Deep Learning 輪読会 2017

第13章　線形因子モデル

2017.12.11

松尾研究室

特任研究員　村上　遥

構成

第13章　線形因子モデル

13.1 確率的PCAと因子分析

13.2 独立成分分析（ICA）

13.3 Slow Feature Analysis

13.4 スパース符号化

13.5 PCAの多様体解釈

はじめに

- 第II部までの議論：教師あり学習の問題の解決方法
- 第III部の議論：教師なし学習の既存のアプローチ
  - 教師あり学習が適さない問題もある（ex.欠損値処理）
  - 現実には大量にラベルデータがあることの方が少ない
- 13章：線形因子モデル
  - 最先端の研究では入力の確率モデル$p_{model}(x)$の構築が必要とされるので、潜在変数を持つ最も簡単な確率モデルとして線形因子モデルを紹介

13章　線形因子モデル　の章について

- この章は、13.1～13.4が複数の主要な線形因子モデルの紹介、
- 13.5が数理的解釈？の話です。

WHY 確率モデル、WHY 線形因子モデル？

- 最先端の研究の多くで入力の確率モデル$p_{model}(x)$の構築が求められる

$$p_{\text{model}}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{h}} p_{\text{model}}(\boldsymbol{x} \mid \boldsymbol{h})$$

- となる潜在変数hによって別の表現が得られる

- 線形因子モデルは、潜在変数を持つ確率モデルの中で最も簡単なモデル
- 混合モデルや深層確率モデルを構築するための要素として使われる
- hの線形変換にノイズを加えることでxを生成する確率的線形符号化関数を用いて定義される
- 単純な同時分布を持つ説明因子を発見できる

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{h} + \boldsymbol{b} + \text{noise} \qquad \mathbf{h} \sim p(\boldsymbol{h}).$$

- 実数値の観測変数　通常ガウス分布で対角　因子分解可能な分布

13.1 確率的PCAと因子分析

- 確率的PCA=主成分分析、因子分析は線形因子モデルの一種
- 違いは①ノイズ分布と②x観測前の潜在変数hにおけるモデル事前分布

- 因子分析：潜在変数の事前分布は分散が１のガウス分布
- 観測変数xiはhが与えられた下で条件付き独立であると仮定
  - ノイズは対角共分散ガウス分布から得られる
  - 潜在変数の役割は、異なる観測変数xi同士の依存関係を捉えること

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{x}; \boldsymbol{b}, \boldsymbol{W}\boldsymbol{W}^\top + \boldsymbol{\psi}).$$

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{x}; \boldsymbol{b}, \boldsymbol{W}\boldsymbol{W}^\top + \sigma^2 \boldsymbol{I})$$

$$\mathbf{x} = \boldsymbol{W}\mathbf{h} + \boldsymbol{b} + \sigma \mathbf{z}$$

13.2 独立成分分析（ICA）

- 最も古い表現学習アルゴリズムの１つ
- ICA：観測信号を、構成するいくつもの潜在的な信号（互いに完全に独立）に分解しようとする、線形因子をモデル化するためのアプローチの１つ
- 方法論が複数あり、どれもICAと呼ばれている

13.2　独立成分分析（ICA）

- 事前分布p(h)を決定（ガウス分布以外、Wが一意にならないため）
- 決定論的にx= Whが生成される　（ノイズを加えるだけの変種もある）
- 非線形変換でp(x)を決定し、最尤法でモデルの学習を行う

- p(h)を独立に選ぶと元の因子が独立に近い形で復元できる
  - cf.)聖徳太子
  - cf.)脳波記録の際のノイズとの分離

- 生成モデルとの違い
  - xとhの変換方法を知っているが、p(x)を表現する術を持たない

13.3　Slow Feature Analysis

- SFA：時間信号からの情報を使って不変特徴量を学習する線形因子モデル

Q. シマウマが走っていく動画の特徴量をどう取るか

- 動画の構成要素（ピクセル）の特徴量？
  - 黒白黒白黒白・・・と激しく変化
- シマウマの位置
  - ゆっくりと変化

- Slowness原理ー動画では重要な特徴はゆっくり変化するという考え方

13.3　Slow Feature Analysis

原理：

$$\lambda \sum_t \underline{L(f(\boldsymbol{x}^{(t+1)}), f(\boldsymbol{x}^{(t)}))}.$$

λ : slowness正規化項の強さを決定するハイパラ

距離を測る損失関数

f : 正規化される特徴抽出器

平均二乗誤差が多い

最適化問題：

$$\min_{\boldsymbol{\theta}} \mathbb{E}_t (f(\boldsymbol{x}^{(t+1)})_i - f(\boldsymbol{x}^{(t)})_i)^2$$

- 制約１：平均ゼロ　$\mathbb{E}_t f(\boldsymbol{x}^{(t)})_i = 0$　問題の解を一意にする
- 制約２：単位分散を持つ

$$\mathbb{E}_t[f(\boldsymbol{x}^{(t)})_i^2] = 1$$

すべての特徴量が０に潰れない様に

- 変化が遅い特徴量から順に学習される
- 複数の特徴量を学習するには更に以下の制約が必要

$$\forall i < j, \mathbb{E}_t[f(\boldsymbol{x}^{(t)})_i f(\boldsymbol{x}^{(t)})_j] = 0.$$

