HOMEWORK 2 : MAP estimation on NMF

Introduction to Probabilistic Graphical Models

Authored by

marion karakouzian
&
yassine benazzou
&
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 $v_{fn}|w_{f,:}, h_{:,n} \sim \mathcal{P}(\langle w_{f,:}, h_{:,n} \rangle)$.

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

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

$S_{fnk}|w_{fk}, h_{kn} \sim \mathcal{P}(w_{fk}, h_{kn})$
$V_{fn}|S_{fnk} \sim \delta(V_{fn} - \sum_k S_{fnk})$

We can apply the EM algorith, 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 : $W_{t+1}, H_{t+1} = \underset{(W,H)}{\text{argmax}} \mathcal{L}_t(W, H)$

$\Rightarrow$ Application of the algorithm :

1. E-step :

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

$$\log P(W,H) = \sum_{fk} (\alpha_w - 1) \times \log w_{fk} - w_{fk} \times \beta_w$$
$$+ \sum_{kn} (\alpha_h - 1) \times \log h_{kn} - h_{kn} \times \beta_h$$
$$+ ...(constant)$$

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

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

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

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

Then, we obtain

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

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,

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

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

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

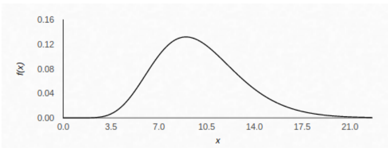
2 Exercise 2

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

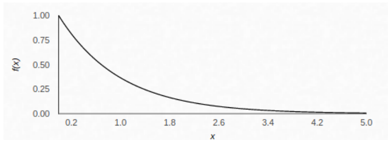
Question 2 We can interpret the W matrix as the projection of the faces of the V matrix on a K dimensional space. Therefore every column of W is a 'feature' or 'base' face. Then the V matrix corresponds to the weightenings between the K faces to obtain the original faces.

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

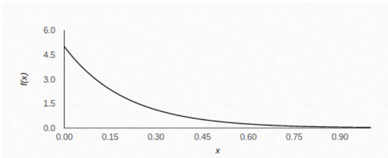
Therefore, we interpret the β_W parameter as a sparsity parameter. We can tune β_W for some specific expected value. We interpret the α_W parameter as a variance parameter, therefore we control the range of values with a high probability.



