Introduction to probabilistic graphical model

Done by MEJRI Aymen

November 2018

1 Question 1 :

1. Draw the directed graphical model :

Answer:

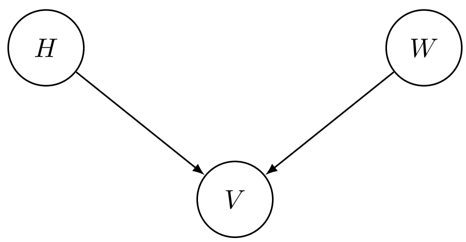

Figure 1: Directed graphical model using matrix representation

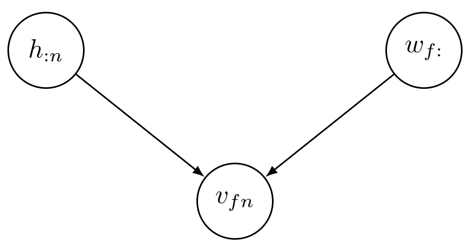

Figure 2: Directed graphical model represented with components

with

$$w_{f:} = \{w_{fk}\}_{k=1}^{K} \text{ and } h_{:n} = \{h_{kn}\}_{k=1}^{K}$$

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

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

Where V , W , and H are the matrices with the form: $V = [v_{fn}]_{f,n}$, $W = [w_{fk}]_{f,k}$, $H = [h_{kn}]_{k,n}$.

Define auxiliary latent random variables (i.e. data augmentation) if necessary (in that case draw the new graphical model).

You need to end up with some 'multiplicative update rules'. Show all your work.

Answer:

Let $V = (v_{fn})_{f,n}$, $W = (w_{fk})_{f,k}$, $H = (h_{kn})_{k,n}$ with $w_{f,:} = \{w_{fk}\}_{k=1}^K$ and $h_{:,n} = \{h_{kn}\}_{k=1}^K$

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

- **First step :** We would have to find $S = (S_{fn})_{f\in\{1..F\},n\in\{1..N\}}$ the latent random variable such that $P(S|V, W, H)$ is tractable:

  We consider $\forall f, n \in$\{1..F\}$\times\{1..N\}$:

  $$S_{f,n} = (s_{f,n,1}, .., s_{f,n,K}) = \{s_{f,n,k}\}_{k=1}^K$$

  such that :

  $$s_{f,n,k}|w_{f,k}, h_{k,n} \sim \mathcal{PO}(s_{f,n,k}, w_{f,k}h_{k,n})$$

  and :

  $$v_{f,n}|S_{f,n} \sim \delta(v_{f,n} - \sum_{k=1}^{K} s_{f,n,k})$$

  The directed graph becomes :

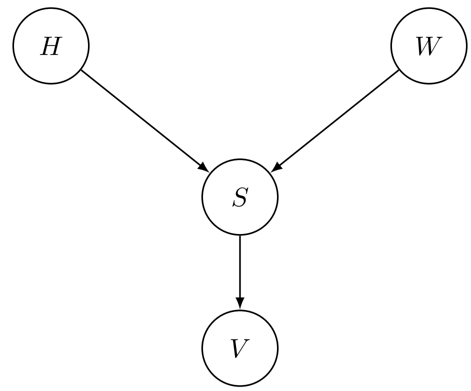

Figure 3: The directed graph with the latent variable

$$P(S|V,W,H) = \prod_{f=1}^{F}\prod_{n=1}^{N} P(S_{f,n}|v_{f,n}, w_{f,:}h_{:,n})$$

$$P(S_{f,n}|v_{f,n}, w_{f,:}h_{:,n}) = \frac{P(S_{f,n}, v_{f,n}, w_{f,:}h_{:,n})}{P(v_{f,n}, w_{f,:}h_{:,n})}$$

We know that :

$$P(S_{f,n}, v_{f,n}, w_{f,:}h_{:,n}) = P(w_{f,:}h_{:,n})P(S_{f,n}|w_{f,:}h_{:,n})P(v_{f,n}|S_{f,n}, w_{f,:}h_{:,n})$$