学習された特徴量が

互いに無相関

13.3　Slow Feature Analysis

- 用途：
- 非線形特徴量の学習
- 利点：損失関数が特定の画素値に大きく依存しない
- どの特徴量をモデルが学習するか理論的に予測可能

- 難点：
- 物体認識や姿勢推定の学習に使われてきたが、最近の研究の基礎にはなっていない
- なぜかは不明

13.4　スパース符号化

- 教師なし特徴量学習と特徴抽出の仕組みとして重点的に研究されてきた線形因子モデル
- スパース符号化：このモデルでhの値を推論する過程
- スパースモデリング：モデルの設計と学習過程

$$p(\boldsymbol{x} \mid \boldsymbol{h}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{W}\boldsymbol{h} + \boldsymbol{b}, \frac{1}{\beta}\boldsymbol{I}).$$

ノイズを加えた線形復号化器を使う

- p(h)は因数分解後のラプラス分布・t分布、コーシー分布など

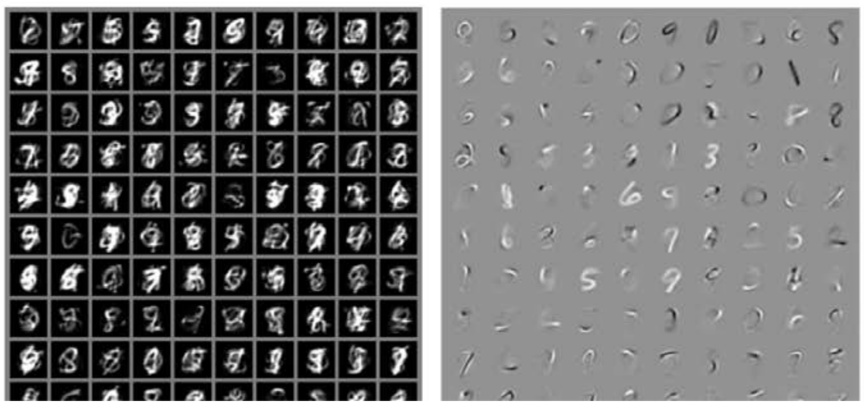
- 最尤法での訓練は困難なため、データの符号化と符号化されたデータをよりよく再構成するような復号化器の学習とを交互に行って学習させる
- =hとWの最小化を交互に行う

$$\boldsymbol{h}^* = f(\boldsymbol{x}) = \arg\max_{\boldsymbol{h}} p(\boldsymbol{h} \mid \boldsymbol{x}).$$

符号化器の最適化アルゴリズム

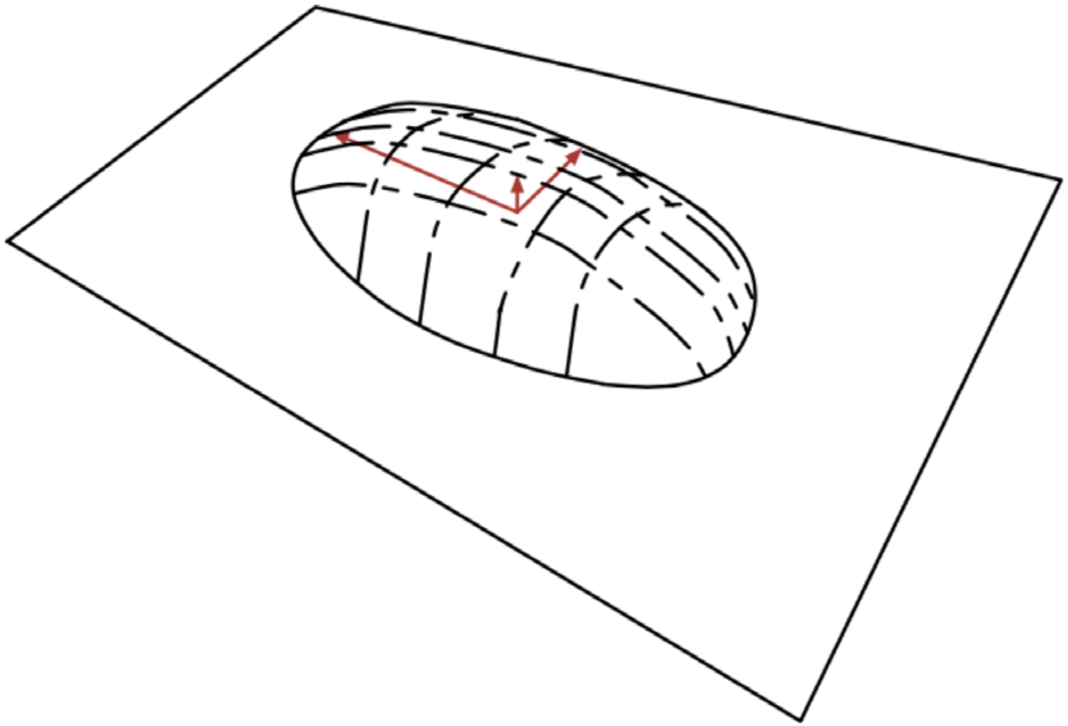
13.4　スパース符号化

- 利点：
- ノンパラメトリックな符号化器と組み合わせると再構成誤差と対数事前分布の組み合わせを原理上、最もよく最小化できる
- 符号化器に汎化誤差がない
- 欠点：ノンパラメトリック符号化器の欠点として、xが与えられた下でのhを計算するのに多くの時間が必要
- 反復アルゴリズムの実行が原因
- ノンパラメトリック符号化器を介する逆伝播が簡単ではない



13.5　PCAの多様体解釈

- PCAと因子分析を含む線形因子モデルは多様体を学習していると解釈可能



参考文献

- Deep Learning
  - Ian Goodfellow, Yoshua Bengio, Aaron Courville
  - 日本語版 https://www.amazon.co.jp/%E6%B7%B1%E5%B1%A4%E5%AD%A6%E7%BF%92-Ian-Goodfellow/dp/4048930621