非凸学習理論チーム
Non-Convex Learning Theory Team

金森 敬文
Takafumi KANAMORI

研究成果報告
- 非規格化統計モデルの学習アルゴリズムと理論解析
- 深層クラスタリングの高精度アルゴリズム
- 非凸モデル統合によるモデル空間拡大と理論解析
- 非もつれ表現の継続学習への応用

非規格化統計モデルの学習アルゴリズムと理論解析

データ $x_1, x_2, \ldots, x_n \sim_{i.i.d.} q(x) \implies q(x)$ を推定.

統計モデル：$p_\theta(x) = \frac{\widetilde{p}_\theta(x)}{Z_\theta}$. ただし Z_θ は計算困難.

目標：Z_θ の計算をしないで最尤推定量を近似的に求める.

アイデア：

変分表現

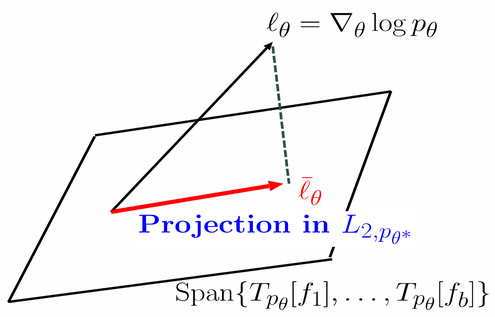
$$\mathrm{KL}(q, p) = \max_{r(x)>0} \int q(x) \log r(x) \mathrm{d}x \text{ s.t. } \mathbb{E}_p[r(x)] = 1$$

経験近似：制約式を Stein operator で処理. 計算可能に！

$$\max_r \underbrace{\int_{\mathcal{X}} \widehat{q}(x) \log r(x) dx \text{ s.t. } \int r(x) p_\theta(x) dx = 1}_{\approx \mathrm{KL}(\widehat{q}, p_\theta)} \longrightarrow \min \text{w.r.t. } \theta.$$

さまざまな理論的成果：

$$\sqrt{n}(\widehat{\theta}_F - \theta^*) \xrightarrow{d} N(0, \Sigma_{F,\theta^*}),$$
$$\Sigma_{F,\theta^*} = (\mathbb{E}_q[\bar{\ell}_\theta \bar{\ell}_\theta^T])^{-1}$$



Variance Monotonicity

$\mathrm{Span}\{T_{p_{\theta^*}}[F_1]\} \subset \mathrm{Span}\{T_{p_{\theta^*}}[F_2]\} \implies \Sigma_{F_1,\theta^*} \succeq \Sigma_{F_2,\theta^*}$

$\nabla_\theta \log p_{\theta^*} \in \mathrm{Span}\{T_{p_{\theta^*}}[F]\} \implies \Sigma_{F,\theta^*} = (\text{Fisher info})^{-1}$

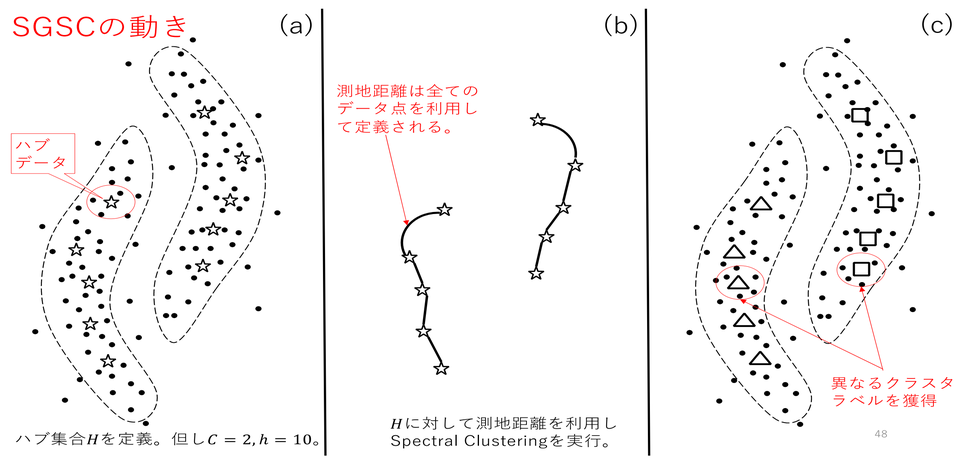
推定された典型画像：確率値が大きい（上段：提案法，下段：既存法）

推定された外れ画像：確率値が小さい（上段：提案法，下段：既存法）

[neurIPS2019: poster]

深層クラスタリングの高精度アルゴリズム

複雑なデータのクラスタリング
- ハブになるサンプル集合から大域的構造を抽出
- クラスタリング：半教師あり学習とみなす(DNNでモデリング)