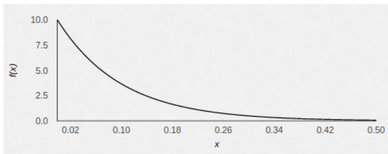
(a) $\alpha = 1$, $\beta = 0.1$



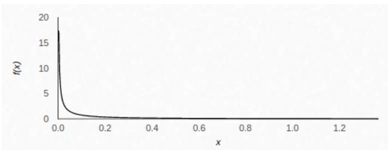
(b) $\alpha = 1$, $\beta = 1$



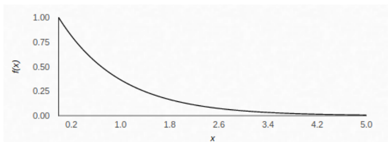
(c) $\alpha = 1$, $\beta = 5$



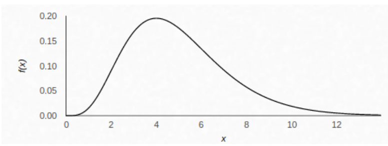
(d) $\alpha = 1$, $\beta = 10$



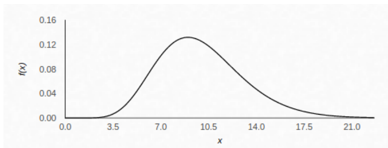
(e) $\alpha = 0.1$, $\beta = 1$



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



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



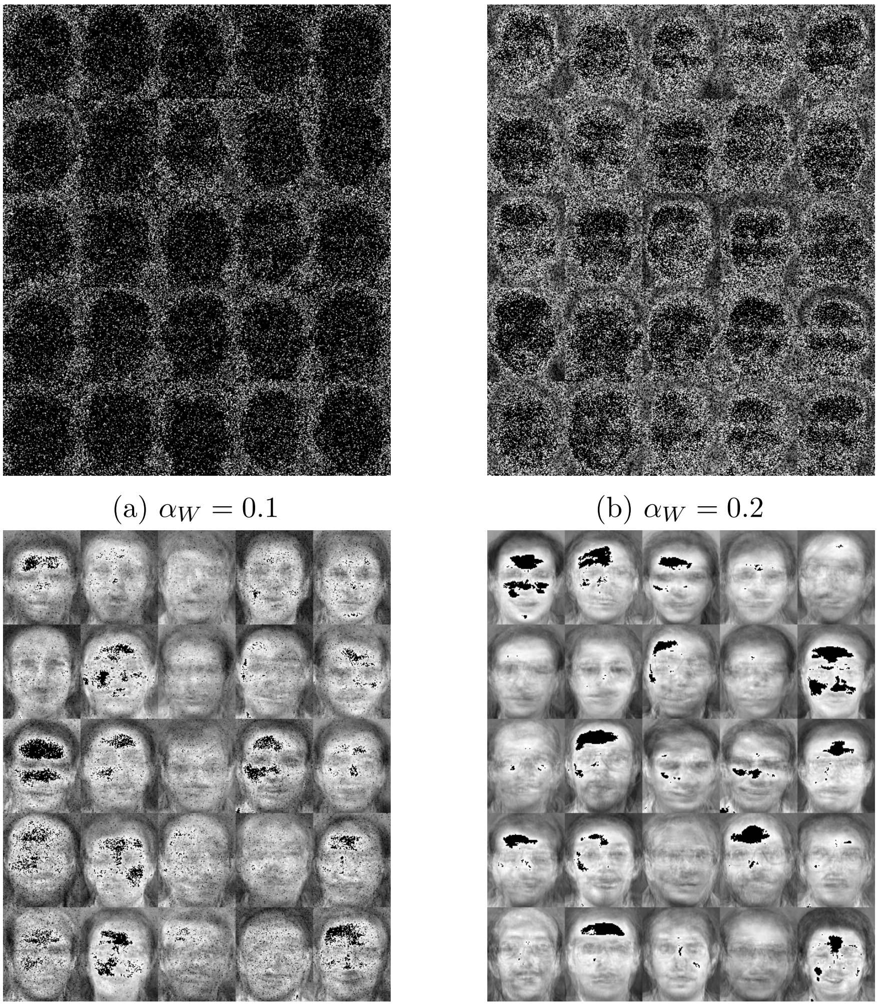
(h) $\alpha = 10$, $\beta = 1$

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

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

In the first figure, we see that increasing α_W decreases the sparcity of the W representation, because the range of the possible values with a higher probability increases.

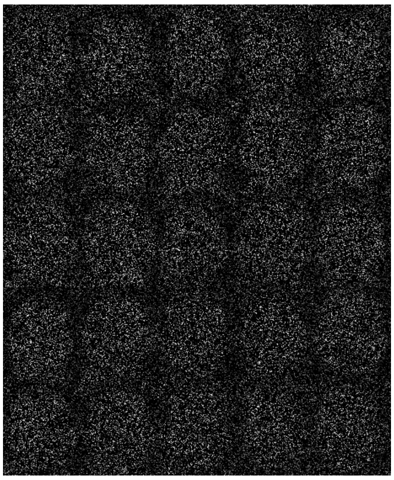
In the second figure, we see that increasing α_H increases the sparsity of the solution. (The weights get bigger the bigger α_H I expected the opposite)



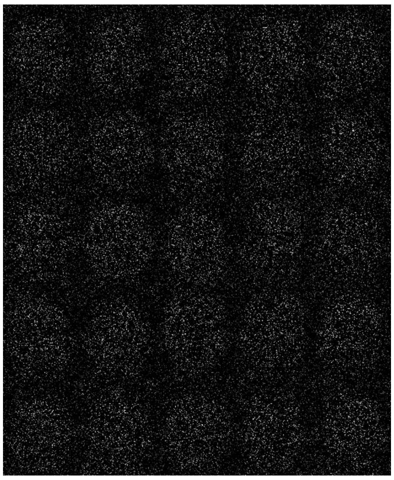
(a) $\alpha_W = 0.1$ (b) $\alpha_W = 0.2$

