Introduction to Probabilistic Graphical Models Maximum A Priori (MAP) Estimation on NMF

Yacine Ben Baccar
yacine.benbaccar@telecom-paristech.fr

November 22, 2018

1 Theoretical Work

1.1 Mathematical Tools

NMF decomposition : V = WH; st. $\begin{cases} V \in \mathbf{R}_+^{N*M} \\ H \in \mathbf{R}_+^{N*R} \\ W \in \mathbf{R}_+^{R*M} \end{cases}$

For our purposes The dataset of images will be the matrix : V, each column V_i is the flattened matrix of each face of the subjects. V has a shape of (92 * 112, 400).

Kullback-Leibler divergence: 'a measure of dissimilarity' :

$$d_{KL}(V||\hat{V}) = \sum_{i,j}(v_{ij}log(\frac{\hat{v_{ij}}}{v_{ij}}) - \hat{v_{ij}} + v_{ij})$$

Relation between the Poisson distribution and the Multinomial distribution $\mathcal{M}$:

$$X \sim \mathcal{PO}(X;\lambda) \Rightarrow p(X) = exp(log\lambda - \lambda - log\Gamma(X+1)) \quad (1)$$

$$S_i \sim \mathcal{PO}(\lambda_i) \iff X = \sum_{i=1}^{n} S_i \sim \mathcal{PO}(\sum_{i=1}^{n}\lambda_i) \quad (2)$$

$$p(S_1..S_n|X) = \mathcal{M}(S_1..S_n|\frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i}, .., \frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i}) \quad (3)$$

$$p(S_1..S_n|X) = \frac{X!}{S_1!..S_n!}(\frac{\lambda_1}{\sum_{i=1}^{n}\lambda_i})^{S_1}..(\frac{\lambda_n}{\sum_{i=1}^{n}\lambda_i})^{S_n} \quad (4)$$

1.2 Definition of the problem

MAP estimation on NMF:

$$f \in [1..F];\ n \in [1..N];\ k \in [1..K]$$

$$w_{fk} \sim \mathcal{G}(w_{fk}; \alpha_w, \beta_w)$$

$$h_{kn} \sim \mathcal{G}(h_{kn}; \alpha_h, \beta_h)$$

$$v_{fn}|wf, :, h_{:,n} \sim \mathcal{PO}(v_{fn}; \sum_{k=1}^{n} w_{fk}h_{kn})$$

We need to find W^* and H^* such that :

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

$$V \approx W^*H^*$$

1.3 The Graphical Models

Question 1:

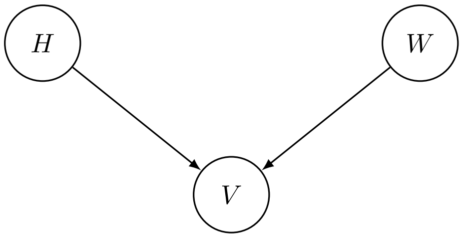

Figure 1: Directed graphical model

From this graphical model we can write the following probability distribution: $p(W, H, V) = p(V|W, H)p(W)p(H)$

1.4 Expectation Maximization (EM-Algorithm)

Question 2:
By implementing the Bayes rule and knowing that we have W $\perp$ H so we get $p(W, H) = p(W)p(H)$ we get the following:

$$\begin{aligned} p(W, H|V) &= \frac{p(V|W, H)p(W, H)}{p(V)} \\ &= \frac{p(V, W, H)}{p(V)} \end{aligned}$$

So we can rewrite the maximization problem:

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

Now let us consider the latent variable $\{s_{fnk}\}_{f,n,k}$

$$s_{fnk}|w_{fn}h_{kn} \sim \mathcal{PO}(s_{fnk}; w_{fk}h_{kn})$$

$$v_{fn}|s_{fn} \sim \delta(v_{fn} - \sum_{k=1}^{K} s_{fnk})$$

$$v_{fn} = \sum_{k=1}^{K} s_{fnk} \Rightarrow v_{fn} \sim \mathcal{PO}(v_{fn}, \sum_{k=1}^{K} w_{fk}h_{kn})$$

