Homework 2

Fubang ZHAO
fubang.zhao@polytechnique.edu

November 7, 2017

MAP estimation on NMF

Given the following probabilistic non-negative matrix factorization (NMF) model: (for f = 1, . . . , F , n = 1, . . . , N , k = 1,...,K)

$$w_{f,k} \sim \mathcal{G}(w_{f,k}; \alpha_w, \beta_w)$$
$$h_{k,n} \sim \mathcal{G}(h_{k,n}; \alpha_h, \beta_h)$$
$$v_{f,n}|w_{f,:}, h_{:,n} \sim \mathcal{PO}(v_{f,n}; \sum_{k=1}^{K} w_{f,k}h_{k,n})$$

where $\mathcal{G}$ and $\mathcal{PO}$ denote the gamma and the Poisson distributions, respectively.

Question 1

Derive an Expectation-Maximization algorithm for finding the maximum a-posteriori estimate (MAP), defined as follows:

$$(W^*, H^*) = \underset{W,H}{argmax} \log p(W, H|V) \tag{1}$$

Here, we define a latent random variable:

$$s_{f,k,n} \sim \mathcal{PO}(s_{f,k,n}; w_{f,k}h_{k,n}), \quad v_{f,n} = \sum_{k} s_{f,k,n} \tag{2}$$

Then we will transform the problem by introducing $S = [s_{fkn}]_{f,k,n}$:

$$
\begin{aligned}
(W^*, H^*) &= \underset{W,H}{argmax} \log p(W, H|V) \\
&= \underset{W,H}{argmax} \log \frac{p(V|W, H)p(W, H)}{p(V)} \\
&= \underset{W,H}{argmax} \log p(V|W, H)p(W, H) \\
&= \underset{W,H}{argmax} \sum_S \log p(V, S|W, H) + \log p(W, H) \\
&= \underset{W,H}{argmax} \sum_S \log p(V|S)p(S|W, H) + \log p(W, H)
\end{aligned}
$$

And $v = s_1 + s_2 + ... + s_K$, then:

$$
\begin{aligned}
\log p(V|W, H) &= \log \sum_S p(V|S)p(S|W, H) \\
&= \log \prod_{f,n} \mathcal{PO}(v_{f,n}; \sum_k w_{f,k}h_{k,n}) \\
&= \sum_f \sum_n (v_{f,n} \log[WH]_{f,n} - [WH]_{f,n} - \log \Gamma(v_{f,n} + 1))
\end{aligned}
$$

Then, according to the Jensen's inequality, we define:

$$
\begin{aligned}
\mathcal{L}_V(W, H) &= \log \sum_S p(V|S)p(S|W, H) \\
&\geq \sum_S q(S) \log \frac{p(V, S|W, H)}{q(S)} \\
&= \mathcal{B}_{EM}[q]
\end{aligned}
$$

To make the bound tight for a particular value of θ, we need for the step involving Jensens inequality in our derivation above to hold with equality. For this to be true, we know it is sufficient that that the expectation be taken over a constant-valued random variable. It means:

$$
q(S) \propto p(V, S|W, H) \tag{3}
$$

Since we know that $\sum q(S) = 1$, we can get that:

$$
\begin{aligned}
\underset{q(S)}{argmax} \mathcal{B}_{EM}[q] &= \frac{p(V, S|W, H)}{\sum_S p(V, S|W, H)} \\
&= \frac{p(V, S|W, H)}{p(V|W, H)} \\
&= p(S|V, W, H)
\end{aligned}
$$

Here,

$$
\begin{aligned}
\underset{W,H}{argmax}\mathcal{B}_{EM} &= \underset{W,H}{argmax}\sum_S q(S)\log\frac{p(V,S|W,H)}{q(S)} \\
&= \underset{W,H}{argmax}\sum_S q(S)\log p(V,S|W,H) \\
&= \underset{W,H}{argmax}\mathbb{E}[logp(V,S|W,H)]_{q(S)^{(n)}}
\end{aligned}
$$

Hence the problem can be maximized iteratively as follows:

$$
\begin{aligned}
&E\ step \quad q(S)^{(n)} = p(S|V,W^{n-1},H^{n-1}) \\
&M\ step \quad (W^n,H^n) = \underset{W,H}{argmax}\mathbb{E}[logp(V,S|W,H)]_{q(S)^{(n)}} + \log p(W,H)
\end{aligned}
$$

