Homework 4
Inference and Representation

Aodong Li
NetID: al5350
al5350@nyu.edu

July 18, 2018

1. (a) The pmf of X_{ij} is

$$P(X_{ij} = x_{ij}) = e^{-WH_{ij}} \frac{WH_{ij}^{x_{ij}}}{x_{ij}!}$$

.

Then the log-likelihood is

$$\begin{aligned} L(W, H) &= \sum_{i,j} \log P(x_{ij}) \\ &= \sum_{i,j} (-WH_{ij} + x_{ij} \log(WH_{ij}) - \log(x_{ij}!)) \\ &\propto \sum_{i,j} (-WH_{ij} + x_{ij} \log(WH_{ij})) \end{aligned}$$

(b)

$$\begin{aligned} &x^{t+1} = \arg\max_x g(x, x^t) \\ \Rightarrow &g(x^{t+1}, x^t) \geq g(x, x^t) \ \forall x \\ \Rightarrow &f(x^{t+1}) \geq g(x^{t+1}, x^t) \geq g(x^t, x^t) = f(x^t) \end{aligned}$$

Thus it is a non-decreasing sequence.

(c)

$$\sum_k y_k = \sum_k c_k \frac{y_k}{c_k}.$$

Because $\sum_k c_k = 1$, this is an average of y_k/c_k. By the concavity of logrithm or the variant of Jensen's inequality,

$$\log(\sum_k y_k) >= \sum_k c_k \log(\frac{y_k}{c_k})$$

(d) Replace the variables in the last section with corresponding variables.

(e)

$$\begin{aligned}
&\sum_{i,j,k} x_{ij} c_{kij} \log(w_{ik} h_{kj}) - w_{ik} h_{kj} \\
=&\sum_{i,j} -WH_{ij} + x_{ij}(\sum_k c_{kij} \log(w_{ik} h_{kj}/c_{kij}) + c_{kij} \log(c_{kij})) \\
\leq&\sum_{i,j} -WH_{ij} + x_{ij}(\log(WH_{ij}) + \sum_k c_{kij} \log(c_{kij})) \\
=&L(W,H) + C
\end{aligned}$$

(f) Set the partial derivative to be zero,

$$\begin{aligned}
\frac{\partial g}{\partial w_{ij}} &= \sum_j x_{ij} c_{kij} \frac{1}{w_{ik}} - \sum_j h_{kj} = 0 \\
w_{ik} &= \frac{\sum_j x_{ij} c_{kij}}{\sum_j h_{kj}}.
\end{aligned}$$

Similarly, we can get

$$h_{kj} = \frac{\sum_i x_{ij} c_{kij}}{\sum_i w_{ki}}.$$

2. (a) After we define an inner product on random variables

$$\langle X, Y \rangle := \mathbb{E}XY,$$

we can use Cauchy-Schwarz inequality and get

$$|\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)| \leq \sqrt{\mathbb{E}(X-\mathbb{E}X)^2\mathbb{E}(Y-\mathbb{E}Y)^2}$$

which is $|\tilde{\sigma}_{ij}| \leq 1$.

(b) The solution of factor analysis are within the covariance matrix

$$\Sigma = AA^\top + D_\epsilon$$

where D_ϵ is the covariance matrix of the noise. If we standardize the covariance matrix into correlation matrix, which is equivalent to standardize the data, the correlation matrix is

$$\tilde{\Sigma} = VAA^\top V^\top + VD_\epsilon V$$

which is just a scaling on all the parameters.

So the solutions are equal when we apply factor analysis on covariance matrix or correlation matrix.

But PCA gives different solutions on different matrix.

In all, factor analysis is modeling the correlation structure while PCA models the covariance structure.

(c) Consider the three independent standard Gaussian random variables Z_1, Z_2, Z_3

$$\begin{aligned} X_1 &\sim Z_1 \\ X_2 &= X_1 + 0.01Z_2 \\ X_3 &= 100Z_3. \end{aligned}$$

PCA will find the first principal component aligns along X_3, while factor analysis finds the leading principal direction $X_1 + X_2$.

(d) Consider the three independent standard Gaussian random variables Z_1, Z_2, Z_3

$$\begin{aligned} X_1 &\sim Z_1 \\ X_2 &= 0.01Z_2 \\ X_3 &= 100Z_3. \end{aligned}$$

PCA will find the first principal component aligns along X_3, while factor analysis fails finding the direction along any variable due to the lack of unicity.

(e) The Gaussian joint likelihood is

$$X \sim \mathcal{N}(0, AA^\top + \mathrm{diag}(\beta))$$

and the log likelihood is

$$\ln P(X|A,\beta) \propto -\frac{N}{2} \ln |AA^\top + \mathrm{diag}(\beta)| - \frac{1}{2} \sum_{n=1}^{N} x^\top (AA^\top + \mathrm{diag}(\beta))^{-1} x.$$

(f) No, the solution is not unique, because A and $AR^\top$ are equivalent in the sense $AA^\top = AR^\top (AR^\top)^\top$

(g) Following the last section and by the facts that $\frac{\mathrm{d}}{\mathrm{d}\,A} \ln |A| = A^{-T}$ and $\frac{\partial}{\partial A} \mathrm{tr}\ AB = \frac{\partial}{\partial A} \mathrm{tr}\ BA = B$, we have

$$\begin{aligned} \ln P(X|A,\beta) &\propto -\frac{N}{2} \ln |AA^\top + \mathrm{diag}(\beta)| - \frac{1}{2} \sum_{n=1}^{N} x^\top (AA^\top + \mathrm{diag}(\beta))^{-1} x \\ &= -\frac{N}{2} \ln |AA^\top + \beta_0 I| - \frac{1}{2} \mathrm{tr}\ ((AA^\top + \beta_0 I)^{-1} \sum_{n=1}^{N} x x^\top). \end{aligned}$$

By taking the derivatives to be zero, we have

$$AA^\top + \beta_0 I = \frac{1}{N} \sum_{n=1}^{N} x^\top x$$