Method	MNIST	Reuters-10k	FC	TM	TR	Average
k-means	0.53	0.53 (0.04)	0.60 (0.05)	0.64 (0.04)	0.35 (0.03)	0.53
SC	0.72	0.62 (0.03)	0.80 (0.04)	0.85 (0.03)	0.96 (0.03)	0.79
IMSAT	**0.98**	0.71 (0.05)	0.70 (0.04)	0.66 (0.05)	0.34 (0.01)	0.68
DEC	0.84	**0.72 (0.05)**	0.72 (0.04)	0.67 (0.03)	0.48 (0.04)	0.69
SpectralNet	0.83	0.67 (0.03)	0.79 (0.03)	0.87 (0.02)	**0.99 (0.01)**	0.83
SEDC	0.89	**0.73 (0.05)**	**0.95 (0.03)**	**0.96 (0.02)**	**0.99 (0.00)**	**0.90**

[Entropy journal, 2019]

非凸モデル統合によるモデル空間拡大と理論解析

- モデル統合：モデルを組み合わせることでモデル空間を拡大.
- 「非凸」な組み合わせ法を提案する.

非凸な組み合わせ法

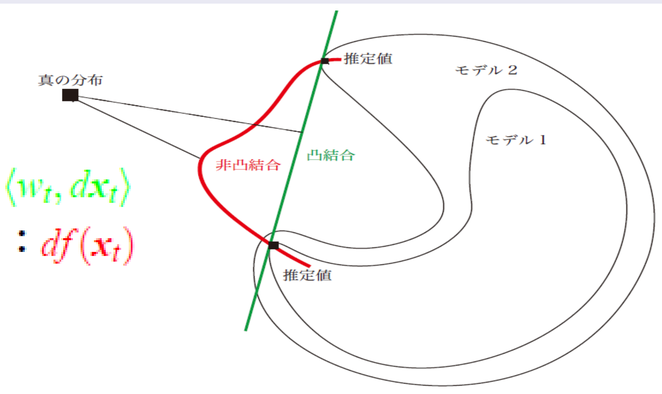
- 前提：各モデルは確率微分方程式でモデリングされる：

$$dx_{j,t} = \mu_j(x_{j,t}, t)\, dt + \sigma_j(x_{j,t}, t)\, dW_t.$$

- 「非凸」組み合わせ $f(x_{1,t}, x_{2,t}, \ldots, x_{J,t})$：伊藤の公式より，従来の凸な組み合わせに，項を一つ加えるだけ！

$$df(x_{1,t}, x_{2,t}, \ldots, x_{J,t}) = \text{凸な組み合わせ} + \text{おつりの項}.$$

- 凸な組み合わせ：$\langle w_t, dx_t \rangle$
- 非凸な組み合わせ：$df(x_t)$



理論的成果

- 予測分布は KL ダイバージェンスの意味で，Exact に Minimax 性をもつ：

$$\min_{p_t} \max_{q_t} \mathbb{E}_{y_t}\left[\mathrm{KL}\left(q_t\left(y_{t+\Delta t} \middle| y_t\right), p_t\left(y_{t+\Delta t} \middle| y_t\right)\right)\right].$$

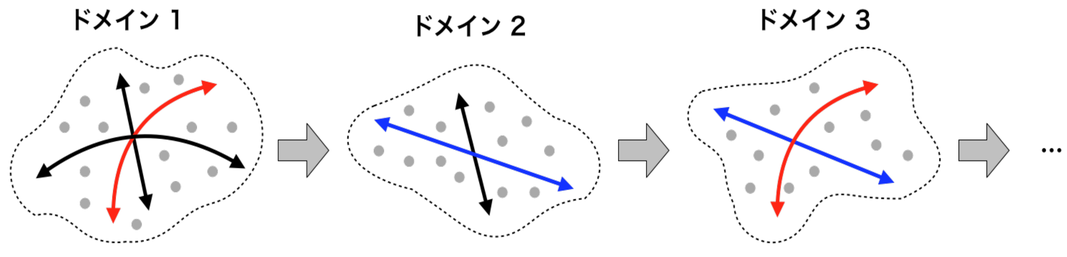
[arXiv:1911.08662]

非もつれ表現の継続学習への応用

■ 継続学習 とは？
- ドメインの系列に対し逐次的に学習を行うこと
- 目的: 過去の知識を用いて新たなタスクでの学習を加速させる
- 問題: 破滅的忘却 (新タスクの学習により旧タスクの知識を喪失)

■ 非もつれ表現
- 仮定: 各ドメインのデータ x は潜在変数 z から生成 x=f(z)
- 分解可能定理[Khemakhem, et al.(2019)]:
  適切な補助変数により潜在変数のアファイン同値類を学習可能

■ 継続学習アルゴリズムとその特徴
- アルゴリズムの概略
  1. 各ドメインで非もつれ表現を学習
  2. 旧ドメインとの潜在変数成分の一致部分を検出
  3. 新規の潜在変数成分に当たる部分のみを学習
- 新規な独立成分のみの学習はサンプル複雑度が低い
- 非もつれ表現は元データよりコンパクトなので低記憶容量で済む



次年度の研究課題

- 非規格化モデルの推定と生成モデルの学習： GANへの展開
- 非凸モデル拡大による統計的推定：機械学習アルゴリズムへの展開
- 多ドメイン深層学習・継続学習：理論基盤の確立