The E Step

To derive the posterior of the latent sources, we observe that

$$
p(S|V,W,H) = \frac{p(V,S|W,H)}{p(V|W,H)} \tag{4}
$$

For this model, we can get:

$$
\begin{aligned}
&\log p(V,S|W,H) \\
&= \sum_f\sum_n\left(\sum_k(-w_{f,k}h_{k,n} + s_{f,k,n}\log(w_{f,k}h_{k,n}) - log\Gamma(s_{f,k,n}+1) + \log\delta(v_{f,n}-\sum_k s_{f,k,n}))\right)
\end{aligned}
$$

According to (4), we can get:

$$
\begin{aligned}
&\log p(S|V,W,H) \\
&= \sum_f\sum_n\left(\sum_k(s_{f,k,n}log\frac{w_{f,k}h_{k,n}}{\sum_{k'}w_{f,k'}h_{k',n}} - log\Gamma(s_{f,k,n}+1) + log\ \Gamma(v_{f,n}+1) + log\delta(v_{f,n}-\sum_k s_{f,k,n}))\right) \\
&= \sum_f\sum_n log\ \mathcal{M}(s_{f,1,n},...s_{f,K,n};v_{f,n},p_{f,1,n},...,p_{f,K,n})
\end{aligned}
$$

where $p_{f,k,n} = w_{f,k}h_{k,n}/\sum_{k'}w_{f,k'}h_{k',n}$ are the cell probabilities. Here, $\mathcal{M}$ denotes a **multinomial distribution** defined by

$$
\begin{aligned}
\mathcal{M}(\mathbf{s};v;\mathbf{p}) &= \binom{v}{s_1s_2\cdots s_K}p_1^{s_1}p_2^{s_2}\cdots p_K^{s_K}\delta(v-\sum_k s_k) \\
&= \delta(v-\sum_k s_k)v!\prod_{k=1}^K\frac{p_k^{s_k}}{s_k!}
\end{aligned}
$$

According to the definition of multinomial distribution, we can get the marginal mean:

$$
\mathbb{E}[s_k] = vp_k \tag{5}
$$

The M Step

We know that the Gamma distribution is

$$\log \mathcal{G}(x;\alpha,\beta) = (\alpha-1)\log x - \frac{x}{\beta} - \log\Gamma(\alpha) - \alpha\log\beta \tag{6}$$

And because W and H are independent, $\log p(W,H) = \log p(W) + \log p(H)$

$$\begin{aligned}
&\mathbb{E}[logp(V,S|W,H)]_{q(S)^{(n)}} + \log p(W,H) \\
&= \sum_f \sum_n \left(\sum_k (-w_{f,k}h_{k,n} + \mathbb{E}[s_{f,k,n}] log(w_{f,k}h_{k,n}) - log\Gamma(\mathbb{E}[s_{f,k,n}]+1) + log\delta(v_{f,n} - \sum_k \mathbb{E}[s_{f,k,n}])) \right) \\
&+ \sum_f \sum_k \left((\alpha_w - 1)\ log\ w_{f,k} - \frac{w_{f,k}}{\beta_w} - \Gamma(\alpha_w) - \alpha_w\ log(\beta_w) \right) \\
&+ \sum_n \sum_k \left((\alpha_h - 1)\ log\ h_{k,n} - \frac{h_{k,n}}{\beta_h} - \Gamma(\alpha_h) - \alpha_h\ log(\beta_h) \right)
\end{aligned}$$

where $\mathbb{E}[s_{f,k,n}] = v_{f,n}p_{f,k,n}$, and we only need to maximize the simpler objective:

$$\begin{aligned}
\mathcal{L}(W,H) = &\sum_f \sum_n \left(\sum_k (-w_{f,k}h_{k,n} + \mathbb{E}[s_{f,k,n}] log(w_{f,k}h_{k,n})) \right) \\
&+ \sum_f \sum_k \left((\alpha_w - 1)\ log\ w_{f,k} - \frac{w_{f,k}}{\beta_w}) \right) \\
&+ \sum_n \sum_k \left((\alpha_h - 1)\ log\ h_{k,n} - \frac{h_{k,n}}{\beta_h}) \right)
\end{aligned}$$

and the solution is given by the following equations

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_{f,k}} &= -\sum_n h^i_{k,n} + \frac{\sum_n \mathbb{E}[s^i_{f,k,n}] + (\alpha_w - 1)}{w_{f,k}} - \frac{1}{\beta_w} \\
w^{i+1}_{f,k} &= \frac{w^i_{f,k} \sum_n \frac{v_{f,n}h^i_{k,n}}{\sum_{k'} w^i_{f,k'}h^i_{k',n}} + (\alpha_w - 1)}{1/\beta_w + \sum_n h^i_{k,n}} \\
\frac{\partial \mathcal{L}}{\partial h_{k,n}} &= -\sum_f w^i_{f,k} + \frac{\sum_f \mathbb{E}[s^i_{f,k,n}] + (\alpha_h - 1)}{h_{k,n}} - \frac{1}{\beta_h} \\
h^{i+1}_{kn} &= \frac{h^i_{k,n} \sum_f \frac{v_{f,n}w^i_{f,k}}{\sum_{k'} w^i_{f,k'}h^i_{k',n}} + (\alpha_h - 1)}{1/\beta_h + \sum_f w^i_{f,k}}
\end{aligned}$$

