
HOMework 2 : MAP estimation on NMF

INTRODUCTION TO PROBABILISTIC GRAPHICAL MODELS

AUTHORED BY

MARION KARAKOUZIAN
&
YASSINE BENZAOU
&
MEHDI ABBANA BENNANI

PROFESSOR
UMUT SIMSEKLI

Ecole Polytechnique

2017

1 Exercise 1

We have that $\hat{V} = WH$, $w_{fk} \sim G(\alpha_w, \beta_w)$, $h_{kn} \sim G(\alpha_h, \beta_h)$ and $v_{fn}|w_{f,:}, h_{:,n} \sim \mathcal{P}(\langle w_{f,:}, h_{:,n} \rangle)$.

$$\begin{aligned}
\langle W^*, H^* \rangle &= \operatorname{argmax}_{(W,H)} \log(P(W, H|V)) \\
&= \operatorname{argmax}_{(W,H)} \log\left(\frac{P(V|W, H) \times P(W, H)}{P(V)}\right) \\
&= \operatorname{argmax}_{(W,H)} \log(P(V|W, H)) + \log(P(W, H)) \\
&= \operatorname{argmax}_{(W,H)} \sum_{fn} \left(V_{fn} \times \log(\hat{V}_{fn}) - \hat{V}_{fn} \right) + \log(P(W, H))
\end{aligned}$$

$$\begin{aligned}
\log(P(W, H)) &= \sum_{fk} \log P(w_{fk}) + \sum_{kn} \log P(h_{kn}) \\
&= \operatorname{argmax}_{(W,H)} \left(\sum_{fn} V_{fn} \times \log(\hat{V}_{fn}) - \hat{V}_{fn} \right) + \sum_{fk} \log \frac{w_{fk}^{\alpha_w-1} \times \exp(-w_{fk} \times \beta_w) \times \beta_w^{\alpha_w}}{\Gamma(\alpha_w)} \\
&\quad + \sum_{kn} \log \frac{h_{kn}^{\alpha_h-1} \times \exp(-h_{kn} \times \beta_h) \times \beta_h^{\alpha_h}}{\Gamma(\alpha_h)} \\
&= \operatorname{argmax}_{(W,H)} \left[\sum_{fn} (V_{fn} \times \log(\hat{V}_{fn}) - \hat{V}_{fn}) + \sum_{fk} ((\alpha_w - 1) \times \log w_{fk} - w_{fk} \times \beta_w) \right. \\
&\quad \left. + \sum_{kn} ((\alpha_h - 1) \log h_{kn} - \beta_h \times h_{kn}) \right]
\end{aligned}$$

$$\begin{aligned}
S_{fkn}|w_{fk}, h_{kn} &\sim \mathcal{P}(w_{fk}, h_{kn}) \\
V_{fn}|S_{fkn} &\sim \delta(V_{fn} - \sum_k S_{fkn})
\end{aligned}$$

We can apply the EM algorithm, that is composed by two steps :

- E-step : We have to compute

$$\begin{aligned}
\mathcal{L}_t(W, H) &= \mathbb{E}[\log(P(V, S, W, H))]_{P(S|V, H_t, W_t)} \\
&= \mathbb{E}_{S|V, H_t, W_t}[\log P(S, V|W, H) \times P(W, H)] \times P(S|V, H_t, W_t)
\end{aligned}$$

- M-step : $W_{t+1}, H_{t+1} = \operatorname{argmax}_{(W,H)} \mathcal{L}_t(W, H)$

\Rightarrow Application of the algorithm :

