
Homework 2

CS420 Machine learning 2021 Spring*
Department of Computer Science and Engineering
Shanghai Jiao Tong University

Submission deadline: 23:59, May 21, 2021, Friday

Submission to:

Please submit your homework in pdf/doc format to Canvas platform.

1 (10 points) PCA algorithm

Give at least two algorithms that could take data set $X = \{x_1, \dots, x_N\}$, $x_t \in \mathbb{R}^{n \times 1}, \forall t$ 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) Dimensionality Reduction by FA

Consider the following Factor Analysis (FA) model,

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mu + \mathbf{e}, \quad (1)$$

$$p(\mathbf{x}|\mathbf{y}) = G(\mathbf{x}|\mathbf{A}\mathbf{y} + \mu, \sigma^2 \mathbf{I}), \quad (2)$$

$$p(\mathbf{y}) = G(\mathbf{y}|0, \mathbf{I}), \quad (3)$$

where the observed variable $\mathbf{x} \in \mathcal{R}^n$, the latent variable $\mathbf{y} \in \mathcal{R}^m$, and $G(\mathbf{z}|\mu, \Sigma)$ denotes Gaussian distribution density with mean μ and covariance matrix Σ . Write a report on experimental comparisons on model selection performance by BIC, AIC on selecting the number of latent factors, i.e., $\mathbf{dim}(\mathbf{y}) = m$.

Specifically, you need to randomly generate datasets based on FA, by varying some setting values, e.g., sample size N , dimensionality n and m , noise level σ^2 , and so on. For example, set $N = 100$, $n = 10$, $m = 3$, $\sigma^2 = 0.1$, $\mu = 0$, and assign values for $\mathbf{A} \in \mathcal{R}^{n \times m}$. The generation process is as follows:

- (1) Randomly sample a \mathbf{y}_t from Gaussian density $G(\mathbf{y}|0, \mathbf{I})$, with $\mathbf{dim}(\mathbf{y}) = m = 3$;

*tushikui@sjtu.edu.cn

- (2) Randomly sample a noise vector \mathbf{e}_t from Gaussian density $G(\mathbf{e}|0, \sigma^2 \mathbf{I})$, with $\sigma^2 = 0.1$, $\mathbf{e}_t \in \mathcal{R}^n$;
- (3) Get $\mathbf{x}_t = \mathbf{A}\mathbf{y}_t + \mu + \mathbf{e}_t$.

Collect all the \mathbf{x}_t as the dataset $X = \{\mathbf{x}_t\}_{t=1}^N$.

The two-stage model selection process for BIC, AIC is as follows:

Stage 1: Run **EM algorithm** on each dataset X for $m = 1, \dots, M$, and calculate the log-likelihood value $\ln[p(X|\hat{\Theta}_m)]$, where $\hat{\Theta}_m$ is the maximum likelihood estimate for parameters;

Stage 2: Select the **optimal m^*** by

$$m^* = \arg \max_{m=1, \dots, M} J(m), \quad (4)$$

$$J_{AIC}(m) = \ln[p(X|\hat{\Theta}_m)] - d_m, \quad (5)$$

$$J_{BIC}(m) = \ln[p(X|\hat{\Theta}_m)] - \frac{\ln N}{2} d_m, \quad (6)$$

where d_m denotes the number free parameters of FA model with m latent factors. You may set $M = 5$, if you generate the dataset X based on $n = 10, m = 3$.

The following codes might be useful.

Python: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FactorAnalysis.html#sklearn.decomposition.FactorAnalysis>