Question 2

1. The code is in the folder with the name ”*hw2_Intro_graphic.ipynb*”.

2. The result is shown below:

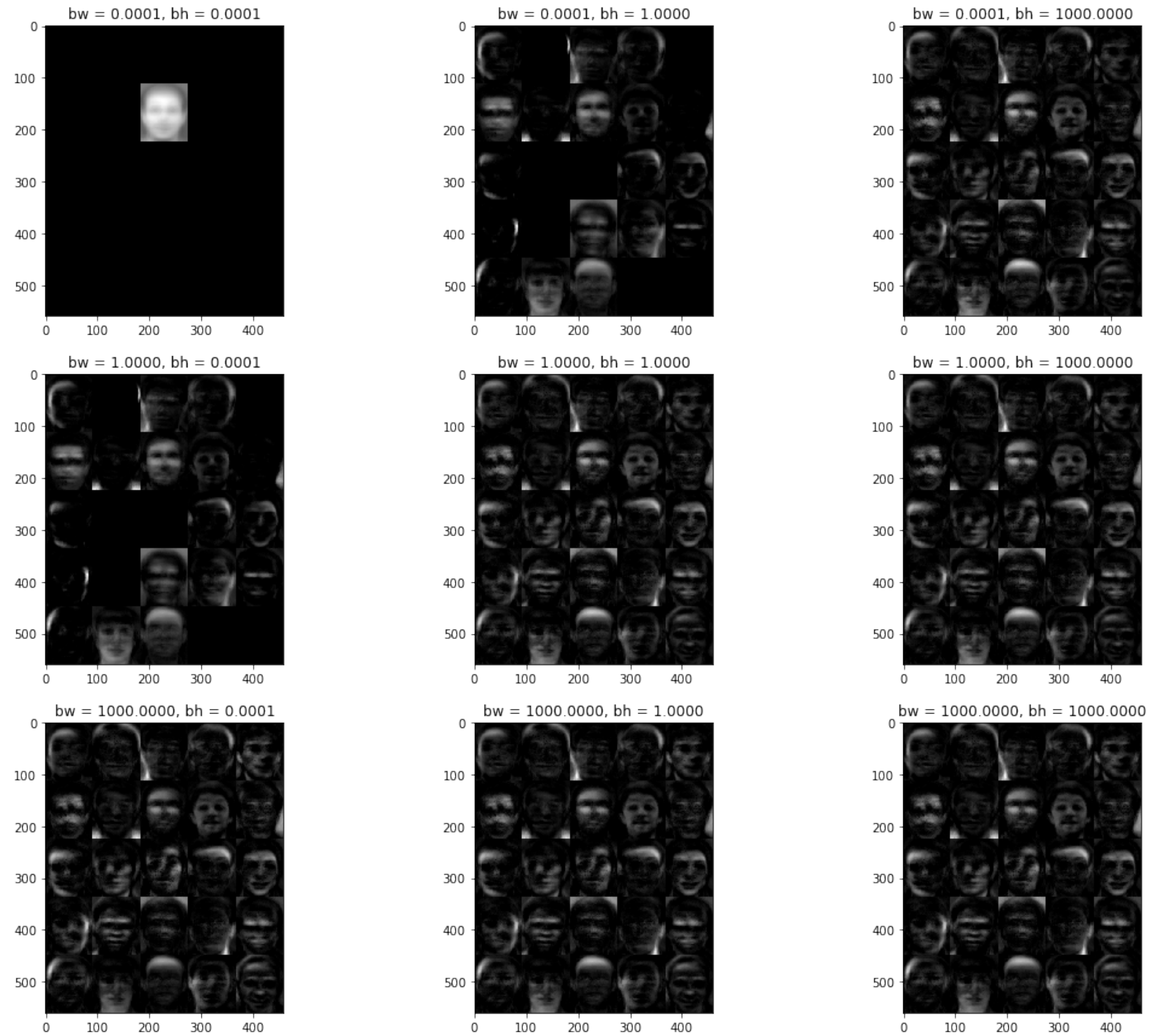

Figure 1: The different W with the changes of aw and ah

Here, we set the range of $\beta_w, \beta_h$is: $[0.001,\ 1,\ 1000]$. According to the result, we find that:

a. when $\beta_w < 1$ *or* $\beta_h < 1$, the characters of faces are not that clear.

b. when $\beta_w >= 1$ *or* $\beta_h >= 1$, the differences between the results are very little, and are much clearer than those with smaller β_w, β_h.

