PGM DM 01 MVA 2019/2020

RAMI Hamza, MOKADIRI Mohammed Amine

ENS Paris Saclay, Cachan, Paris

November 22, 2019

1 Learning in discrete graphical models

Consider the following model: z and x are discrete variables taking respectively M and K different values with $p(z=m)=\pi_m$ and $p(x=k|z=m)=\theta_m k$

Let's take an i.i.d. sample of observations $(Z_1, Z_2, ..., Z_n,$ we have $:P(Z_i = m) = \pi_m,$ with m $\in [m_i, 1 \leq i \leq M]$.

The likelihood for $(z_1, ..., z_n)$ can be written as follow :

$$l_{(z_1,z_2,...,z_n)}(\pi) = \prod_{k=1}^{n} p(z_k|\pi) = \prod_{i=1}^{M} \pi_{m_i}^{\sum_{k=1}^{n} \mathbb{1}_{z_k=m_i}} \tag{1}$$

And the Log-likelihood is :

$$L_{(z_1,z_2,...,z_n)}(\pi) = \sum_{i=1}^{M} (\sum_{k=1}^{n} \mathbb{1}_{z_k=m_i}) log(\pi_{m_i})$$

maximizing the loglikelihood can be solved by solving the problem $\begin{cases} \max L(\pi) \\ s.t : \\ \sum \pi_{m_i} = 1 \end{cases}$

we consider the Lagrange multiplier : $h_\lambda(\pi) = L(\pi) + \lambda(1 - \sum_{i=1}^{M} \pi_{m_i})$

$$\frac{\partial h_\lambda}{\partial \pi_{m_i}}(\pi) = \frac{1}{\pi_{m_i}} \sum_{k=1}^{n} \mathbb{1}_{z_k=m_i} - \lambda = 0$$

$\pi_{m_i}\lambda = \sum_{k=1}^{n} \mathbb{1}_{z_k=m_i} \implies \sum_{i=1}^{M} \pi_{m_i} = \sum_{i=1}^{M} \sum_{k=1}^{n} \mathbb{1}_{z_k=m_i} = \sum_{k=1}^{n} \sum_{i=1}^{M} \mathbb{1}_{z_k=m_i}$
since $\sum_{i=1}^{M} \pi_{m_i} = 1$ and $\sum_{i=1}^{M} \mathbb{1}_{z_k=m_i} = 1$, So finally $\lambda = n$
and

$$\hat{\pi}_{m_i} = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{z_k=m_i}$$

Let's take an i.i.d. sample of observations $(X_1, X_2, ..., X_n,$ we have $:P(X_i = k|z = m) = \theta_{mk},$ with m $\in [m_i, 1 \leq i \leq M]$ and k $\in [k_i, 1 \leq i \leq K]$.

The likelihood for $(x_1, ..., x_n)$ can be written as follow :

$$l_{(x_1,x_2,...,x_n)}(\theta) = \prod_{i=1}^{n} p(x_i|z,\theta,\pi) = \prod_{k,m} \theta_{mk}^{\sum_{i=1}^{n} \mathbb{1}_{x_i=k,z=m}} \tag{2}$$

The loglikelihood can be written as :

$$L_{(x_1,x_2,...,x_n)}(\theta) = \sum_{k,m} (\sum_{i=1}^{n} \mathbb{1}_{x_i=k,z=m}) log(\theta_{mk}) \tag{3}$$

maximizing the loglikelihood can be solved by solving the problem $\begin{cases} \max L(\theta) \\ s.t : \\ \sum \theta_{mk}\pi_m = 1 \end{cases}$

Setting the derivatives to zero as the previous question, we find :

$$\hat{\theta}_{mk} = \frac{\sum_{i=1}^{n} \mathbb{1}_{x_i=k,z=m}}{\hat{\pi}_m \sum_{k,m}(\sum_{i=1}^{n} \mathbb{1}_{x_i=k,z=m})} \tag{4}$$

2 Linear classification

2.1 Generative model (LDA)

a- Let us consider an i.i.d sample of observations $\{(x_i, y_i)\}_{1\leq i\leq n}$ with $\mathrm{x}_i \in R^2, y_i \in \{0,1\}$

Suppose that $y \sim Bernoulli(\pi)$ & $x|y = i \sim \mathcal{N}(\mu_i, \Sigma)$
The likelihood can be written as follow:

$$\begin{aligned}
l(\pi, \mu_0, \mu_1, \Sigma) &= \prod_{i=1}^{n} p(x_i, y_i) \\
&= \prod_{i=1}^{n} p(x_i|y_i)p(y_i) \\
&= \prod_{i=1}^{n} \frac{\pi^{y_i}(1-\pi)^{1-y_i}}{2\pi \det(\Sigma)^{1/2}} \exp(-\frac{1}{2}(x_i - \mu_{y_i})^T\Sigma^{-1}(x_i - \mu_{y_i}))
\end{aligned}$$

The Log-likelihood can be written as follow:

$$\begin{aligned}
L(\pi, \mu_0, \mu_1, \Sigma) &= \sum_{i=1}^{n} \log(p(x_i, y_i)) \\
&= \sum_{i=1}^{n} -\frac{1}{2}(x_i - \mu_{y_i})^T\Sigma^{-1}(x_i - \mu_{y_i}) \sum_{i=1}^{n} y_i log(\pi) + \sum_{i=1}^{n}(1-y_i)log(1-\pi) - \frac{n}{2}log(\det(\Sigma)) + C^{te}
\end{aligned}$$

Setting the derivatives to zeros we get :

$$\hat{\pi} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{5}$$

$$\hat{\mu_k} = \frac{\sum_{i=1}^{n} x_i \mathbb{1}_{yi} = k}{\sum_{i=1}^{n} \mathbb{1}_{y_i=k}} \tag{6}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_k i)(x_i - \mu_k i)^T \text{ with } ki = \mathbb{1}_{yi=1} \tag{7}$$

b- the form of $p(y = 1|x)$

$$\begin{aligned}
p(y = 1|x) &= \frac{p(x|y = 1)p(y = 1)}{p(x|y = 1)p(y = 1) + p(x|y = 0)p(y = 0)} \\
&= \frac{1}{1 + \frac{(1-\pi)f_0(x)}{\pi f_1(x)}} \\
&= \sigma(w^T x + b)
\end{aligned}$$

with :

$$w = \Sigma^{-1}(\mu_1 - \mu_0), b = \log(\frac{\pi}{1-\pi}) - \frac{1}{2}(\mu_1 + \mu_0)^T\Sigma^{-1}(\mu_1 - \mu_0)$$