

非ガウスデータに関する次元圧縮とクラスタリングの同時最適化と 工学的パターン認識への応用

Simultaneous dimension reduction and clustering for non-Gaussian high-dimensional data and application to pattern recognition

渡辺一帆[†], 赤穂昭太郎^{*}, 大町真一郎^{**}, 岡田真人 (P)^{†‡}

Kazuho Watanabe, Shotaro Akaho, Shinichiro Omachi, and Masato Okada

[†] 東京大学大学院新領域創成科学研究科, ^{*} 産業技術総合研究所脳神経情報研究部門

^{**} 東北大学大学院工学研究科, [‡] 理化学研究所脳科学総合研究センター

kazuho@mns.k.u-tokyo.ac.jp

Abstract— In this paper, we introduce a latent variable model for the exponential principal component analysis (e-PCA) and present a simultaneous clustering and dimensionality reduction technique using the mixture of exponential family distributions whose parameters are constrained to a low dimensional subspace. We derive a variational Bayesian learning algorithm using Laplace's method to intractable integration and apply the present method to the MNIST handwritten digits data.

Keywords— Dimensionality reduction, Clustering, Mixture of exponential family distributions, e-PCA

1 はじめに

近年マルチニューロンデータをクラスタリングして得られた構造と, 入力画像などの構造を比較することで, 外界の情報がどのような形で神経ダイナミクスに埋め込まれているかを議論できるようになりつつある [1].

一般に神経系のデータでは, データの次元に比べてサンプル数が十分とはいえない. また発火率は非負の値を取り, スパイク数は非負の整数値をとるために, 集団ベクトルの各要素は一般にはガウス分布に従わない. 我々は, これらの少数データと非ガウス性の問題を解決するクラスタリング手法を提案した [2]. この手法では, 混合指数分布を Akaho によって提案された e-PCA で次元圧縮する [3]. 次元圧縮とクラスタリングを同時に行うので, サンプル数に応じてクラスタリングにとって適切な部分空間を選ぶことができる.

今回その能力を調べるために, 提案手法を手書きの数字の MNIST データベースに関するパターン認識課題に適用した. パターン認識に用いられる特徴ベクトルの各要素は非ガウス分布しており, 今回の MNIST データベースは我々の手法の性能を検証するよい例になっている. 混合指数分布と e-PCA にもとづく次元圧縮を適用することにより, 混合正規分布を用いた従来手法 [4] より

りも認識率が向上することがわかった.

2 指数型分布族と e-PCA

確率ベクトル $\mathbf{x} \in \Xi$ の分布で確率密度関数 (もしくは確率関数) が次の形を持つものを指数型分布族という.

$$p(\mathbf{x}|\theta) = \exp\{\theta \cdot \mathbf{F}(\mathbf{x}) + F_0(\mathbf{x}) - G(\theta)\} \quad (1)$$

ここで $\theta = (\theta_1, \dots, \theta_M)^T \in \Theta$ は自然パラメータと呼ばれ, $\theta \cdot \mathbf{F}(\mathbf{x})$ は θ と $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \dots, F_M(\mathbf{x}))^T$ の内積, $\theta \cdot \mathbf{F}(\mathbf{x}) = \sum_{j=1}^M \theta_j F_j(\mathbf{x})$ である.

e-PCA は指数型分布の空間中にパラメータのサンプルが多数与えられたときに, それらを良く説明する低次元部分空間を抽出する方法である.

$$\tilde{\theta}(\mathbf{w}) = \mathbf{U}\mathbf{w} + \mathbf{u}_0 = \sum_{j=1}^L w_j \mathbf{u}_j + \mathbf{u}_0 \quad (2)$$