3. The result is shown below:

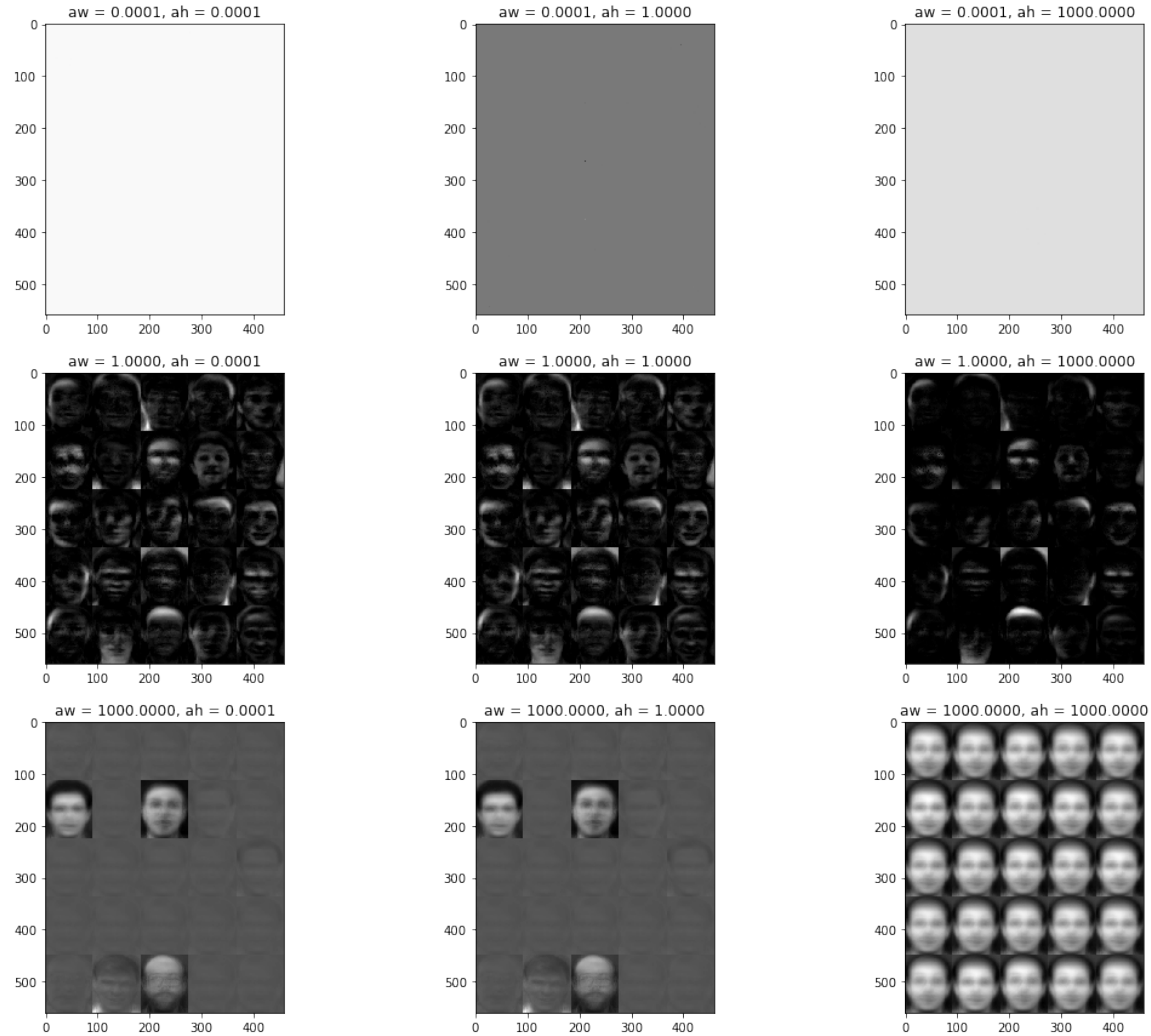

Figure 2: The different W with the changes of bw and bh

Here, we set the range of α_w, α_h is: [0.001, 1, 1000]. According to the result, we find that:

a. when $\alpha_w = 0.0001$, the characters of faces almost disappear. It is almost no character left.

b. when $\alpha_w = 1000$, the characters are also not clear. Especially, when $\alpha_w = \alpha_h = 1000$, the 25 faces are almost the same one.

c. When $\alpha_w = \alpha_h = 1$, the result is the best. The characters of faces are the most recognizable.