GOLOR MLE

* gradient derivation for vertex coloring using the gradient formula for log-linear model.

$$P = \frac{1}{Z} \prod_{ij \in E} \mathbb{1}_{x_i \neq x_j} \prod_{i \in V} \exp(\omega_{x_i})$$

since $\mathbb{1}_{x_i \neq x_j}$ is always 1 we use simplified prob. formula

$$P = \frac{1}{Z} \prod_{i \in V} \exp(\omega_{x_i})$$

for log linear model. of form

$$P = \frac{1}{Z(\theta)} \prod_{c} \exp(\langle \theta, f_c(x_c) \rangle)$$

we define $\theta = \omega$ vector. $\omega = \begin{bmatrix} \omega_{x_1} \\ \omega_{x_2} \\ \vdots \\ \omega_{x_n} \end{bmatrix}$ for n colors.

c is all set of vertices

$$f_c(x_c) = f_i(x_i) = \begin{bmatrix} \mathbb{1}_{x_i = 1} \\ \mathbb{1}_{x_i = 2} \\ \vdots \\ \mathbb{1}_{x_i = n} \end{bmatrix}$$ for n colors & vertex i.

Now, general gradient is given by

$$\nabla_{\theta} \log Z(\theta) = \sum_{c} \sum_{m} \left(f_c(x_c^m) - \sum_{x_c} p_c(x_c | \theta) f_c(x_c) \right)$$

Similarly for our case,

$$\nabla_{\omega} \log \ell(\omega) = \sum_{i \in V} \sum_{m} \left[f_i(x_i) - \sum_{x_i} p_i(x_i | \omega) f_i(x_i) \right]$$

$$= \sum_{i \in V} \underbrace{\begin{bmatrix} \# x_i = 1 \\ \# x_i = 2 \\ \# x_i = 3 \\ \vdots \\ \# x_i = n \end{bmatrix}}_{\text{in } m \text{ samples}} - \sum_{i \in V} m . \sum_{x_i} P(x_i / w) \, f_i(x_i)$$

$$= \sum_{i \in V} \begin{bmatrix} \# x_i = 1 \\ \# x_i = 2 \\ \vdots \\ \# x_i = n \end{bmatrix} - \sum_{i \in V} m . \begin{bmatrix} P(x_i = 1 / w) \\ P(x_i = 2 / w) \\ P(x_i = 3 / w) \\ \vdots \\ P(x_i = n / w) \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \# 1 \\ \# 2 \\ \vdots \\ \# n \end{bmatrix}}_{\substack{\text{for all vertices,} \\ \text{\# of color 1 in} \\ \text{all samples}}} - m . \underbrace{\begin{bmatrix} P(x_1 = 1/w) + P(x_2 = 1/w) \dots P(x_v = 1/w) \\ \vdots \\ P(x_1 = n/w) + P(x_2 = n/w) + \dots P(x_v = n/w) \end{bmatrix}}_{\substack{\text{Sum of all the marginal} \\ \text{prob of all the vertices for} \\ \text{given domain / vertex color.}}}$$

---

once we have the gradient, we start with random w & keep adding gradient with some step size/learningrate)

$$w_{new} = w_{old} + \text{learningrate} \times \text{gradient}$$

gradient = observation − m. sum of marginals.

★ Gradient derivation for vertex coloring

FOR vertex coloring problem, our prob is given by

$$p(x) = \frac{1}{Z} \prod_{(i,j)\in E} 1_{x_i \neq x_j} \prod_{i\in V} \exp(\omega_{x_i})$$

$$Z = \sum_{x} \prod_{(i,j)\in E} 1_{x_i\neq x_j} \prod_{i\in V} \exp(\omega_{x_i})$$

$$\log(z) = \log\left[\sum_{x} \prod_{i,j\in E} 1_{x_i\neq x_j} \prod_{i\in V} \exp(\omega_{x_i})\right]$$

Since we limit all the possible assignments $[X]$ to valid assignments where $1_{x_i\neq x_j} = 1$ always, we can rewrite

$$\log(z) = \log\left[\sum_{x} \prod_{i\in V} \exp(\omega_{x_i})\right]$$

$$= \log \sum_{x} \exp\left(\sum_{i\in V} \omega_{x_i}\right)$$

Now, log likelihood of m samples is given by

$$\log l = \sum_{m} \log p(x)$$

$$= \sum_{m}\left[\log\left[\prod_{i,j\in E} 1_{x_i\neq x_j}\right] + \log\left[\prod_{i\in V} \exp(\omega_{x_i})\right] - \log(z)\right]$$

$$= \sum_{m}\sum_{ij} \log 1_{x_i\neq x_j} + \sum_{m}\sum_{i\in V} \omega_{x_i} - \sum_{m} \log(z)$$

Since all our samples are valid assignments, term 1 becomes constant.

