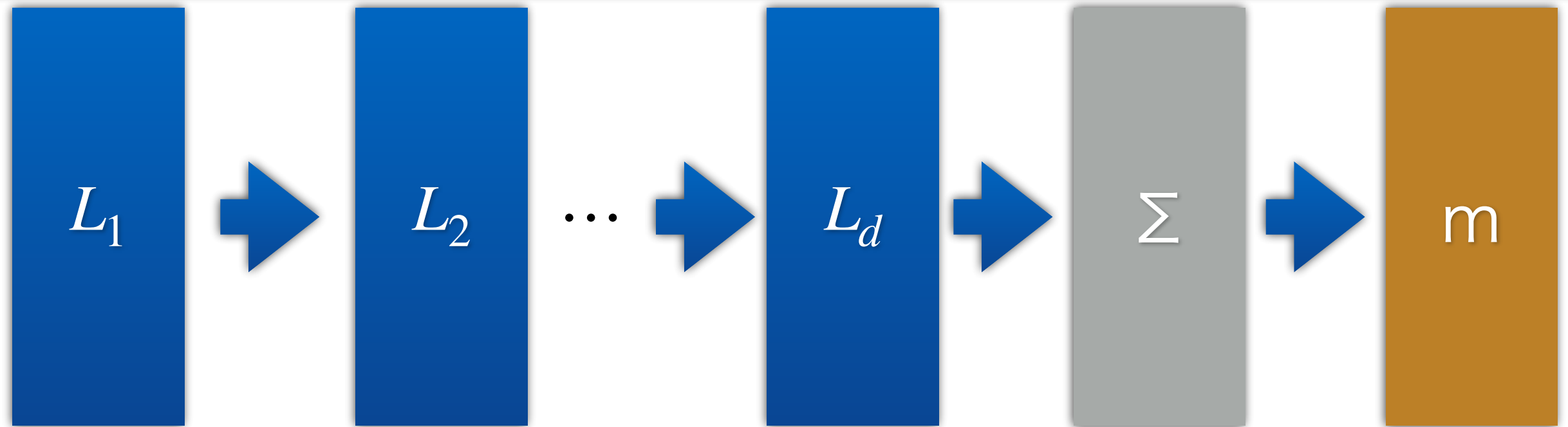
$\phi(x) = [x, x^2, x^3]$, $\rho(u, v, w) = (u^3 - 2w + 3uv)/6 + u$ と置
くと $\rho(\phi(x_1) + \phi(x_2) + \phi(x_3)) = f(x_1, x_2, x_3)$ が成り立つ.

$$f(x_1, \dots, x_M) = \rho\left(\sum_{i=1}^m \phi(x_i)\right) \text{ の }$$

深層ニューラルネット的解釈



不変深層ニューラルネット



- ・ L_i は同変な層
- ・ m はmulti layer perceptron
- ・ Σ は軌道の和を取る関数

不変深層学習

- ・ Lの部分はlem3のパラメータがありmの部分は普通のパラメータがある。これらのパラメータに対しback propagationで学習可能
- ・ 結果として得られる関数は置換不変 (Σ の部分でそうなる)
- ・ 今回はLの部分のパラメータのみ圧縮されている

まとめ

- ・ 不変のケースも同変のケースも各層に対応する行列を制限することでDNNが表現する関数自体を不変/同変にした。
- ・ パラメータが具体的に与えられているため、その範囲で学習ができる。
- ・ 不変ケースは数学的背景があるが、同変ケースは同変な関数を近似できる保証がないのでは？

Universal approximation theorem (Sannai-Takai-Matthieu)

定理2.2 G を S_M の部分群とする。 K を $\mathbb{R}^M$ のコンパクトな部分集合で G の作用で安定なものとする。任意の連続で G 同変な写像 $F : K \rightarrow \mathbb{R}^M$ と任意の $\varepsilon > 0$ に対し、（STMの意味での） G 不変深層ニューラルネット H が存在し、 $\|F - H\|_\infty < \varepsilon$ となる。

