119010143
Li Chenyi

1. Proof full conditional for node i in DGM

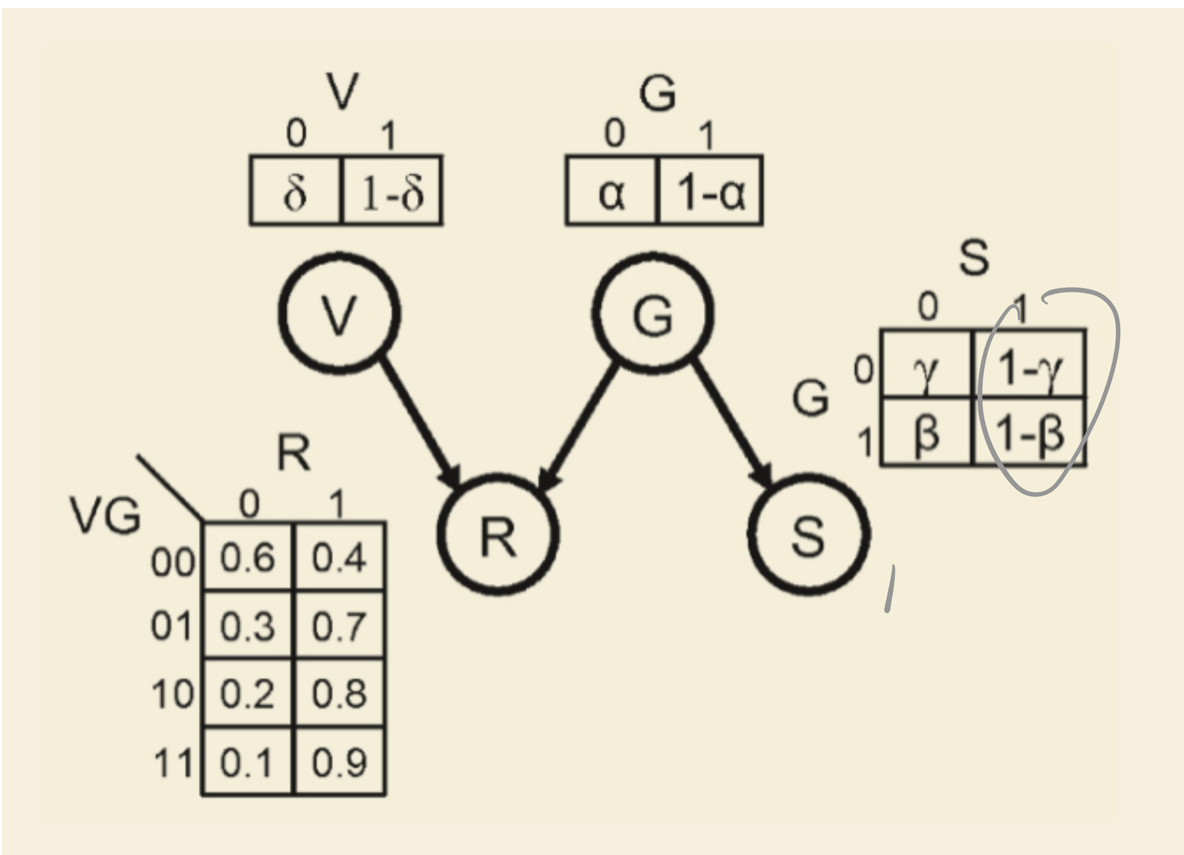
x_i

node x_i

$$P(x_i | x_{-i}) = \frac{P(x)}{P(x_{-i})} = \frac{P(x)}{\int P(x)\,dx_i} = \frac{\prod_{j=1}^{p} P(x_j | x_{parents(x_j)})}{\int \prod_{j=1}^{p} P(x_j | x_{parents(j)})\,dx_i}$$

$$= P(x_i | x_{parents(i)})\, P(x_{child(i)} | x_i)$$

2.