We can deduce from the graph that:

$$P(S_{f,n}, v_{f,n}, w_{f,:}h_{:,n}) = P(w_{f,:}h_{:,n})P(S_{f,n}|w_{f,:}h_{:,n})P(v_{f,n}|S_{f,n})$$

so:

$$\begin{aligned}
P(S_{f,n}|v_{f,n}, w_{f,:}h_{:,n}) &= \frac{P(S_{f,n}, v_{f,n}, w_{f,:}h_{:,n})}{P(v_{f,n}, w_{f,:}h_{:,n})} \\
&= \frac{P(w_{f,:}h_{:,n})P(S_{f,n}|w_{f,:}h_{:,n})P(v_{f,n}|S_{f,n})}{P(v_{f,n}, w_{f,:}h_{:,n})} \\
&= \frac{P(S_{f,n}|w_{f,:}h_{:,n})P(v_{f,n}|S_{f,n})}{P(v_{f,n}|w_{f,:}h_{:,n})}
\end{aligned}$$

Let's note: $\hat{v}_{fn} = \sum_{k=1}^{K} w_{f,k}h_{k,n}$

We have:

$$
\begin{aligned}
P(S_{f,n}|w_{f,:}h_{:,n}) =& P(s_{f,n,1},..s_{f,n,K}|w_{f,:}h_{:,n}) \\
=& \prod_{k=1}^{K} P(s_{f,n,k}|w_{f,k}h_{k,n}) \\
=& \prod_{k=1}^{K} \frac{(w_{f,k}h_{k,n})^{s_{fnk}}}{s_{fnk}!} \exp\left(-w_{f,k}h_{k,n}\right)
\end{aligned}
$$

and :

$$
P(v_{f,n}|w_{f,:}h_{:,n}) = \frac{(\hat{v}_{f,n})^{v_{fn}}}{v_{fn}!} \exp\left(-\hat{v}_{fn}\right)
$$

then :

$$
P(S_{f,n}|v_{fn}, w_{f,:}h_{:,n}) = \frac{v_{fn}!}{\prod_{k=1}^{K} s_{fnk}!} (\prod_{k=1}^{K} (\frac{w_{f,k}h_{k,n}}{\hat{v}_{fn}})^{s_{fnk}}) \delta(v_{f,n} - \sum_{k=1}^{K} s_{f,n,k})
$$

In the case of $v_{fn} \neq \sum_{k=1}^{K} s_{f,n,k}$:

$$
P(S_{f,n}|v_{fn}, w_{f,:}h_{:,n}) = 0
$$

Now,let's assume that $v_{fn} = \sum_{k=1}^{K} s_{f,n,k}$. We can conclude that :

$$
P(S_{f,n}|v_{fn}, w_{f,:}h_{:,n}) = \frac{v_{fn}!}{\prod_{k=1}^{K} s_{fnk}!} (\prod_{k=1}^{K} (\frac{w_{f,k}h_{k,n}}{\hat{v}_{fn}})^{s_{fnk}})
$$

$$
S_{fn}|v_{f,n}, w_{f,:}h_{:,n} \sim \mathcal{M}(v_{fn}, p_1, p_2, .., p_k)
$$

with $p_k = \frac{w_{f,k}h_{k,n}}{\hat{v}_{fn}} \forall k \in \{1..K\}$

$$
\begin{aligned}
P(S|V, W, H) &= \prod_{f=1}^{F} \prod_{n=1}^{N} P(S_{f,n}|v_{f,n}, w_{f,:}h_{:,n}) \\
&= \prod_{f=1}^{F} \prod_{n=1}^{N} \frac{v_{fn}!}{\prod_{k=1}^{K} s_{fnk}!} \prod_{k=1}^{K} (\frac{w_{f,k}h_{k,n}}{\hat{v}_{fn}})^{s_{fnk}}
\end{aligned}
$$

- **Second step:** this step corresponds to the *E-step*.
  We have to compute $\mathcal{L}_t(W,H) = \mathbf{E}(\log(P(S,V,W,H))_{P(S|V,W^t,H^t)}$

$$\begin{aligned}
\mathcal{L}_t(W,H) =& \mathbf{E}(\log(P(S,V,W,H))_{P(S|V,W^t,H^t)} \\
=& \mathbf{E}(\log(P(W,H) + \log(P(S|W,H) + log(P(V|S))_{P(S|V,W^t,H^t)} \\
=^+& \mathbf{E}(\log(P(S|W,H))_{P(S|V,W^t,H^t)} + \log(P(W,H)) \\
=^+& \mathbf{E}(\log(P(S|W,H))_{P(S|V,W^t,H^t)} + \log(P(W)P(H)) \\
=^+& \mathbf{E}(\log(P(S|W,H))_{P(S|V,W^t,H^t)} + \log(P(W)) + \log(P(H))
\end{aligned}$$