定義 （STMの意味での G 不変深層ニューラルネット）

各層のノードの集合（ベクトル空間の基底集合）への G 作用を自然な作用ではなく、任意の作用まで一般化する。他は今までと同じ。

証明の方針

- ・ 不変関数と同変関数の間にある対応を深層ニューラルネットレベルにリフトし、不変関数のUniversal approximationに帰着
- ・ 不変関数の時のZaheerらの定理 (S_M のケース)あるいはMaronらの定理 (有限群G)を適応する

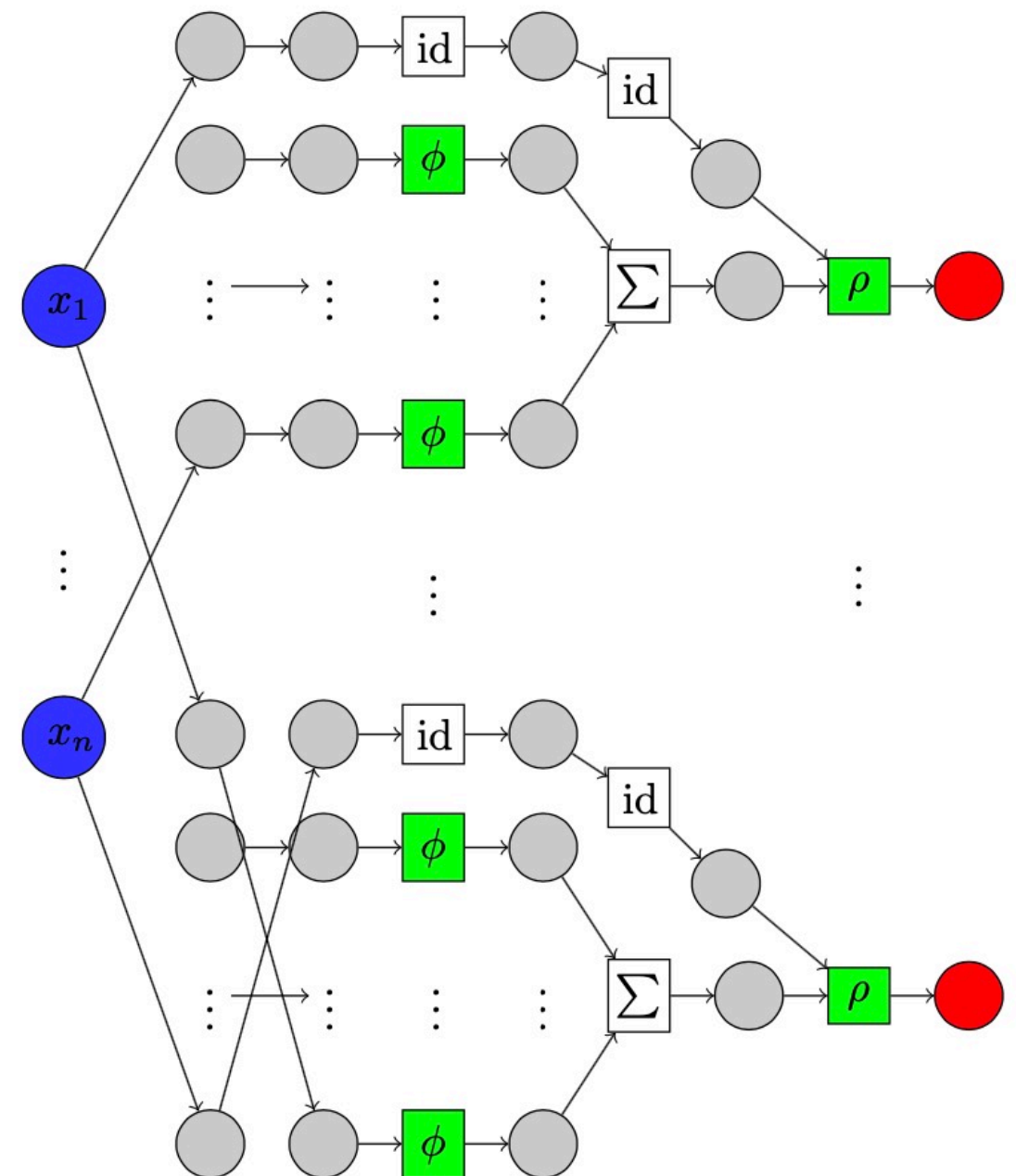
同変関数の表現定理

系3.1 (Sannai-Takai-Matthieu): 写像 $F: \mathbb{R}^M \rightarrow \mathbb{R}^M$ が S_M -同変になるための必要十分条件は $\text{Stab}(1)$ -不変関数 $f: \mathbb{R}^M \rightarrow \mathbb{R}$ が存在し、
 $F = (f \circ (1\ 1), f \circ (1\ 2), \dots, f \circ (1\ M))^T$ の形で書けること。ここで互換 $(i\ j)$ は $\mathbb{R}^M \rightarrow \mathbb{R}^M$ の線形写像とみなしている。