(a) $P(S=1 \mid V=1) = \dfrac{P(S=1, V=1)}{P(V=1)} = \dfrac{(1-\delta)(2-\gamma-\beta)}{1-\delta} = 2-\gamma-\beta$

$P(V=1) = 1-\delta$

$$P(S=1, V=1) = \sum_{G} P(S=1, V=1, G)$$
$$= \sum_{G} P(V=1) \cdot P(S=1, G \mid V=1)$$
$$= (1-\delta) \sum_{G} \cdot P(S=1, G)$$
$$= (1-\delta)(1-\gamma+1-\beta)$$
$$= (1-\delta)(2-\gamma-\beta)$$

(b) $P(S=1 \mid V=0) = \dfrac{P(S=1, V=0)}{P(V=0)} = \dfrac{\delta(2-\gamma-\beta)}{\delta} = 2-\beta-\gamma$

$$P(S=1, V=0) = \sum_{G} P(S=1, V=0, G)$$
$$= \sum_{G} P(V=0) \cdot P(S=1, G \mid V=0)$$
$$= \delta \cdot (2-\gamma-\beta)$$

$P(S=1 \mid V=0) = P(S=1 \mid V=1)$ same

reason: $P(S=1 \mid V=0) = \dfrac{P(S=1, V=0)}{P(V=0)} = \dfrac{\sum_{G} P(V=0)\, P(G, S=1 \mid V=0)}{P(V=0)}$

$$= \sum_G P(G, S=1 | V=0) = \sum_G P(G, S=1)$$

$$P(S=1 | V=1) = \sum_G \frac{P(G, S=1 | V=1)\ P(V=1)}{P(V=1)} = P(G, S=1)$$

(c) $\alpha, \beta, \gamma = \underset{\alpha, \beta, \gamma}{\operatorname{argmax}}\ P(D | \alpha, \beta, \gamma)$

3. K-means cost function

Algorithm:

initialize μ.

repeat until convergence:

$z_i = \arg\min_k \|x_i - \mu_k\|_2^2$ for $i=1$ to n

$\mu_k = \frac{1}{N_k} \sum_{i: z_i=k} x_i$;

Proof: $J_w(z) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i: z_i=k} \sum_{i': z_{i'}=k} (x_i - x_{i'})^2 = \sum_{k=1}^{K} n_k \sum_{i: z_i=k} (x_i - \bar{x}_k)^2$ (1)

We claim $\sum_i (x_i - \mu)^2 = ns^2 + n(\bar{x} - \mu)^2$ $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (*)

① left, for any k

$$\sum_{i: z_i=k} \sum_{i': z_{i'}=k} (x_i - x_{i'})^2 = \sum_{i=z_i=k} \sum_{i': z_{i'}=k} (x_{i'} - x_i)^2$$

$$\overset{by(*)}{=} \sum_{i=z_i=k} \left(n_k \cdot S_k^2 + n_k(\bar{x}_k - x_i)^2\right)$$

$$= n_k^2 S_k^2 + n_k \cdot \sum_{i: z_i=k} (\bar{x}_k - x_i)^2$$

$$= n_k^2 S_k^2 + n_k \cdot \sum_{i: z_i=k} (x_i - \bar{x}_k)^2$$

$$\overset{by(*)}{=} n_k^2 S_k^2 + n_k \cdot (n_k S_k^2 + n_k(\bar{x}_k - \bar{x}_k))$$

$$= 2n_k^2 S_k^2 \quad (**)$$

② right, for any k

$$n_k \sum_{i: z_i = k} (x_i - \bar{x}_k)^2 = n_k \left(n_k \cdot S_k^2 + n_k (\bar{x}_k - \bar{x}_k) \right)$$
$$= n_k^2 S_k^2$$

$\frac{1}{2} \times (**) = n_k^2 S_k^2 = (***)$

$\therefore$ (1) is proved ■

4. proof: (a)

suppose we know soft assignment for Data $\{x_1, \dots x_n\}$

$\{\gamma_1, \gamma_2, \dots \gamma_n\}$

① for one class. k

for $X = (x_1, x_2, \dots, x_L) \in \{0,1\}^L$.