1. E-step :

$$\begin{aligned}
\log P(V, S|W, H) &= \log[P(V|S, W, H) \times P(S|W, H)] \\
&= \sum_{fn} \log(\delta(V_{fn} - \sum_k S_{fkn})) + \log P(S|W, H) \\
&= \sum_{fkn} (S_{fkn} \times \log(w_{fk} \times h_{kn}) - w_{fk} \times h_{kn} - \log \Gamma(S_{fkn})) \\
&\quad + \sum_{fn} \log \left(\delta(V_{fn} - \sum_k S_{fkn}) \right)
\end{aligned}$$

$$\begin{aligned}
\log P(W, H) &= \sum_{fk} (\alpha_w - 1) \times \log w_{fk} - w_{fk} \times \beta_w \\
&\quad + \sum_{kn} (\alpha_h - 1) \times \log h_{kn} - h_{kn} \times \beta_h \\
&\quad + \dots(\text{constant})
\end{aligned}$$

According to the course, we have $\log P(S|V_t, W_t, H_t) = \log \prod_{fn} M(S_{fn}, V_{fn}, [\Pi_1, \Pi_k]^{(t)})$

with $\Pi_k^{(t)} = \frac{w_{fk}^{(t)} \times h_{kn}^{(t)}}{\hat{v}_{fn}^{(t)}}$, so we obtain

$$\begin{aligned}
\mathcal{L}_t(W, H) &= \sum_{fkn} \mathbb{E}[S_{fkn}] \times \log(w_{fk} \times h_{kn}) - w_{fk} \times h_{kn} \\
&\quad + \sum_{fk} (\alpha_w - 1) \times \log w_{fk} - w_{fk} \times \beta_w + \sum_{kn} \log h_{kn} - h_{kn} \times \beta_h \\
&\quad + \text{constant}
\end{aligned}$$

2. M-step : To compute the argmax of \mathcal{L}_t , we use the derivative :

$$\frac{\partial \mathcal{L}_t}{\partial w_{fk}} = \sum_n \left(\frac{\mathbb{E}[S_{fkn}]}{w_{fk}} - h_{kn}^{(t)} \right) + \frac{\alpha_w - 1}{w_{fk}} - \beta_w = 0$$

Then, we obtain

$$\begin{aligned}
w_{fk}^{(t+1)} &= \frac{\sum_n \mathbb{E}[S_{fkn}] + \alpha_w - 1}{\beta_w + \sum_n h_{kn}^{(t)}} \\
&= \frac{\sum_n \left(v_{fn} \times \frac{w_{fk}^{(t)} \times h_{kn}^{(t)}}{\hat{v}_{fn}^{(t)}} \right) + \alpha_w - 1}{\beta_w + \sum_n h_{kn}^{(t)}} \\
&= \frac{w_{fk}^{(t)} \times \sum_n \frac{v_{fn}}{\hat{v}_{fn}^{(t)}} \times h_{kn}^{(t)} + \alpha_w - 1}{\beta_w + \sum_n h_{kn}^{(t)}}
\end{aligned}$$

$$\text{Likewise for h, } h_{kn}^{(t+1)} = \frac{h_{kn}^{(t)} \times \sum_f \frac{v_{fn}}{\hat{v}_{fn}^{(t)}} \times w_{fk}^{(t+1)} + \alpha_h - 1}{\beta_h + \sum_f w_{fk}^{(t+1)}}$$

Finally, we have $\hat{V} = WH$, so,

$$\begin{aligned}
W &\leftarrow \frac{W \cdot [(V/\hat{V})H^T] + (Q_w - 1)O_{FK}}{O_{FN}H^T + \beta_w O_{FK}} \\
H &\leftarrow \frac{H \cdot [W^T(V/\hat{V})] + (Q_h - 1)O_{KN}}{W^T O_{FN} + \beta_h O_{KN}}
\end{aligned}$$

with \cdot the element-wise matrix multiplication and $/$ the element-wise division.

where $O_{KN} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$ with K lines and N columns.

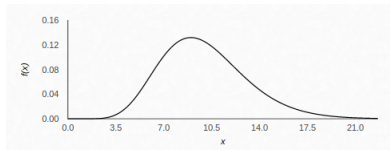
2 Exercise 2

Question 1 Please refer to the "EM_algorithm.m" matlab code.