- ・ $\text{Stab}(1)$: 1 を固定する置換のなす S_M の部分群
- ・ S_M 同変な関数は互換と $\text{Stab}(1)$ 不変な関数の合成で書ける
- ・ 一般の部分群 G に対しても互換を別の表現に置き換えた表記がある

$F = (f \circ (1\ 1), f \circ (1\ 2), \dots, f \circ (1\ M))^T$ の 深層ニューラルネット表現

- 第1層で互換に対応する部分を表現
- コピーの一つ一つは f を近似している



部分群とテンソル表現

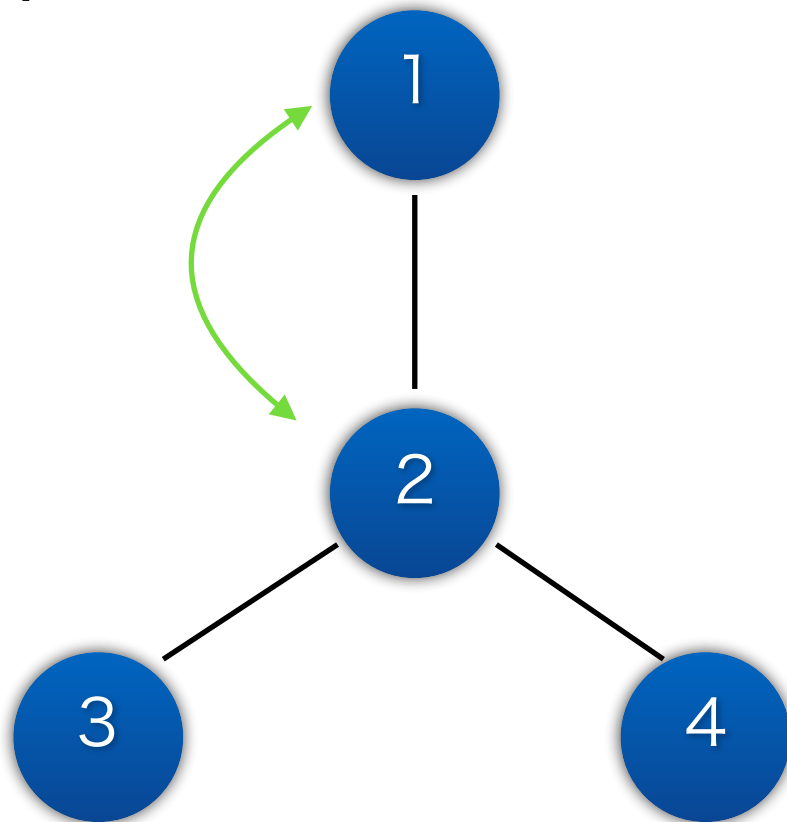
部分群とその作用

- ・ 群 G の部分集合 H が G の演算で群となる時、 H を G の部分群という
- ・ S_m の部分群 G を考える
- ・ G は S_m の部分集合であるから S_m の自然な作用が G の作用を誘導する。これを部分作用と呼ぶ。
- ・ 部分群を考えることでどのような状況を含むことができるか？