(c) $\alpha_W = 1$ (d) $\alpha_W = 10$

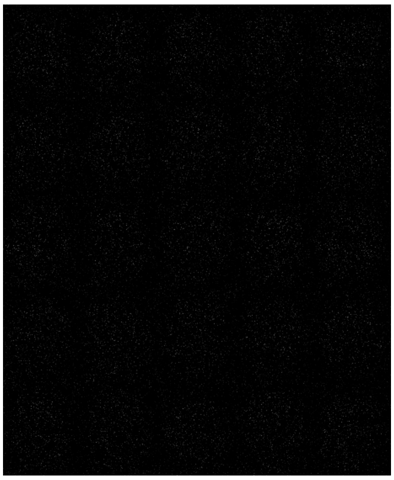
Figure 2: W matrix for $\beta_w = 1$ and $\beta_h = 1$, $\alpha_H = 0.1$



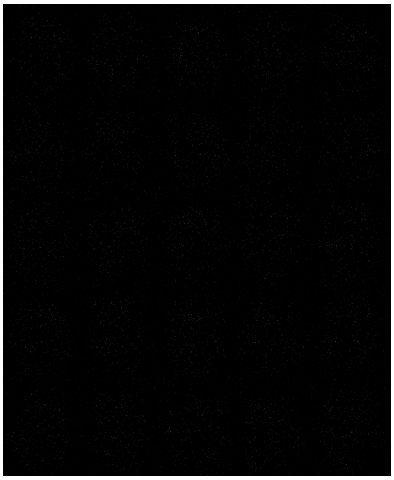
(a) $\alpha_W = 0.1$, $\alpha_H = 0.5$



(b) $\alpha_W = 0.1$, $\alpha_H = 1$



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



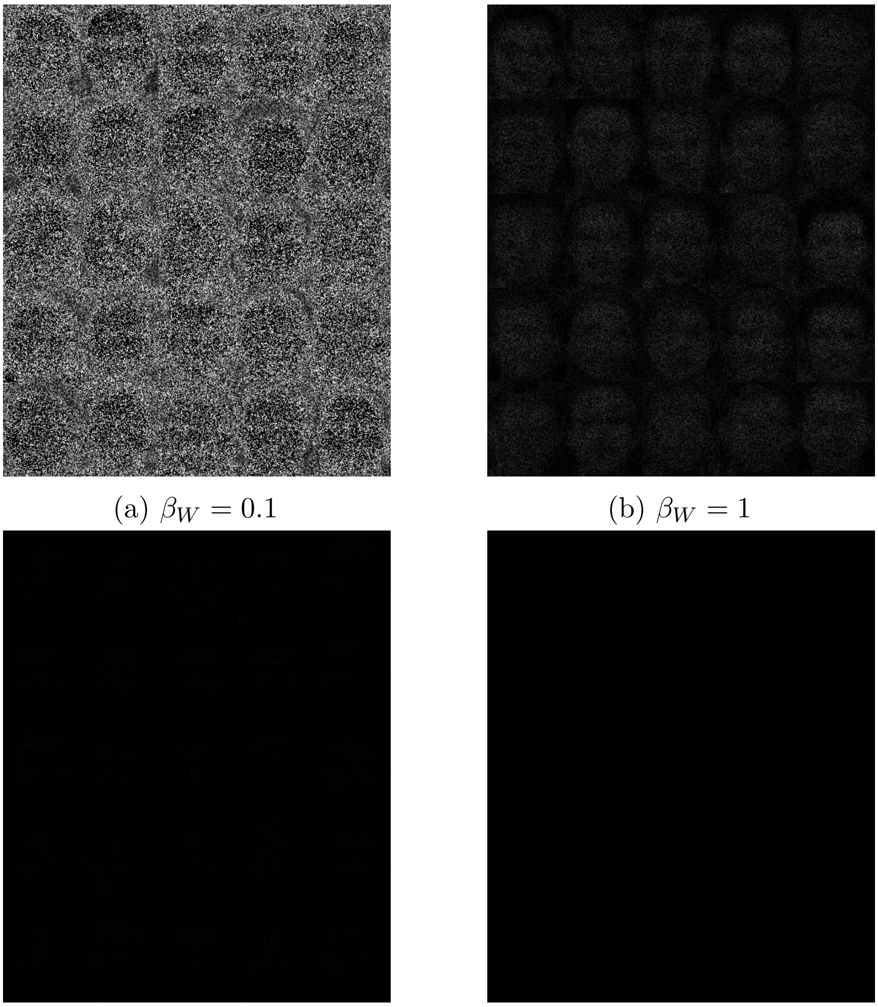
(d) $\alpha_W = 0.1$, $\alpha_H = 10$

Figure 3: W matrix for $\alpha_w = 1$ and $\alpha_h = 1$ and $\beta_W = 0.1$

Question 3 The following figure shows the most relevant values of β, for $\alpha_w = 1$ and $\alpha_h = 1$

In the first figure, as expected, by increasing β_W, the expectation of the prior decreases therefore the sparsity increases.