により Θ の L 次元部分空間を表す. ここで行列 \mathbf{U} は L 個の基底となる M 次元ベクトル $\mathbf{u}_1, \dots, \mathbf{u}_L$ を列ベクトルとして並べた行列であり $\mathbf{w} = (w_1, w_2, \dots, w_L) \in R^L$ は $\tilde{\theta}$ の低次元表現である. 集合 $\{\tilde{\theta}(\mathbf{w}); \mathbf{w} \in R^L\}$ は e-平坦部分空間 [3] となり全ての $\theta \in \Theta$ に対し, カルバック情報量 $K(\theta|\tilde{\theta}) \equiv \int p(\mathbf{x}|\theta) \log \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\tilde{\theta})} d\mathbf{x}$ の最小値を与える $\tilde{\theta}$ が一意に存在することが知られている. この事実に基づき, e-PCA は自然パラメータの例 $\Theta^n = \{\theta^{(i)}\}_{i=1}^n$ が学習データとして与えられたとき, 目的関数 $\sum_{i=1}^n K(\theta^{(i)}|\tilde{\theta}(\mathbf{w}^{(i)}))$ を最小化するようにデータの低次元表現 $\mathbf{W}^n = \{\mathbf{w}^{(i)}\}_{i=1}^n$ と基底ベクトル \mathbf{U} 及び \mathbf{u}_0 を求める. e-PCA のための逐次最適化法が与えられている [3].

3 制約付き混合分布

$\mathbf{x} \in \Xi$ として n 個のデータ $X^n = \{\mathbf{x}^{(i)}\}_{i=1}^n$ が与えられたとする. このとき確率的 PCA を指数型分布に拡張することにより, データ \mathbf{x} と潜在変数 $\mathbf{w} = (w_1, w_2, \dots, w_L) \in R^L$ の同時分布として以下の確率分布を考える.

$$p(\mathbf{x}, \mathbf{w}) = p(\mathbf{x}|\tilde{\theta}(\mathbf{w}))p(\mathbf{w}) \quad (3)$$

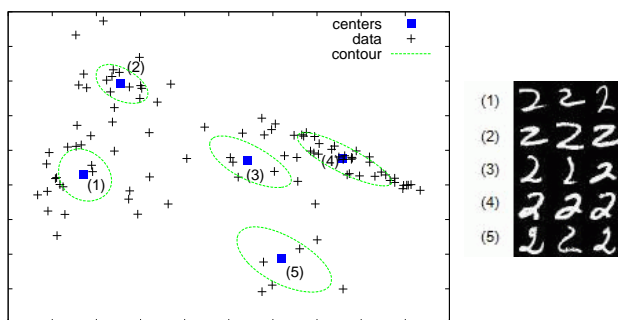


図 1: 左: データ 2 次元表現, 得られたクラスタ中心及び中心へのカルバック情報量の等高線. 右: 各クラスタ中心に最も近い 3 つのサンプル

ここで各 $F(\mathbf{x}^{(i)})$ を期待値を持つ (1) 式の分布の自然パラメータを $\theta^{(i)}$ とする対応を用いると, $p(\mathbf{w})$ を一様分布として上式から求めた $\mathbf{W}^n = \{\mathbf{w}^{(i)}\}_{i=1}^n$ の事後分布の最大解は e-PCA による解に一致する. ここではデータのクラスタリングを行なうため, (3) において $p(\mathbf{w})$ を離散分布 $\sum_{k=1}^K a_k \delta(\mathbf{w} - \mathbf{w}_k)$ (δ はデルタ関数, a_k は $a_k \geq 0, \sum_{k=1}^K a_k = 1$ を満たす実数) とし, \mathbf{w} について周辺化するとパラメータが低次元空間に制限されている混合指数型分布

$$p(\mathbf{x}|\omega) = \sum_{k=1}^K a_k p(\mathbf{x}|\tilde{\theta}(\mathbf{w}_k))$$

となる. ここで $\omega = \{\{a_k\}_{k=1}^K, \{\mathbf{w}_k\}_{k=1}^K; 0 \leq a_k \leq 1, \sum_{k=1}^K a_k = 1, \mathbf{w}_k \in R^L\}$ はパラメータ.

4 変分ベイズ法によるクラスタリング

z_k をデータ \mathbf{x} が k 番目のコンポーネントから得られたものであるとき 1, それ以外るとき 0 となる変数として, 潜在変数 $\mathbf{z} = (z_1, z_2, \dots, z_K)$ を導入すると, 次の同時分布を得る.

$$p(\mathbf{x}, \mathbf{z}|\omega) = \prod_{k=1}^K \{a_k p(\mathbf{x}|\tilde{\theta}(\mathbf{w}_k))\}^{z_k}. \quad (4)$$