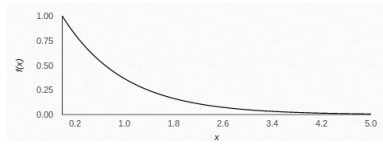
Question 2 We can interpret the W matrix as the projection of the faces of the V matrix on a K dimensional space. Therefore every column of W is a 'feature' or 'base' face. Then the V matrix corresponds to the weightenings between the K faces to obtain the original faces.

The following figures show the Γ distribution for different values of α and $\beta = 1$.

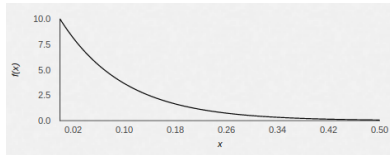
Therefore, we interpret the β_W parameter as a sparsity parameter. By putting a strong prior in some region, we can tune the sparsity of the image. For the α_W parameter, we interpret it as the strenght of the sparsity (?)



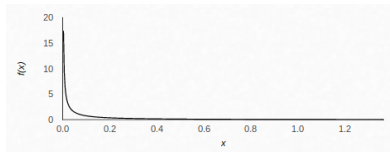
(a) $\alpha = 1, \beta = 0.1$



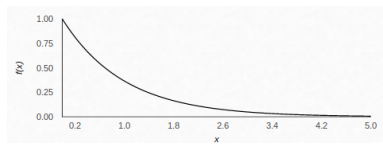
(b) $\alpha = 1, \beta = 1$



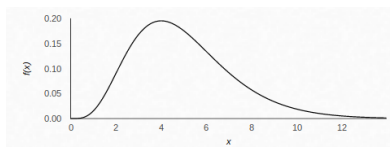
(c) $\alpha = 1, \beta = 10$



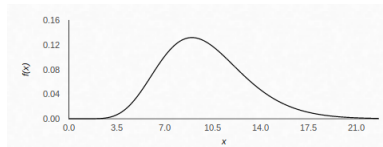
(d) $\alpha = 0.1, \beta = 1$



(e) $\alpha = 1, \beta = 1$



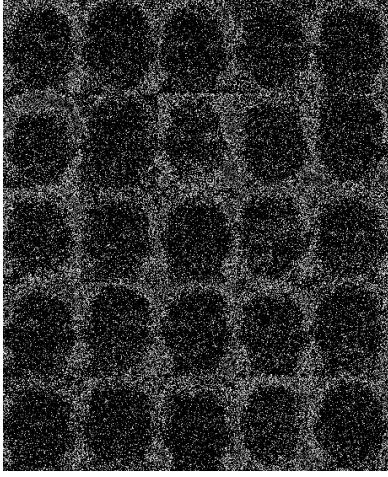
(f) $\alpha = 5, \beta = 1$



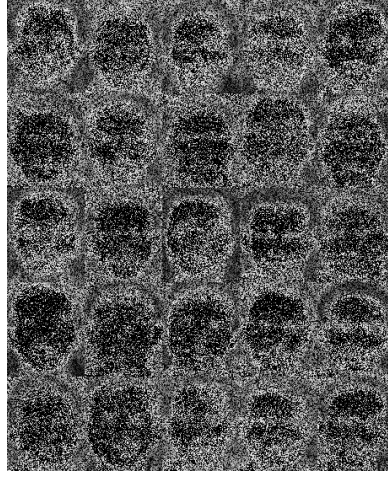
(g) $\alpha = 10, \beta = 1$

Figure 1: Gamma distribution for different values of α and β

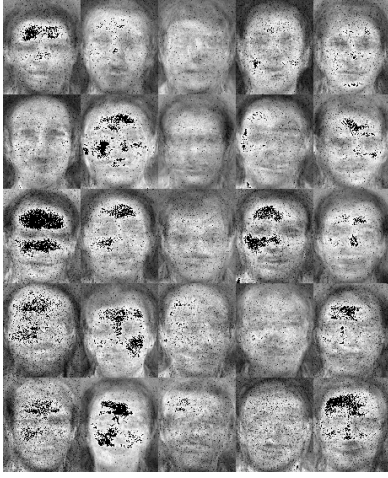
The following figure shows the most relevant values of α , for $\beta_w = 1$ and $\beta_h = 1$



