Homework 2

CS420 Machine learning 2018 Spring[*]
Department of Computer Science and Engineering
Shanghai Jiao Tong University

Submission deadline: 20:00, May 17, 2018, Thursday

Submission to:

Please submit your homework in pdf format to the CS420 folder in the following FTP. File name should be like this: `015033910032_chenyajing_hw1.pdf`.

ftp://public.sjtu.edu.cn
username: cyj907
password: public

1 (10 **points**) PCA algorithm

Give at least two algorithms that could take data set $X = \{x_1, \ldots, x_N\}$, $x_t \in \mathbb{R}^{n \times 1}, \forall t$ as input, and output the first principal component $\mathbf{w}$. Specify the computational details of the algorithms, and discuss the advantages or limitations of the algorithms.

2 (10 **points**) Factor Analysis (FA)

Calculate the Bayesian posterior $p(\mathbf{y}|\mathbf{x})$ of the Factor Analysis model $\mathbf{x} = \mathbf{A}\mathbf{y} + \mu + \mathbf{e}$, with $p(\mathbf{x}|\mathbf{y}) = G(\mathbf{x}|\mathbf{A}\mathbf{y} + \mu, \Sigma_e), p(\mathbf{y}) = G(\mathbf{y}|0, \Sigma_y)$, where $G(\mathbf{z}|\mu, \Sigma)$ denotes Gaussian distribution density with mean μ and covariance matrix Σ.

3 (10 **points**) Independent Component Analysis (ICA)

Explain why maximizing non-Gaussianity could be used as a principle for ICA estimation.

4 (30 **points**) Causal discovery algorithms

Apply one causal discovery algorithm on a real world problem. You need to specify the details of the problem, collect the data by yourself or from a public website, briefly summarize what algorithm you use, and explain the results.

You may use any causal discovery algorithm described in the following paper [Spirtes et al., 2016], and use the software packages in Page 26 of the paper.

- Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. Applied Informatics, 3:3, 2016
  `https://applied-informatics-j.springeropen.com/track/pdf/10.1186/s40535-016-0018-x`

---

[*]tushikui@sjtu.edu.cn

Page 26:
The following software packages are available online:

- The Tetrad project webpage (Tetrad implements a large number of causal discovery methods, including PC and its variants, FCI, and LiNGAM): `http://www.phil.cmu.edu/tetrad/`.
- Kernel-based conditional independence test Zhang et al. (2011): `http://people.tuebingen.mpg.de/kzhang/KCI-test.zip`.
- LiNGAM and its extensions, Shimizu et al. (2006, 2011): `https://sites.google.com/site/sshimizu06/lingam`.
- Fitting the nonlinear additive noise model Hoyer et al. (2009): `http://webdav.tuebingen.mpg.de/causality/additive-noise.tar.gz`
- Distinguishing cause from effect based on the PNL causal model, Zhang and Hyvärinen (2009, 2010): `http://webdav.tuebingen.mpg.de/causality/CauseOrEffect_NICA.rar`
- Probabilistic latent variable models for distinguishing between cause and effect, Mooij et al. (2010): `http://webdav.tuebingen.mpg.de/causality/nips2010-gpi-code.tar.gz`
- Information-geometric causal inference, Daniusis et al. (2010); Janzing et al. (2012): `http://webdav.tuebingen.mpg.de/causality/igci.tar.gz`

5 (20 bonus points) Causal tree reconstruction

Give an algorithm to reconstruct a causal tree based on star-decomposable necessary and sufficient conditions.