but:

$$\begin{aligned}
\log(P(S|W,H)) &= \log(\prod_{f=1}^{F}\prod_{n=1}^{N} P(S_{f,n}|w_{f,:}h_{:,n})) \\
&= \sum_{f=1}^{F}\sum_{n=1}^{N} \log(P(S_{f,n}|w_{f,:}h_{:,n})) \\
&= \sum_{f=1}^{F}\sum_{n=1}^{N}\sum_{k=1}^{K} \log(P(s_{f,n,k}|w_{f,k}h_{k,n})) \\
&= \sum_{f=1}^{F}\sum_{n=1}^{N}\sum_{k=1}^{K} \log(\frac{(w_{f,k}h_{k,n})^{s_{f,n,k}}}{s_{f,n,k}!}\exp(-w_{f,k}h_{k,n})) \\
&= \sum_{f=1}^{F}\sum_{n=1}^{N}\sum_{k=1}^{K} s_{f,n,k}\log(w_{f,k}h_{k,n}) - \log(s_{f,n,k}!) - w_{f,k}h_{k,n} \\
&=^+ \sum_{f=1}^{F}\sum_{n=1}^{N}\sum_{k=1}^{K} s_{f,n,k}\log(w_{f,k}h_{k,n}) - w_{f,k}h_{k,n}
\end{aligned}$$

As we know:

$$h_{kn} \sim \mathcal{G}(h_{k,n}, \alpha_h, \beta_h)$$
$$w_{fk} \sim \mathcal{G}(W_{f,k}, \alpha_w, \beta_w)$$

Then:

$$\begin{aligned}
\log P(H) &= \sum_{n=1}^{N}\sum_{k=1}^{K} log(P(h_{k,n})) \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} (\alpha_h - 1)\log(h_{k,n}) + \alpha_h\log(\beta_h) - \beta_h h_{k,n} \\
&=^+ \sum_{n=1}^{N}\sum_{k=1}^{K} (\alpha_h - 1)\log(h_{k,n}) - \beta_h h_{k,n}
\end{aligned}$$

$$\log P(W) = \sum_{f=1}^{F}\sum_{k=1}^{K} log(P(w_{f,k}))$$
$$= \sum_{f=1}^{F}\sum_{k=1}^{K}(\alpha_w - 1)\log(w_{f,k}) + \alpha_w \log(\beta_w) - \beta_w w_{f,k}$$
$$=^{+} \sum_{f=1}^{F}\sum_{k=1}^{K}(\alpha_w - 1)\log(w_{f,k}) - \beta_w w_{f,k}$$

Finally :

$$\mathcal{L}_t(W,H) =^{+} \sum_{f=1}^{F}\sum_{n=1}^{N}\sum_{k=1}^{K}(\mathbf{E}(s_{f,n,k})_{P(s_{fnk}|v_{fn},w_{f,k}^t h_{k,n}^t)} \log(w_{f,k}h_{k,n}) - w_{f,k}h_{k,n})$$
$$+ \sum_{f=1}^{F}\sum_{k=1}^{K}(\alpha_w - 1)\log(w_{f,k}) - \beta_w w_{f,k} + \sum_{n=1}^{N}\sum_{k=1}^{K}(\alpha_h - 1)\log(h_{k,n}) - \beta_h h_{k,n}$$
$$=^{+} \sum_{f=1}^{F}\sum_{n=1}^{N}\sum_{k=1}^{K}(v_{f,n}\frac{w_{f,k}^t h_{k,n}^t}{\hat{v}_{f,n}^t} \log(w_{f,k}h_{k,n}) - w_{f,k}h_{k,n})$$
$$+ \sum_{f=1}^{F}\sum_{k=1}^{K}(\alpha_w - 1)\log(w_{f,k}) - \beta_w w_{f,k} + \sum_{n=1}^{N}\sum_{k=1}^{K}(\alpha_h - 1)\log(h_{k,n}) - \beta_h h_{k,n}$$