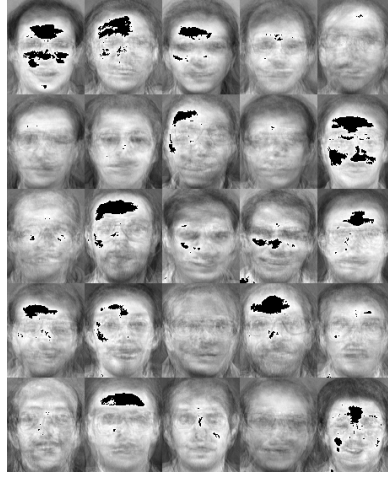
(a) $\alpha_W = 0.1, \alpha_H = 0.1$



(b) $\alpha_W = 5, \alpha_H = 0.1$

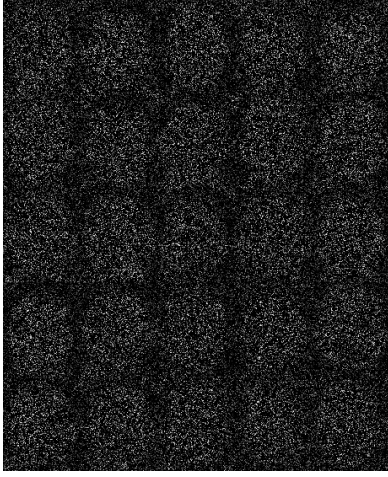


(c) $\alpha_W = 0.1, \alpha_H = 10$

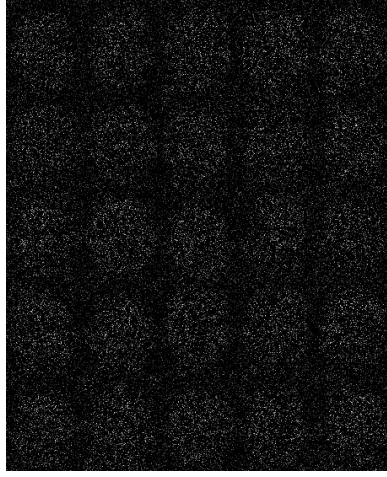


(d) $\alpha_W = 5, \alpha_H = 0.1$

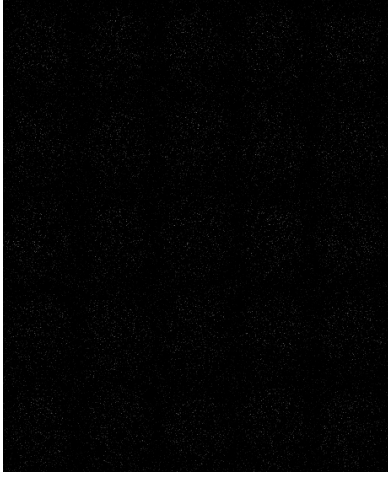
Figure 2: W matrix for $\beta_w = 1$ and $\beta_h = 1$



(a) $\alpha_W = 1, \alpha_H = 5$



(b) $\alpha_W = 1, \alpha_H = 10$



(c) $\alpha_W = 5, \alpha_H = 5$



(d) $\alpha_W = 5, \alpha_H = 0.1$

Figure 3: W matrix for $\beta_w = 1$ and $\beta_h = 1$

3. Let's print the results for different values of alpha. The following figure shows the most relevant values of β , for $\alpha_w = 1$ and $\alpha_h = 1$