グラフのノード置換と隣接行列

グラフ（の同型類）と隣接行列の関係を考える

グラフとノードのラベリングを固定すると隣接行列が定まるが、ラベリングを置換すると隣接行列もそれに応じて変化する



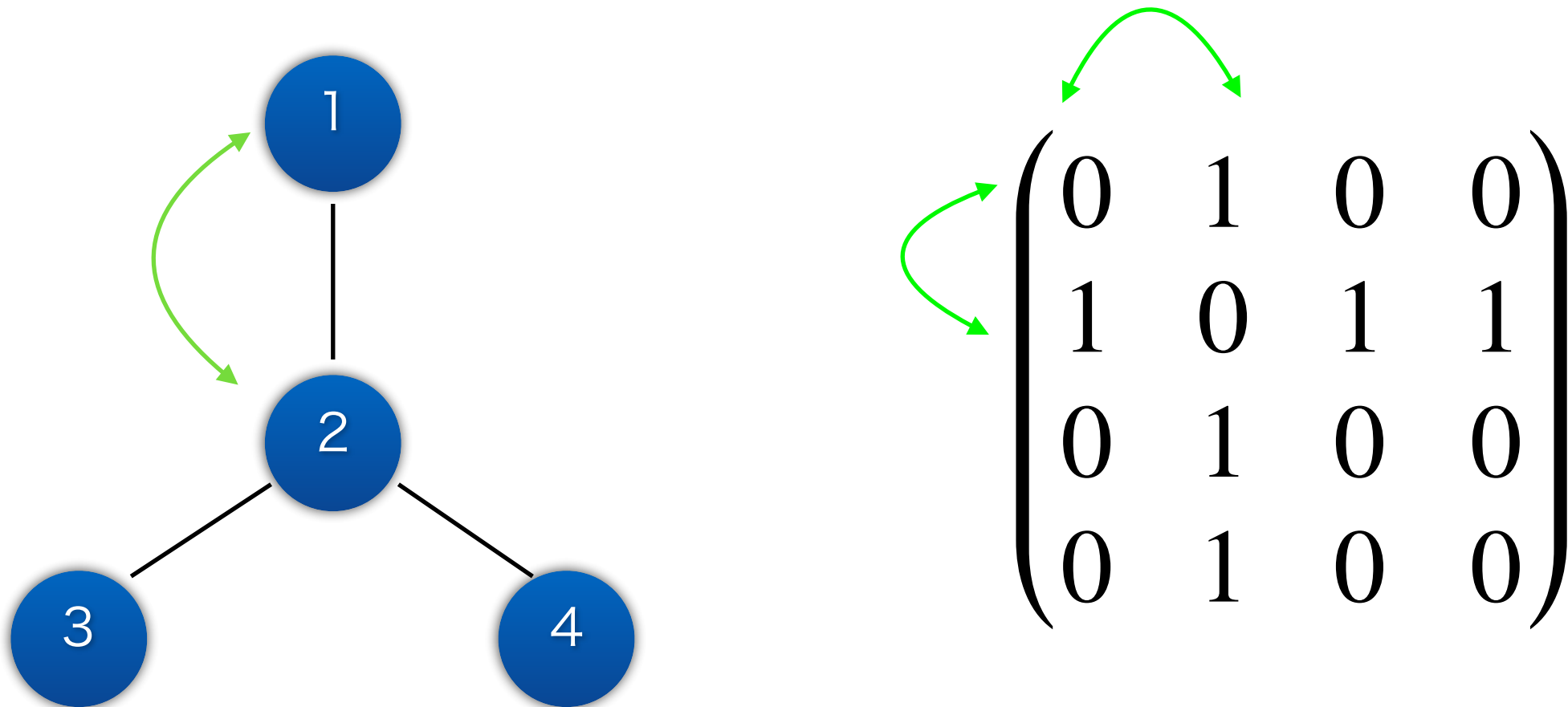
$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