- **Third step** : Corresponding to the *M-step*:
  We have to find W,H such that :

  $$W^{t+1}, H^{t+1} = \underset{W,H}{\operatorname{argmax}} \mathcal{L}_t(W,H)$$

  for that, we should compute :

  $$W^{t+1} = \underset{W}{\operatorname{argmax}} \mathcal{L}_t(W,H^t)$$

  $$H^{t+1} = \underset{H}{\operatorname{argmax}} \mathcal{L}_t(W^t,H)$$

  to find W^{t+1} you have to derive the expression $\mathcal{L}_t(W,H^t)$ with respect to W:
  $\nabla_W \mathcal{L}_t(W,H^t)$=$(\frac{\partial}{\partial w_{fk}}\mathcal{L}_t(W,H^t))_{1\leq f\leq F, 1\leq k\leq K}$

  $$\frac{\partial}{\partial w_{fk}}\mathcal{L}_t(W,H^t)_{|w_{fk}=w_{fk}^{t+1}} = \sum_{n=1}^{N}((v_{f,n}\frac{w_{f,k}^t h_{k,n}^t}{\hat{v}_{f,n}^t}\frac{1}{w_{f,k}^{t+1}}) - h_{k,n}) + \frac{(\alpha_w - 1)}{w_{f,k}^{t+1}} - \beta_w = 0$$

  Then :

$$[(\alpha_w - 1) + \sum_n (v_{fn} \frac{w_{fk}^t h_{kn}^t}{\hat{v}_{fn}^t})] \frac{1}{w_{fk}^{t+1}} = \beta_w + \sum_k h_{kn}^t$$

$$w_{fk}^{t+1} = \frac{(\alpha_w - 1) + \sum_n (v_{fn} \frac{w_{fk}^t h_{kn}^t}{\hat{v}_{fn}^t})}{\beta_w + \sum_k h_{kn}^t}$$

following the same approach to find H^{t+1}, we obtain:

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

If we try to write the solution with matrices, we get:

$$W^{t+1} = \frac{(\alpha_w - 1)\mathbf{I}_w + W^t \odot ((V/\hat{V}^t)(H^t)^T)}{\beta_w \mathbf{I}_w + \mathbf{I}_v (H^t)^T}$$

$$H^{t+1} = \frac{(\alpha_h - 1)\mathbf{I}_h + H^t \odot ((W^t)^T (V/\hat{V}^t)))}{\beta_h \mathbf{I}_h + (W^t)^T \mathbf{I}_v}$$

Where $\odot$ correspond to the the elementwise product and / correspond of the elementwise division.
The matrices $\mathbf{I}_h$, $\mathbf{I}_w$ and $\mathbf{I}$ are all-ones matrices belonging respectively to $\mathbf{R}^{K \times N}$, $\mathbf{R}^{F \times K}$ and $\mathbf{R}^{F \times N}$.

2 Question 2 :

1. Implement the EM algorithm that you developed in Question 1.
   ***Answer* :**

   The implementation of the algorithm is done in the notebook Project.ypinb sent with this report.

2. Run the algorithm on the face dataset. Set K = 25, $\alpha_w = \alpha_h = 1$. Try different values for β_w and β_h . Visualize the columns of estimated W matrices. What do you observe when you change the parameters?
   ***Answer* :**

   - For the high values of β_w and β_h, we can distinguish some features of the face like the eyes, the jaw, the nose, ... This is clearly visible for the figure having β_w = 100 and β_h= 100 but by further increasing the beta values. We can see that the images become totally dark. so the increase of β_w and β_h values to make the matrix hollow.
   - For low values of beta we get some fuzzy images. This can be explained by the fact that the matrices W and H are dense.

3. Run the algorithm with K = 25, $\beta_w = \beta_h = 1$. Try different values for α_w and α_h . Visualize the columns of estimated W matrices. What do you observe when you change the parameters?
   Answer :

   - We can see that for low values of α_w and α_h (of the order of 1e-2) we observe a total opacity.
   - for alpha values greater than 1 we observe some facial feature. the increase in the value of α_h tends to make the image dark. This shows that the increase of α_h makes the matrix W sparse.
   - By increasing the value of α_w the image becomes more and more clear.

4. Now try changing the number of components K. What do you observe?
   Answer :

   NB: the variation of K could give us a meaningless result. sometimes we have a sake of interpretability.

   - The more we increases K the longer is the execution time of the algorithm. this is explained by the fact that the model becomes more and more complicated(harder to estimate, more parameters to transmit,...)
   - We can also notice that by increasing K, facial features become sharper even if the image is dark.So a greater K leads to a better approximation. Moreover with increasing the value of K we can observe more details in the picture that's means that we have more features.