And Therefore, based on what we saw during the last lecture, $p(s_{fn1},..,s_{fnK}|v_{fn}) = \mathcal{M}(s_{fn1},..,s_{fnK}|\pi_{fn1},..,\pi_{fnK})$ with:

$$\pi_{fnk} = \frac{w_{fk}h_{kn}}{\sum_{k=1}^{K} w_{fk}h_{kn}}$$

So now we, get a new directed graph where the hidden variable S appears:

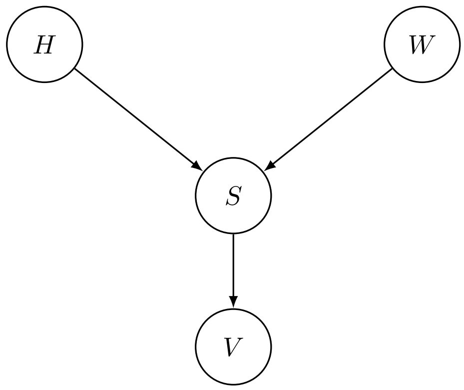

Figure 2: New Directed graphical model

E-Step: $\theta = (W, H)$

$$\mathcal{L}_t(\theta) = \mathrm{E}[\log p(V, S, \theta)]_{p(S|V,\theta^t)}$$
$$\iff \mathcal{L}_t(w_{f:}, w_{:n}) = \mathrm{E}[\log p(v_{fn,s_{fn},w_{f:},h_{:n}})]_{p(S_{fn}|V_{fn},\theta^t)}$$

Preparing the formulas for w_{fk} and h_{kn}

$$w_{fk} \sim \mathcal{G}(w_{fk}; \alpha_w, \beta_w)$$
$$h_{kn} \sim \mathcal{G}(h_{kn}; \alpha_h, \beta_h)$$
$$\log p(w_{fk}) = \alpha_w \log \beta_w + (\alpha_w - 1)\log w_{fk} - \beta_w w_{fk} - \log \Gamma(\alpha_w)$$
$$\log p(h_{kn}) = \alpha_h \log \beta_h + (\alpha_h - 1)\log h_{kn} - \beta_h h_{kn} - \log \Gamma(\alpha_h)$$

For our purposes, and when we inject both of these expressions inside the $\mathrm{argmax}_{(W,H)}$ operator will write them in the following fashion:

$$\log p(w_{fk}) \stackrel{+}{=} \sum_{f,k} (\alpha_w - 1)\log w_{fk} - \beta_w w_{fk}$$

$$\log p(h_{kn}) \stackrel{+}{=} \sum_{k,n} (\alpha_h - 1)\log h_{kn} - \beta_h h_{kn}$$

E-Step:

$$\theta^{t+1} = \underset{\theta}{\mathrm{argmax}}\, \mathcal{L}_t(\theta)$$

$$p(V,S,W,H) = p(S|W,H)p(V|S)p(W)p(H)$$
$$\Rightarrow \mathcal{L}_t(\theta) = \mathrm{E}[\log p(V|S) + \log p(S|W,H) + \log p(W) + \log p(H)]$$
$$\Rightarrow \operatorname*{argmax}_\theta \mathcal{L}_t(\theta) = \operatorname*{argmax}_\theta \mathrm{E}[\log p(S|H,W) + \log p(W) + \log p(V)]$$

Let's now denote $\theta_{fnk} = (w_{fn}, h_{kn})$ So the maximization problem will be element wise on θ_{fnk}

$$\operatorname*{argmax}_\theta \mathcal{L}_t(\theta) = \operatorname*{argmax}_\theta \mathrm{E}[\sum_{f,n,k} \log p(s_{fnk}|\theta_{fnk}) + \sum_{f,k} \log p(w_{fk}) + \sum_{k,n} \log p(h_{kn})]$$

