On the Consistency of Maximum Likelihood Estimation of Probabilistic Principal Component Analysis

Arghya Datta [1], Sayak Chakrabarty [2]

[1]Université de Montréal
[2]Northwestern University

NeurIPS 2023

Table of Contents

Table of Contents

The model: PPCA

- **Probabilistic Principal Component Analysis (PPCA) Model Description:** PPCA is a generative framework where data points $\{x_i\}_{i=1}^n \in \mathbb{R}^p$ are independently derived via the equation $x = \mathbf{W}z + \epsilon$, where ϵ follows a normal distribution $\mathcal{N}(0, \sigma^2)$.

  In this model, $\mathbf{W}$ is a $p \times q$ loading matrix, and $z \in \mathbb{R}^q$ is a latent variable independently sampled from $\mathcal{N}(0, I_q)$. Here, p and q are integers with $p \gg q$ and $\text{rank}(\mathbf{W}) = q$. The objective is to estimate the loading matrix $\mathbf{W}$ and the variance σ^2 given the datapoints whose marginal distribution is $\mathcal{N}(0, \mathbf{W}\mathbf{W}^T + \sigma^2 I_p)$.

ML solution of the PPCA model

- **Maximum Likelihood Estimation for Observed Data**: For n observed data points $x_1, x_2, \ldots, x_n$, the maximum likelihood estimators are defined as $\widehat{\mathbf{W}} = \mathbf{U}(\Delta_q - \widehat{\sigma}^2 I_q)^{1/2}\mathbf{R}$ and $\widehat{\sigma}^2 = \frac{1}{p-q}\sum_{j=q+1}^{p} \delta_j$, where $\mathbf{R}$ is a $q \times q$ rotational matrix. The columns of $\mathbf{U}$ are the first q eigenvectors of the sample covariance matrix $S_x = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$, and $\Delta_q = \text{diag}(\delta_1, \ldots, \delta_q)$ represents the top q eigenvalues of S_x in descending order.

No theoretical guarantee of the ML solution

- **Addressing the Lack of Theoretical Guarantees in PPCA:** Despite its prevalence as a method for dimensionality reduction, the maximum likelihood (ML) solution for the PPCA model lacked theoretical underpinnings.

Table of Contents

The Main Challenge

- **Addressing Identifiability Through Parameter Space Modification:** The issue of non-identifiability can be eliminated by modifying the geometry of the parameter space, denoted as $\Theta := \mathbb{R}^{p \times q} \times \mathbb{R}_+$. This is achieved by transitioning through a topological quotient of the Euclidean space.
- **Conceptual Inspiration and Technical Challenges:** The conceptual framework for this approach is drawn from Redner [1981]. However, employing a quotient parameter space introduces additional technical complexities that demand careful attention and were not comprehensively tackled in Redner [1981].

Key Ideas and Methodology

Given a topological space Θ and an equivalence relation $\sim$ on it, one can naturally topologize the quotient space $\Theta/\sim$, which makes the quotient map $\pi : \Theta \to \Theta/\sim$ continuous. A point $\theta \in \Theta$ can be represented as $(\mathbf{W}_\theta, \sigma_\theta^2)$. We consider the closed subset defined by

$$C := \{\theta \in \Theta : \mathbf{W}_\theta \mathbf{W}_\theta^T + \sigma_\theta^2 I_p = \mathbf{W}_0 \mathbf{W}_0^T + \sigma_0^2 I_p\} \subset \Theta.$$

The equivalence relation defined by $\theta \sim \phi$ if $\theta = \phi$ or $\theta, \phi \in C$ allows us to get rid of the identifiability issue. We aim to prove consistency in the quotient space Θ/C.

Three Major Roadblocks

- There is no *obvious* metric structure in $\Theta/\sim$ for Wald's framework to apply.
- The topology generated by the (pseudo) metric in $\Theta/\sim$ is different from the quotient topology of $\Theta/\sim$, in general.
- Interpreting Wald's conditions within the quotient parameter space $\Theta/\sim$ is in general hard, as the metric structure in Θ/C is abstract and elusive. We can, however, show that in our case the metric structure can be put into a tractable form.

Our Contributions

- **Establishing Consistency in PPCA:** Consider $\hat{\theta} = (\widehat{\mathbf{W}}, \widehat{\sigma}^2)$ as the sequence of maximum likelihood estimates in the PPCA model. We prove that $[\hat{\theta}] \xrightarrow{\mathbb{P}} [\theta_0]$.
- **Covariance Estimation:** Furthermore, as an application of the previous result we obtain $\widehat{\mathbf{W}}\widehat{\mathbf{W}}^T + \widehat{\sigma}^2 I_p \xrightarrow{\mathbb{P}} \mathbf{W}_0\mathbf{W}_0^T + \sigma_0^2 I_p$.
- **Strong Consistency:** We can prove a.s. versions of the previous results if $\theta_0 = (\mathbf{W}_0, \sigma_0^2)$ is contained in a compact subset of the parameter space Θ.

Our Contributions Continued...

- **Broad Applicability:** The consistency results obtained are not limited to Maximum Likelihood Estimators (MLE) but extend to a wide range of estimators within the PPCA model.
- **Elaborating Quotient Topological Space Theory:** This work presents a thorough and rigorous exploration of the theory of quotient topological spaces, particularly in statistical applications, addressing gaps previously present in the academic literature.

Table of Contents

Potential future research directions

- Opens the door for proving more theoretical results like the asymptotic distribution of the MLE in the quotient space.
- Flexible methodology: The quotient space framework does not depend much on the statistical model. It is readily applicable to models where ambiguity is present due to identifiability (eg. Matrix factorization).

Thank you for listening!