変分ベイズ法は, 次式で定義される変分自由エネルギー $\tilde{F}[q]$ を最小にするようにパラメータの近似事後分布 $q(\omega)$ と潜在変数 $Z^n = \{\mathbf{z}^{(i)}\}_{i=1}^n$ の近似事後分布 $q(Z^n)$ を交互最適化する学習法である.

$$\tilde{F}[q] = \sum_{Z^n} \int q(Z^n) q(\omega) \log \frac{q(Z^n) q(\omega)}{p(\omega) \prod_{i=1}^n p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}|\omega)} d\omega$$

($p(\omega)$ はパラメータの事前分布). 前節の制約付き混合モデルの変分ベイズ法のアルゴリズムには, 一般の \mathbf{U} に対し $\alpha \in R^L$ と $\xi \in R$ の関数 $\Phi(\alpha, \xi) = \log \int \exp\{\xi(\alpha \cdot \mathbf{w} - G(\tilde{\theta}(\mathbf{w})))\} d\mathbf{w}$ の偏導関数 $\frac{\partial \Phi}{\partial \alpha}, \frac{\partial \Phi}{\partial \xi}$ の計算が必要となるが, ラプラス近似を用いた近似法を与えることができる [2].

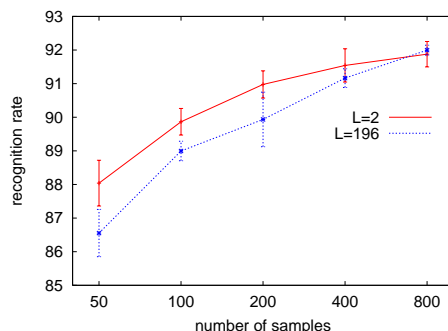


図 2: 学習例数に対する認識率

5 次元圧縮

これまででは部分空間の基底ベクトル \mathbf{U} を固定してきたが, 推定中のクラスタの中心 $\nu_k = \frac{1}{n_k} \sum_{i=1}^n \langle z_k^{(i)} \rangle_{q(Z^n)} \mathbf{F}(\mathbf{x}^{(i)})$ ($n_k = \sum_{i=1}^n \langle z_k^{(i)} \rangle_{q(Z^n)}$ とおいた) に e-PCA を適用することで, 部分空間を同時に推定することができる. すなわち, $\sum_{k=1}^K n_k K(\theta(\nu_k) || \hat{\theta}(\langle \mathbf{w}_k \rangle_{q(\mathbf{w}_k)}))$ を最小化するように各基底ベクトルを更新する ($\theta(\nu_k)$ は (1) 式で期待値 ν_k を持つパラメータ θ への対応).

また, 各データの低次元表現は他の変数 (\mathbf{w}^K, Z^n など) に関して周辺化した事後分布 $p(\mathbf{W}^n | X^n)$ による平均として求めることができる.

6 手書き数字認識データへの適用

本手法を手書き数字認識データ (MNIST データ) に適用する実験を行なった. データは 196 次元の非負整数値のベクトルに変換されており, $\mathbf{F}(\mathbf{x}) = \mathbf{x}, G(\theta) = \prod_{j=1}^M e^{\theta_j}$ としたポアソン分布の混合モデルを用いた. 図 1 に数字 '2' のデータに対する適用例 (コンポーネント数 $K = 5, L = 2$ 次元に圧縮した場合) を示す. 10 個の数字のクラスに対し $K = 5$ の混合ポアソン分布を学習し, 10000 個のテストデータに対する認識率を評価した. 図 2 に全 196 次元を用いた場合と $L = 2$ 次元に圧縮した場合について, 各クラスの学習例数を変えたときの認識率の結果を示す. 学習例数が少ないときの次元圧縮の効果がわかる. 具体的に適切な次元数を求める手法としては, クロスバリデーションを用いることができる. また変分自由エネルギーを最大にする次元数を求めることで, モデル選択することも可能であることを数値実験で確認している.

参考文献

- [1] Matsumoto, Okada, Y. Sugase-Miyamoto, Yamane and K. Kawano, *Cerebral Cortex*, 15, 1103, 2005
- [2] Watanabe, Akaho and Okada, *Proc. of ITS/L*, to appear, 2008.
- [3] Akaho, *Proc. of IJCNN*, 129, 2004.
- [4] Omachi, Omachi and Aso, *IEICE Transactions on Information and Systems*, E90-D, 1160, 2007.