$$\begin{aligned} = \operatorname*{argmax}_\theta \ \mathrm{E}[&\sum_{f,n,k} s_{fnk}\log(\theta_{fnk}) - \theta_{fnk} - \log\Gamma(s_{fnk}+1) \\ &+ \sum_{f,k}(\alpha_w - 1)\log(w_{fk}) - \beta_w w_{fk} \\ &+ \sum_{k,n}(\alpha_h - 1) log(h_{k,n}) - \beta_h h_{k,n}] \end{aligned}$$

$$\begin{aligned} = \operatorname*{argmax}_\theta \ &\sum_{f,n,k} \mathrm{E}(s_{fnk})\log(\theta_{fnk}) - \theta_{fnk} \\ &+ \sum_{f,k}(\alpha_w - 1)\log(w_{fk}) - \beta_w w_{fk} \\ &+ \sum_{k,n}(\alpha_h - 1)\log(h_{kn}) - \beta_h h_{kn} \end{aligned}$$

$$\begin{aligned} = \operatorname*{argmax}_\theta \ &\sum_{f,n,k} v_{fn}\pi^t_{fkn}\log(\theta_{fnk}) - \theta_{fnk} + \sum_{f,k}(\alpha_w - 1)\log(w_{fk}) - \beta_w w_{fk} \\ &+ \sum_{k,n}(\alpha_h - 1)\log(h_{kn}) - \beta_h h_{kn} \end{aligned}$$

With $\pi^t_{fkn} = \frac{w^t_{fk}h^t_{kn}}{\hat{v}^t_{fn}}$ such that : $\hat{v}^t_{fn} = \sum_{k=1}^K w_{fk}h_{kn}$

We need to compute the maximum for both w_{fk} and h_{kn}:

$$\frac{\partial \mathcal{L}_t}{\partial w_{fk}}_{|w_{fk}=w^{t+1}_{fk}} = \frac{\alpha_w - 1}{w^{t+1}_{fk}} - \beta_w - \sum_{n=1}^N h_{kn} + \sum_{n=1}^N v_{fn}\frac{\pi^t_{fkn}}{w^{t+1}_{fk}} = 0$$

$$\Rightarrow w^{t+1}_{fk} = \frac{(\alpha_w - 1) + \sum_{n=1}^N v_{fn}\pi^t_{fkn}}{\beta_w + \sum_{n=1}^N h^t_{kn}}$$

$$\frac{\partial \mathcal{L}_t}{\partial h_{kn}}_{|h_{kn}=h^{t+1}_{kn}} = \frac{\alpha_h - 1}{h^{t+1}_{kn}} - \beta_h - \sum_{f=1}^F h_{kn} + \sum_{f=1}^F v_{fn}\frac{\pi^t_{fkn}}{h^{t+1}_{kn}} = 0$$

$$\Rightarrow h^{t+1}_{kn} = \frac{(\alpha_h - 1) + \sum_{f=1}^F v_{fn}\pi^t_{fkn}}{\beta_h + \sum_{f=1}^F h^t_{kn}}$$

2 Implementation

When we try to implement the EM-algorithm element wise we will need an important computation time to get the results, that's why for the implementation we opted for the matricial expression of the two previous results and so we get.

$$W^{t+1} = \frac{(\alpha_w - 1)\mathbf{1}_{F,K} + W^t \odot ((V \oslash \hat{V})(H^t)^T)}{\beta_w \mathbf{1}_{F,K} + \mathbf{1}_{F,N}(H^t)^T} \tag{5}$$

$$H^{t+1} = \frac{(\alpha_h - 1)\mathbf{1}_{K,N} + H^t \odot ((W^t)^T (V \oslash \hat{V}))}{\beta_h \mathbf{1}_{K,N} + (W^t)^T \mathbf{1}_{F,N}} \tag{6}$$

With $\mathbf{1}_{M,N} \in \mathrm{R}^{M*N}$ and $(\mathbf{1}_{M,N})_{m,n} = 1$.

References

[1] Ferdinando S Samaria and Andy C Harter, *"Parameterisation of a stochastic model for human face identification" Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994.* IEEE, 1994, pp. 138–142.

[2] Ali Taylan Cemgil, *"Bayesian inference for nonnegative matrix factorisation models," Computational Intelligence and Neuroscience,* vol. 2009, 2009.