グラフのノード置換と隣接行列

この関係はノードの置換群 S_M の $\mathbb{R}^{M^2} \cong GL(M, \mathbb{R})$ への非自明な作用を与える。

この作用は S_{M^2} の自然な作用の部分作用となっている

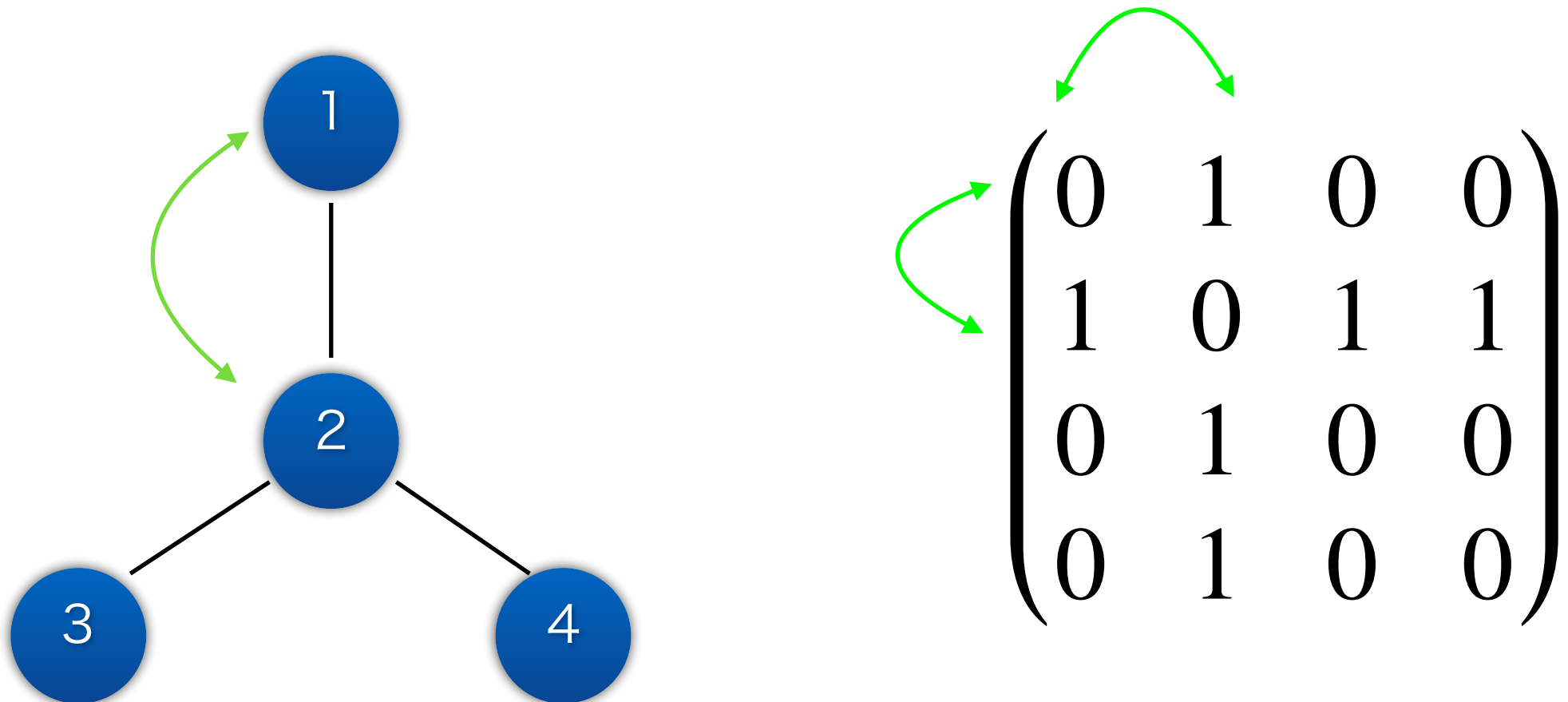
部分群による非自明な作用で役に立つ例となっている



ノード置換の 二階テンソル表現

この作用は行列の (i, j) 成分と $(\sigma(i), \sigma(j))$ 成分を置換する作用となっている

つまり行列を二階テンソルとみなした時、自然な作用を二階テンソルしたものと同値になっている



テンソルの定義

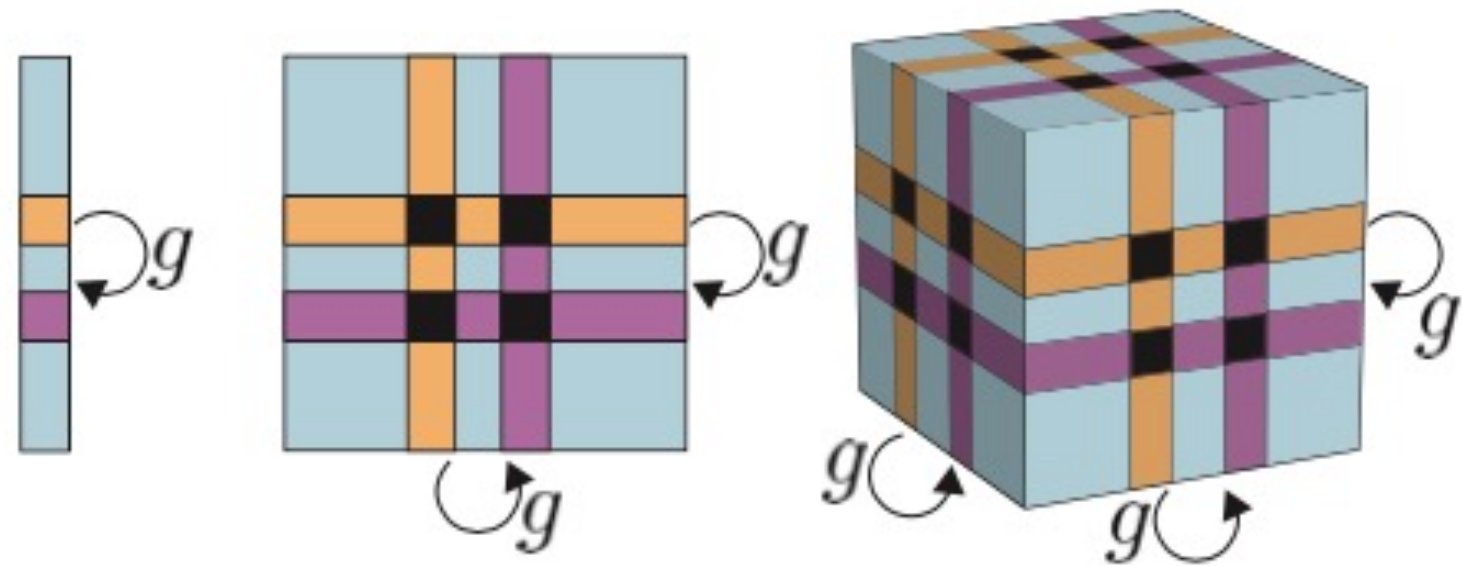
$\mathbb{R}^m \times \mathbb{R}^m$ 上の同値関係 $\sim$ を以下で定義する： $(\lambda x, y) \sim (x, \lambda y)$ ただし λ は 0 でない実数

この同値関係の商集合は M^2 次元のベクトル空間 $\mathbb{R}^{M^2}$ となり、 $\overline{(e_i, e_j)}$ が基底となる。この基底を簡単のため e_{ij} と書く

同様に $\mathbb{R}^M \times \cdots \times \mathbb{R}^M$ に対し、 $(\lambda x_1, x_2, \cdots, x_l) \sim (x_1, \cdots, \lambda x_i, \dots, x_l)$ という同値関係を考えることで l 階のテンソル $\mathbb{R}^{m^l}$ が定義され、その基底を $e_{i_1, \dots, i_l}$ と書く