$$= C + \sum_{m}\sum_{i\in V} \omega_{x_i} - \sum_{m} \log(z)$$

To find gradient, we take derivative w.r.t ω_1 & then extend it to all ωs

$$\frac{\partial \log L}{\partial \omega_1} = 0 + \sum_m \sum_{\substack{i \in V \\ st:\, x_i = 1}} 1 - \sum_m \frac{\partial}{\partial \omega_1} \log(z)$$

$$= 0 + \#1 \text{ in samples} - \sum_m \frac{1}{z} \sum_x \frac{\partial}{\partial \omega_1} \exp\left(\sum_{i \in V} \omega x_i\right)$$

$$= \#1 \text{ in samples} - \sum_m \frac{1}{z} \sum_x \left[\exp\left(\sum_{i \in V} \omega x_i\right) \cdot \sum_{i \in V} 1_{x_i = 1}\right]$$

$$= \#1 \text{ in samples} - \sum_m \sum_x \left[\frac{1}{z} \exp\left(\sum_{i \in V} \omega_{x_i}\right) \cdot \sum_{i \in V} 1_{x_i = 1}\right]$$

$$= \#1 \text{ in samples} - m \cdot \sum_{\substack{i \in V \\ st\ x_i = 1}} \sum_{x_i = 1} \frac{1}{z} \exp\left(\sum_{i \in V} \omega_{x_i}\right)$$

$$= \#1 \text{ in samples} - m \cdot \sum_{\substack{i \in V \\ st.\ x_i = 1}} \frac{\exp\left(\sum_{i \in V} \omega_{x_i}\right)}{z}$$

$$= \#1 \text{ in samples} - m \cdot \underbrace{\sum_{\substack{i \in V \\ st\ x_i = 1}} P(x_i = 1)}_{\text{sum of marginal of all vertex having color 1}}$$

$$= \#1 \text{ in samples} - m \cdot \sum_{i \in V} P(x_i = 1)$$

Similarly

$$\frac{\partial \log L}{\partial \omega_2} = \#2 \text{ in samples} - m \cdot \sum_{i \in V} P(x_i = 2)$$

So we got gradient in all directions.

E^{step} $\quad P(x_m^k \mid X_{obs}^k, \theta) = q$ COLOR EM

M^{step}.

$$\theta^{t+1} = \arg\max_{\theta} \sum_{k=1}^{K} \sum_{x_m} q(x_m) \log P(x_o^k\, x_m^k \mid \theta)$$

$$\frac{\partial \omega^{t+1}}{\partial \omega_1^t} = \sum_{k=1}^{K} \sum_{x_m} q(x_m) \frac{\partial}{\partial \omega_1^t} \left(\log\left[\prod_{i,j \in E} 1_{x_i \neq x_j}\right] + \log\left[\prod_{i \in V_m} \exp(\omega_{x_i})\right] + \log\left[\prod_{i \in V_o} \exp(\omega_{x_i})\right] - \log(Z) \right)$$

$$= \sum_{k=1}^{K} \sum_{x_m} q(x_m) \left[0 + \frac{\partial}{\partial \omega_1^t} \sum_{i \in V_m} \omega_{x_i} + \frac{\partial}{\partial \omega_1^t} \sum_{i \in V_o} \omega_{x_i} - \frac{1}{Z} \sum_{x} \frac{\partial}{\partial \omega_1} \exp\left(\sum_{i \in V} \omega_{x_i}\right) \right]$$

$$= \sum_{k} \sum_{x_m} q_k(x_m) \left[\#1 \text{ in missing values} + \#1 \text{ in sample } k - \underbrace{\frac{1}{Z} \sum_{x} \exp\left(\sum_{i \in V} \omega_{x_i}\right) \cdot \sum_{i \in V} 1_{x_i = 1}} \right]$$

$$= \sum_{k} \sum_{x_m} q_k(x_m) \#1 \text{ in } x_m + \sum_{k} \sum_{x_m} \#1 \text{ in } k (q x_m) - \sum_{k} \sum_{x_m} q(x_m) \frac{\sum_{x_i=1} \exp\left(\sum_{i \in V} \omega_{x_i}\right)}{Z}$$

$$= \sum_{k} \sum_{x_m} q_k(x_m) \#1 \text{ in } x_m + \sum_{k} \#1 \text{ in } k \underbrace{\sum_{x_m} q_k(x_m)}_{=1} - \sum_{i \in V} P(x_i = 1 \mid w) \sum_{k} \underbrace{\sum_{x_m} q_k(x_m)}_{=1}$$

$$= \sum_{k} \sum_{x_m} q(x_m) \#1 \text{ in } x_m + \sum_{k} \#1 \text{ in } k - \sum_{i \in V} P(x_i = 1 \mid w)\, K$$

Similarly for all colors.

Δ_{ω_i} = Expected number of missing variables taking color i + # of observed variables in sample having color i − K · sum of marginals of all vertices having color i