MLE using GOP

November 11, 2017

1 Model

We will try to maximize likelihood for simple model with only one hidden variable and n observed variables.

$$p(x) = \theta_h(x_h)\prod_i \theta_{ih}(x_i, x_h)$$

Likelihood of M samples is,

$$L(\theta) = \prod_{m=1}^{M}\sum_{x_h} p(x_1^m, ..., x_n^m, x_h)$$

Writing it in variational form,

$$\begin{aligned}
\log L(\theta) &= \max_b \sum_m \langle b_h^m, \log\theta'_h\rangle + H(b_h^m) \\
&\quad \text{where, } \theta'_h(x_h) = \theta_h(x_h) * \prod_i \theta_{ih}(x_i^m, x_h) \\
\log L(\theta) &= \max_b \sum_m \sum_{x_h} b_h^m(x_h)\log(\theta_h(x_h)) + \sum_m \sum_{x_h} \sum_i b_h^m(x_h)\log(\theta_{ih}(x_i^m, x_h)) + \sum_m H(b_h^m)
\end{aligned}$$

Define $\bar{b}_h(x_h)$ and $\bar{b}_{ih}(x_i, x_h)$ and rewrite the formulation,

$$\begin{aligned}
\log L(\theta) &= \max_b \sum_{x_h} \bar{b}_h(x_h)\log(\theta_h(x_h)) + \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h)\log(\theta_{ih}(x_i, x_h)) + \sum_m H(b_h^m) \\
&\quad \text{where, } \bar{b}_h(x_h) = \frac{1}{M}\sum_m b_h^m(x_h) \\
&\quad \bar{b}_{ih}(x_i, x_h) = \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h)
\end{aligned}$$

2 Problem

We want to solve following problem,

$$-\log L = \min_{\theta, b} -\sum_{x_h} \bar{b}_h(x_h)\log(\theta_h(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h)\log(\theta_{ih}(x_i, x_h)) - \sum_m H(b_h^m)$$

with respect to following constraints,

$$\begin{aligned}
&\bar{b}_h(x_h) = \sum_m b_h^m(x_h), \forall x_h \\
&\bar{b}_{ih}(x_i, x_h) = \sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h), \forall x_i, x_h \\
&\sum_{x_h} \theta_h(x_h) = 1 \\
&\sum_{x_i} \theta_{ih}(x_i, x_h) = 1 \\
&\sum_{x_h} b_h^m(x_h) = 1, \forall m \in M
\end{aligned}$$

3 GOP formulation

3.1 Primal problem

We need to solve following primal problem for fixed θ_k at k'th iteration,

$$\begin{aligned}
P^k(\theta^k) = \min_b \Bigg\{ &- \sum_{x_h} \bar{b}_h(x_h) \log(\theta_h^k(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h) \log(\theta_{ih}^k(x_i, x_h)) - \sum_m H(b_h^m) \\
&+ \sum_{x_h} \lambda_h(x_h) \left(\bar{b}_h(x_h) - \sum_m b_h^m(x_h) \right) + \sum_i \sum_{x_i, x_h} \lambda_{ih}(x_i, x_h) \left(\bar{b}_{ih}(x_i, x_h) - \sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h) \right) \\
&+ \sum_{m=1}^{M} \lambda_m \left(\sum_{x_h} b_h^m(x_h) - 1 \right) + \lambda_{\theta_h} \left(\sum_{x_h} \theta_h^k(x_h) - 1 \right) + \sum_{i, x_h} \lambda_{\theta_{ih}}(x_h) \left(\sum_{x_i} \theta_{ih}^k(x_i, x_h) - 1 \right) \Bigg\}
\end{aligned}$$

Taking derivatives to get λ values,

$$\begin{aligned}
\frac{\partial P^k(\theta^k)}{\partial \bar{b}_h(x_h)} &= -\log(\theta_h^k(x_h)) + \lambda_h^k(x_h) = 0 \Rightarrow \lambda_h^k(x_h) = \log(\theta_h^k(x_h)) \\
\frac{\partial P^k(\theta^k)}{\partial \bar{b}_{ih}(x_i, x_h)} &= -\log(\theta_{ih}^k(x_i, x_h)) + \lambda_{ih}^k(x_i, x_h) = 0 \Rightarrow \lambda_{ih}^k(x_i, x_h) = \log(\theta_{ih}^k(x_i, x_h)) \\
\frac{\partial P^k(\theta^k)}{\partial b_h^m(x_h)} &= 0 \Rightarrow \lambda^m = \log(\theta_h^k(x_h)) + \sum_i \log(\theta_{ih}^k(x_i^m, x_h)) - \log(b_h^m(x_h)) - 1 \ , \forall x_h \\
\frac{\partial P^k(\theta^k)}{\partial \theta_h^k(x_h)} &= 0 \Rightarrow \lambda_{\theta_h} = \frac{\bar{b}_h(x_h)}{\theta_h^k(x_h)} \ , \forall x_h \\
\frac{\partial P^k(\theta^k)}{\partial \theta_{ih}^k(x_i, x_h)} &= 0 \Rightarrow \lambda_{\theta_{ih}} = \frac{\bar{b}_{ih}(x_i, x_h)}{\theta_{ih}^k(x_i, x_h)} \ , \forall x_i, x_h
\end{aligned}$$

3.2 Relaxed dual problem

At k'th iteration, we solve following relaxed dual problem,

$$\begin{aligned}
&\min_{\theta, \mu_B} \mu_B \\
&\text{s.t. } \mu_B \geq L^k(b^{B_l}, \theta, \lambda^k)\big|_{b^k}^{lin}
\end{aligned}$$

Linearization of lagrange function around b^k is defined as,

$$L^k(b^{B_l}, \theta, \lambda^k)|_{b^k}^{lin} =$$