テンソル自然表現

G を S_M の部分群とする。 $x = (x_{i_1, \dots, i_l}) \in \mathbb{R}^{M^l}$ と $g \in G$ に対し、 $g \cdot x = (x_{g^{-1}(i_1), \dots, g^{-1}(i_l)})$ と定義する。この作用は自然な作用 $S_M \curvearrowright \mathbb{R}^M$ の l 階のテンソルと同値になっている。



テンソルネットワーク

定義: 同変テンソル層とはアファイン写像

$L : \mathbb{R}^{m^k \times a} \rightarrow \mathbb{R}^{m^l \times b}$ であって全ての $g \in G$ と全ての

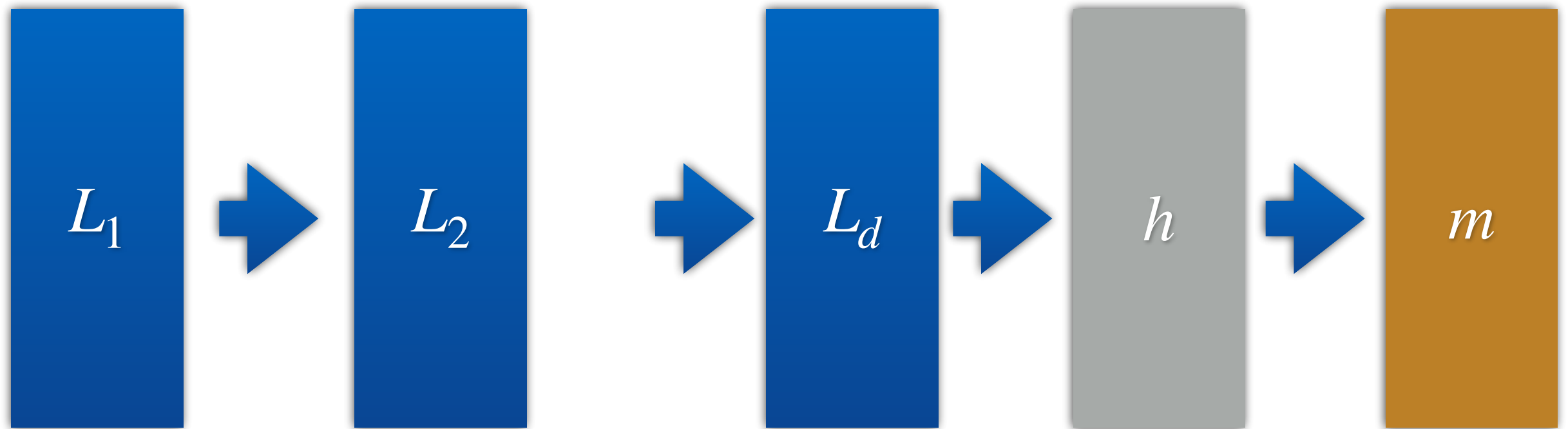
$x \in \mathbb{R}^{m^k \times a}$ に対し、 $L(g \cdot x) = g \cdot L(x)$ を満たすもの。

不変テンソル層とはアファイン写像 $h : \mathbb{R}^{m^k \times a} \rightarrow \mathbb{R}^b$

であって、全ての $g \in G$ と全ての $x \in \mathbb{R}^{m^k \times a}$ に対し

$h(g \cdot x) = h(x)$ を満たすもの。

Maron's model



定義 G 不変テンソルネットワークとは関数

$F : \mathbb{R}^{m^k \times a} \rightarrow \mathbb{R}$ であって $F = m \circ h \circ L_d \circ \sigma \cdots \sigma \circ L_1$ の形のもの。ここで L_i は同変テンソル層で σ は活性化関数、 h は不変テンソル層で、 m は Multi layer perceptron とする。

