Theoretical Connection between Locally Linear Embedding, Factor Analysis, and Probabilistic PCA

Benyamin Ghojogh*, Ali Ghodsi**, Fakhri Karray*, Mark Crowley*

* Department of Electrical & Computer Engineering
** Department of Statistics and Actuarial Science
University of Waterloo, ON, Canada

June 1, 2022

Twitter: @comppthink
PubPub (7eqtuyyc): caiac.pubpub.org/pub/7eqtuyyc
What is Dimensionality Reduction?

- **Big Idea:** extracting informative low-dimensional features from high-dimensional data.
- **Also known as:** Manifold Learning, Feature Extraction, Finding a "projection" to a simpler space
- **Useful for:** Data Preprocessing and Reduction, Visualization, Improved Performance on high-dimensional data, ML on Embedded Systems, ..., many more.

Dimensionality Reduction methods can be *divided into three broad categories*:

1. **Spectral methods**:
   - Examples: PCA and LLE
2. **Probabilistic methods**:
   - Examples: Probabilistic PCA, Factor Analysis
3. **Neural network-based methods**:
   - Examples: Restricted Boltzmann Machine, Variational Autoencoder
Building a Bridge

In this work we build a bridge between the *spectral* and *probabilistic* approaches to Dimensionality Reduction.
In particular, we look at these three methods:

1. Factor Analysis
2. Probabilistic PCA
3. LLE

We show:
- how these methods are all tightly related,
- and how this relationship explains their different properties.
Factor analysis [1, 2] assumes that every data point $x_i$ is generated from a latent factor $w_i$ [3].

$$x_i := \Lambda w_i + \mu + \epsilon,$$  \hspace{1cm} (1)

$$\mathbb{P}(x_i | w_i, \Lambda, \mu, \Psi) = \mathcal{N}(x_i; \Lambda w_i + \mu, \Psi).$$  \hspace{1cm} (2)
Probabilistic PCA

Probabilistic PCA [4, 5] is a special case of factor analysis where the variance of noise is equal in all dimensions of data space with covariance between dimensions, i.e. [3]:

$$\Psi = \sigma^2 I.$$  \hfill (3)

Therefore:

$$x_i := \Lambda w_i + \mu + \epsilon,$$  \hfill (4)

$$P(\epsilon) = \mathcal{N}(0, \sigma^2 I),$$  \hfill (5)

$$P(x_i | w_i, \Lambda, \mu, \sigma^2 I) = \mathcal{N}(x_i; \Lambda w_i + \mu, \sigma^2 I).$$  \hfill (6)
Locally Linear Embedding (LLE)

LLE [6, 7] has two main steps [8]:
- linear reconstruction
- linear embedding

Linear reconstruction of LLE can be seen stochastically where every point \( x_i \) is conditioned on and generated by its reconstruction weights \( w_i \) as a latent factor:

\[
x_i = X_i w_i + \mu, \quad (7)
\]
\[
P(w_i) = \mathcal{N}(w_i; 0, \Omega_i). \quad (8)
\]

The covariance \( \Omega_i \) can be learned by Expectation Maximization (EM).
- If \( \Omega_i = \sigma_i I \) is assumed (like in Probabilistic PCA), we’ll have close-form solution (as in the Probabilistic PCA).
- See our paper at the conference for more details.
Connection of LLE, Factor Analysis, and Probabilistic PCA

- Comparing Eqs. (1) and (7):

\[ x_i := \Lambda w_i + \mu + \epsilon, \quad \text{(factor analysis, probabilistic PCA)} \]
\[ x_i = X_i w_i + \mu, \quad \text{(LLE)} \]

shows that data point \( x_i \) is conditioned on some latent variable \( w_i \) (using a transformation matrix), in all methods of factor analysis, probabilistic PCA, and LLE.

- In factor analysis and probabilistic PCA: \( x_i := \Lambda w_i + \mu + \epsilon \).
  - Global matrix \( \Lambda \)
  - So it is data-independent (it is the same matrix for all data points).

- In LLE: \( x_i = X_i w_i + \mu \).
  - local matrix \( X_i \)
  - So it is data-dependent (it is different for every data point).

This explains why factor analysis and probabilistic PCA are linear methods and LLE is a nonlinear algorithm.
Thank You

- “Theoretical Connection between Locally Linear Embedding, Factor Analysis, and Probabilistic PCA”
- Benyamin Ghojogh, Ali Ghodsi, Fakhri Karray, Mark Crowley
- PubPub (7eqtuyyc): caiac.pubpub.org/pub/7eqtuyyc

Acknowledgements

- Research funded in part by:
  - NSERC - Discovery Grant Program, CRD Grants on Digital Pathology
  - NRC - Collaboration Centre Programs
References

[1] B. Fruchter, Introduction to factor analysis. Van Nostrand, 1954.

[2] D. Child, The essentials of factor analysis. Cassell Educational, 1990.

[3] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, “Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: Tutorial and survey,” arXiv preprint arXiv:2101.00734, 2021.

[4] S. Roweis, “EM algorithms for PCA and SPCA,” Advances in neural information processing systems, vol. 10, pp. 626–632, 1997.

[5] M. E. Tipping and C. M. Bishop, “Probabilistic principal component analysis,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 61, no. 3, pp. 611–622, 1999.

[6] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, no. 5500, pp. 2323–2326, 2000.

[7] L. K. Saul and S. T. Roweis, “Think globally, fit locally: unsupervised learning of low dimensional manifolds,” Journal of machine learning research, vol. 4, no. Jun, pp. 119–155, 2003.

[8] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, “Locally linear embedding and its variants: Tutorial and survey,” arXiv preprint arXiv:2011.10925, 2020.