Review of Extended Abstract

Probabilistic PCA

Presenter: Roxana Shaikh
Supervisor: Moritz Herrmann
Reviewer: Lisa Wimmer

January 15^{th}, 2021

Review

Overall impression

The extended abstract contains many important aspects and shows you have already gained a solid understanding of PPCA. In particular, properties and drawbacks of the presented methods are pointed out and used to motivate their respective extensions/modifications. This helps to establish a thread of arguments and yields a convincing story line. In places, however, your explanations are not perfectly clear to me yet (see below). Keeping a stronger focus on the potential audience might lead to better accessibility for readers.

Stated objective

The objective is stated clearly and explicitly. It appears to me just a little biased in its current wording, making it sound as if PPCA dominated PCA in all aspects.

Proposed structure

What structure is given seems coherent, but some points strike me as missing entirely (which may, of course, be due to the brevity of an abstract):

- Given that the context of this seminar is manifold learning, some motivation on dimensionality reduction would seem in order.
- In a report on PPCA I would expect a solid introduction of standard PCA.
- Tipping and Bishop (1999) themselves, and many others (e.g., Burges (2004), Bouveyron et al. (2011)), stress the similarity between PPCA and factor analysis, so this seems worth mentioning in motivating PPCA.
- Some comparison on related work could help to put PPCA into context.
- There is no reference to the practical implementation part.
- You do not list a discussion. Many hints at strengths and drawbacks are given in the text, but a dedicated chapter for synopsis could help to provide a clear structure.

Content outline

- **Objective.** I am not sure I fully understand what you describe as the principal strength/drawback here. It seems to me you might refer to work by Shlens (2005). Suggesting a probabilistic model so users may incorporate prior knowledge is not entirely convincing to me. Personally, I would rather follow Tipping and Bishop (1999) and motivate the probabilistic extension mainly via inference options and comparability to other probabilistic approaches, compatibility to Bayesian frameworks, and the possibility to construct model mixtures.
- **Probabilistic PCA.** I would caution against this introduction of PCA in the context of non-linear methods – after all, PPCA is still linear (and non-linearity is achieved, in a way, only by a mixture of these linear models (Tipping and Bishop, 1999)). In the second sentence, you give a summary of PPCA which, I believe, is inspired by Bishop (2006). While this is certainly correct, I am not certain it the most concise way to describe PPCA. At least to me, the

meaning is not intuitively clear at first glance. Lastly, I would expect some reference to the generative nature of PPCA.

- **MLPCA.** The abstract does not mention the central role Gaussians play, especially in the MLE part.
- **MPPCA.** Here, I miss the motivation for generating mixtures of PPCA analyzers (you describe it as "inevitable", by which you perhaps mean that it results as a logical next step). The second sentence, I am afraid, I do not quite understand.

Further comments

The abstract contains several spelling mistakes. For the report, I would recommend a thorough spell check as such errors might give otherwise well-researched work a sloppy impression.

References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer.

Bouveyron, C., Celeux, G. and Girard, S. (2011). Intrinsic dimension estimation by maximum likelihood in isotropic probabilistic pca, *Pattern Recognition Letters* **32**(14): 1706–1713.

Burges, C. J. C. (2004). Geometric methods for feature extraction and dimensional reduction: A guided tour, *Technical Report MSR-TR-2004-55*, Microsoft Research.

Shlens, J. (2005). A tutorial on principal component analysis, *Technical report*.

Tipping, M. E. and Bishop, C. M. (1999). Mixtures of probabilistic principal component analysers, *Neural Computation* **11**(2): 443–482.