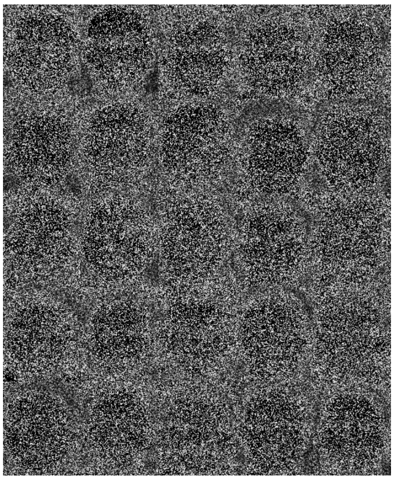
In the second figure, by increasing β_H the expectation of the weight decreases therefore the output value of the pixel decreases and the sparsity increases.



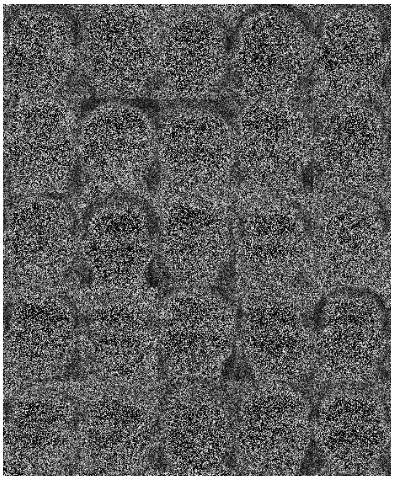
(a) $\beta_W = 0.1$ (b) $\beta_W = 1$

(c) $\beta_W = 5$ (d) $\beta_W = 10$

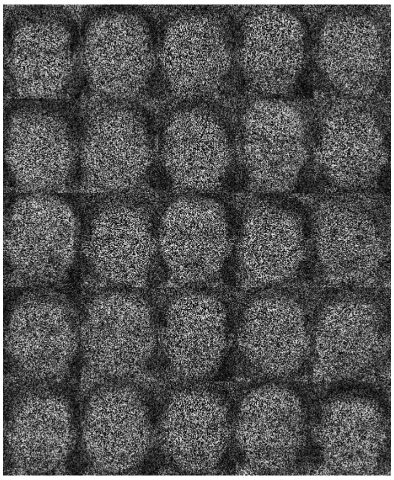
Figure 4: W matrix for $\alpha_w = 1$ and $\alpha_h = 1$ and $\beta_H = 0.1$



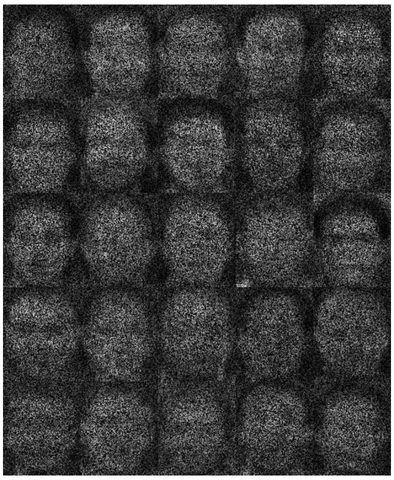
(a) $\beta_H = 0.1$



(b) $\beta_H = 1$



(c) $\beta_H = 5$



(d) $\beta_H = 10$

Figure 5: W matrix for $\alpha_w = 1$ and $\alpha_h = 1$ and $\beta_W = 0.1$