$x_i \sim Ber(p_i)$

$P_k = (\mu_1, \cdots \mu_L)$

$X \sim Ber(P_k)$ $\quad P(X, P_k) = \prod_{i=1}^{L} \mu_i^{x_i} (1-\mu_i)^{1-x_i}$

② class $k \sim$ Categorical (W_k)

$\Rightarrow$ ∴ Bernoulli mixture model

$$P_{BMM}(X; K, P_1, P_2, \dots, W) = \sum_{k=1}^{K} W_k \cdot P_{BM}(X; P_k)$$

$$= \sum_{k=1}^{K} W_k \cdot \prod_{i=1}^{L} \mu_i^{x_i} (1-\mu_i)^{1-x_i}$$

$$P(X; K, P, W) = \prod_{i=1}^{n} P(X_i; K, P, W)$$

$$= \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} w_k \cdot P(x_i \mid \mu_k) \right\}$$

$$= \prod_{i=1}^{n} \sum_{k=1}^{K} W_k \cdot \prod_{i=1}^{L} \mu_i^{x_i} (1-\mu_i)^{1-x_i}$$

log-likelihood $J = \sum_{i=1}^{n} \log \sum_{k=1}^{K} P(x_i \mid \mu_k)$.

suppose x from z_n

$$P(X, Z \mid \mu, \pi) = \prod_{i=1}^{n} \pi_{z_n} P(x_n \mid \mu_{z_n})$$

we want to solve

$$\max_{\mu,\pi} \sum_{i=1}^{n}\sum_{j=1}^{k} \gamma_{ij} \log P(x_i|\mu_k) \quad = g(\mu,\pi)$$

where $\gamma_{ij} = \dfrac{\pi_j P(x_i|\mu_j)}{\sum_{j=1}^{k} \pi_j P(x_i|\mu_j)}$

$$g(\mu,\pi) = \sum_{i=1}^{n}\sum_{j=1}^{k} \gamma_{ij} \log(\pi_j \, \mu_j^{x_i} (1-\mu_j)^{1-x_i}) \; (*)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{k} \gamma_{ij}[\ln \pi_j + x_i \log \mu_j + (1-x_i)\log(1-\mu_j)]$$

where $\sum_{k=1}^{K} \pi_k = 1$

$$\frac{\partial g(\mu,\pi)}{\partial \mu_k} = \sum_{i=1}^{n} \gamma_{ik}\left(\frac{x_i}{\mu_k} - \frac{1-x_i}{1-\mu_k}\right) = 0$$

$$= \sum_{i=1}^{n} \gamma_{ik} \frac{x_i - x_i\mu_k - (\mu_k - \mu_k x_i)}{\mu_k(1-\mu_k)}$$

$$= \sum_{i=1}^{n} \frac{\gamma_{ik}(x_i - \mu_k)}{\mu_k \cdot (1-\mu_k)} = 0$$

$$\therefore \mu_k = \frac{1}{N_k} \sum_{i=1}^{n} \gamma_{ik} x_i$$

$$N_k = \sum_{i=1}^{n} \gamma_{ik}$$

(b)

$$\mu_k \sim \beta(\alpha, \beta)$$

$$P(\mu_k) = \mu_k^{\alpha-1} (1-\mu_k)^{\beta-1}$$

$$(*) = \sum_{i=1}^{n} \sum_{j=1}^{K} \gamma_{ij} \log \left(\pi_j \cdot (\mu_k^{\alpha-1})^{x_i} ((1-\mu_k)^{(\beta-1)})^{1-x_i} \right)$$

let $\mu_k = \mu_{k0}$

$1 - \mu_k = \mu_{k1}$

$$(*) = \sum_{i=1}^{n} \sum_{j=1}^{K} \gamma_{ij} \Big(\log \pi_j + (\alpha-1)(x_i) \log \mu_{k0}$$

$$+ (\beta-1)(1-x_i) \log \mu_{k1} \Big)$$

$$\frac{\partial (*)}{\partial \mu_{k0}} = \sum_{i=1}^{n} \gamma_{ij} \frac{(\alpha-1) x_i}{\mu_{k0}} = 0 \quad \mu_{k0} = \frac{\sum_{i=1}^{n} \gamma_{ik} x_i + \alpha - 1}{\sum_{i=1}^{n} \gamma_{ik} + \alpha + \beta - 2}$$

$$\partial (*) = \sum \gamma_{ij} \frac{(\beta-1)(1-x_i)}{} = 0 \quad \mu_{k} = \frac{\sum_{i=1}^{n} \gamma_{ik} (1-x_i) + \beta - 1}{}$$

$$\frac{\quad}{\partial \mu_{k1}} \quad \sum_{i=1} \quad \mu_{k1} \quad \sum_{i=1} \gamma_{ik}+\alpha+\beta-2$$