Universality of invariant tensor networks

定理1 (On the universality of invariant networks):

$K \subset \mathbb{R}^m$ をコンパクト集合とし $G \subset S_m$ を部分群とする。連続な G 不変関数 $f: K \rightarrow \mathbb{R}$ は G 不変テンソルネットワークで任意精度で近似可能

この結果と同変関数の不変関数による表現を組み合わせることで同変関数の Universal approximation theorem が得られる。

不変学習 の汎化誤差

基本領域を 使った汎化誤差の評価

定理(Sannai-Imaizumi) $f: \mathbb{R}^M \rightarrow \mathbb{R}$ を深層ニューラルネットであって関数として S_M 自然な作用に対して不変なものであるとする。この時、十分小さい正の ε に対し、確率 $1 - O(\varepsilon)$ で次の不等式が成立する。

$$R(f) \leq R_n(f) + \sqrt{\frac{C}{m! n \varepsilon^p}} + O(\log(1/\varepsilon))$$

ここで $R_n(f) = n^{-1} \sum_{i=1}^n L(Y_i, f(X_i))$, $R(f) = \mathbb{E}[L(Y, f(X))]$, C は n と m に依らない定数とする。

ポイントクラウドの時は $m = 1000$ くらいのオーダーなので大きく改善している

仲間外れ探しの時でも $16! = 20922789888000$ の改善

証明のアイデア

(1) $\{\mathbb{R}^M$ からの不変連続関数 $\} \cong \{\Delta$ からの連続関数 $\}$

となるような $\Delta \subset \mathbb{R}^M$ を構成。

(2) $\Delta \cap [0,1]^M$ の covering 数を計算して covering 数と汎化誤差の一般論から不等式を得る。

Δ の意味と構成

- ・ 一般の群 G に対し、一つ目の要請を満たす Δ は存在する（商空間 $\mathbb{R}^M/G$ を取れば良い）。
- ・ S_M の時は $\Delta = \{(x_1, \dots, x_M) \mid x_1 \geq x_2 \geq \dots \geq x_M\}$ と取れば $\mathbb{R}^M$ の中に実現することができる。
- ・ $\text{Sort} : \mathbb{R}^M \rightarrow \Delta$ (大きい順にソートする写像) が構成でき $f \mapsto f \circ \text{Sort}$ が右から左への対応を与えている（左から右は制限）
- ・ ちなみにこの対応を深層ニューラルネットレベルで見ることで置換不変の場合のUniversal approximationも出る

まとめと今後の課題

- ・ (有限) 群対称性という事前知識からそれにfitする深層モデルを設計し、ある程度の理論保証 (UATと汎化誤差不等式) をつけた。
- ・ ハイパラの調整については全く手をつけていない。
- ・ ニューラルネットの積分表現論+対称性
- ・ SO_2, SO_3 などの回転群の時にも良いモデルを設計することはできるか？
- ・ 有限群の場合でもテンソルネットは少し大きすぎるように思える。群を限定しても良いのでより小さいものを取れないか？
- ・ 対称性+ノイズのような状況を取り扱う方法はあるか？
- ・ データから背後に潜む対称性を発見する方法はあるか？

[Reference]

- Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, Alexander Smola. Deep Sets , NIPS, 2017. arXiv:1703.06114.
- Haggai Maron, Heli Ben-Hamu, Nadav Shamir, Yaron Lipman. Invariant and Equivariant Graph Networks . ICLR 2019. arXiv:1812.09902
- Haggai Maron, Ethan Fetaya, Nimrod Segol, Yaron Lipman, On the Universality of Invariant Networks, ICML2019, arXiv: 1901.09342
- Akiyoshi Sannai, Yuuki Takai, Matthieu Cordonnier. Universal approximations of permutation invariant/equivariant functions by deep neural networks , arXiv:1903.01939.
- Improved Generalization Bound of Permutation Invariant Deep Neural Networks, Akiyoshi Sannai, Masaaki Imaizumi, arXiv:1910.06552