MLE using GOP

November 14, 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,

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

Taking log of likelihood and normalizing it, we get,

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

Writing it in variational form,

$$\begin{aligned}
\log L(\theta) &= \max_b \frac{1}{M} \sum_m \langle b_h^m, \log \theta_h' \rangle + H(b_h^m) \\
&\text{where, } \theta_h'(x_h) = \theta_h(x_h) * \prod_i \theta_{ih}(x_i^m, x_h) \\
\log L(\theta) &= \max_b \frac{1}{M} \sum_m \sum_{x_h} b_h^m(x_h) \log(\theta_h(x_h)) + \frac{1}{M} \sum_m \sum_{x_h} \sum_i b_h^m(x_h) \log(\theta_{ih}(x_i^m, x_h)) + \frac{1}{M} \sum_m H(b_h^m)
\end{aligned}$$

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

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

2 Problem

We want to solve following problem,

$$-\log L = \min_{\theta, b} - \sum_{x_h} \bar{b}_h(x_h) \log(\theta_h(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}(x_i, x_h) \log(\theta_{ih}(x_i, x_h)) - \frac{1}{M} \sum_m H(b_h^m)$$

with respect to following constraints,

$$\bar{b}_h(x_h) = \frac{1}{M}\sum_m b_h^m(x_h), \forall x_h$$

$$\bar{b}_{ih}(x_i, x_h) = \frac{1}{M}\sum_{\substack{m \\ x_i^m = x_i}} b_h^m(x_h), \forall x_i, x_h$$

$$\sum_{x_h} \theta_h(x_h) = 1$$

$$\sum_{x_i} \theta_{ih}(x_i, x_h) = 1$$

$$\sum_{x_h} b_h^m(x_h) = 1, \forall m \in M$$

3 GOP formulation

3.1 Primal problem

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

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

Taking derivatives to get λ values,

$$\frac{\partial P^k(\theta^k)}{\partial \bar{b}_h(x_h)} = -\log(\theta_h^k(x_h)) + \lambda_h^k(x_h) = 0 \Rightarrow \lambda_h^k(x_h) = \log(\theta_h^k(x_h))$$

$$\frac{\partial P^k(\theta^k)}{\partial \bar{b}_{ih}(x_i, x_h)} = -\log(\theta_{ih}^k(x_i, x_h)) + \lambda_{ih}^k(x_i, x_h) = 0 \Rightarrow \lambda_{ih}^k(x_i, x_h) = \log(\theta_{ih}^k(x_i, x_h))$$

$$\frac{\partial P^k(\theta^k)}{\partial b_h^m(x_h)} = 0 \Rightarrow 1 + \log b_h^m(x_h) - \frac{1}{M}\lambda_h(x_h) - \frac{1}{M}\sum_i \lambda_{ih}(x_i^m, x_h) + \lambda_m = 0 \ , \forall x_h$$

so $b_h^m(x_h)$ has closed form solution,

$$b_h^m(x_h) = \frac{\exp(\frac{1}{M}\left(\lambda_h(x_h) + \sum_i \lambda_{ih}(x_i^m, x_h)\right))}{\exp(\lambda_m + 1)}$$

Here $exp(\lambda_m + 1)$ is normalization factor for $b_h^m(x_h)$, such that,

$$\exp(\lambda_m + 1) = \sum_{x_h} \exp\left(\frac{1}{M}\left(\lambda_h(x_h) + \sum_i \lambda_{ih}(x_i^m, x_h)\right)\right)$$

$$\frac{\partial P^k(\theta^k)}{\partial \theta_h^k(x_h)} = 0 \Rightarrow \theta_h^k(x_h) = \frac{\bar{b}_h(x_h)}{\lambda_{\theta_h}} \ , \forall x_h$$

Summing over x_h on both sides,

$$\sum_{x_h} \theta_h^k(x_h) = \frac{\sum_{x_h} \bar{b}_h(x_h)}{\lambda_{\theta_h}}$$

Both sums equal 1, so

$\lambda_{\theta_h} = 1$

Similarly,

$$\frac{\partial P^k(\theta^k)}{\partial \theta_{ih}^k(x_i, x_h)} = 0 \Rightarrow \lambda_{\theta_{ih}} = \frac{\bar{b}_{ih}(x_i, x_h)}{\theta_{ih}^k(x_i, x_h)} \ , \forall x_i, x_h$$

3.2 Relaxed dual problem

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

$$\min_{\theta, \mu_B} \mu_B$$

Subject to,

$$\begin{aligned}
\mu_B \geq & L(b^{B_j}, \theta, \lambda^k)|_{b^k}^{lin} \\
& \Delta_{b_i} L(b, \theta, \lambda^k)|_{b^k} \leq 0 \text{ if } b_i^{B_j} = b_i^U \\
& \Delta_{b_i} L(b, \theta, \lambda^k)|_{b^k} \geq 0 \text{ if } b_i^{B_j} = b_i^L
\end{aligned}$$

where, $\forall j \in UL(k, K)$

$k = 1, 2, ..., K-1$

UL(k,K) is set of lagrange functions from k'th iteration whose qualifying constraints satisfy at the current fixed value θ^K for current primal problem.

$$\begin{aligned}
\mu_B \geq & L(b^{B_l}, \theta, \lambda^K)|_{b^K}^{lin} \\
& \Delta_{b_i} L(b, \theta, \lambda^K)|_{b^K} \leq 0 \text{ if } b_i^{B_l} = b_i^U \\
& \Delta_{b_i} L(b, \theta, \lambda^K)|_{b^K} \geq 0 \text{ if } b_i^{B_j} = b_i^L
\end{aligned}$$

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

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

After plugging in values of b^k, λ^k and simplifying,

$$\begin{aligned}
L^k(b, \theta, \lambda^k)|_{b^k}^{lin} = & -\sum_{x_h} \bar{b}_h^k(x_h) \log(\theta_h(x_h)) - \sum_i \sum_{x_i, x_h} \bar{b}_{ih}^k(x_i, x_h) \log(\theta_{ih}(x_i, x_h)) + \frac{1}{M} \sum_m \sum_{x_h} b_h^{m^k}(x_h) \log b_h^{m^k}(x_h) \\
& + \lambda_{\theta_h}^k \left(\sum_{x_h} \theta_h(x_h) - 1 \right) + \sum_{i, x_h} \lambda_{\theta_{ih}}^k(x_h) \left(\sum_{x_i} \theta_{ih}(x_i, x_h) - 1 \right) \\
& - \sum_{x_h} (\bar{b}_h(x_h) - \bar{b}_h^k(x_h)) \log \theta_h(x_h) - \sum_i \sum_{x_i, x_h} (\bar{b}_{ih}(x_i, x_h) - \bar{b}_{ih}^k(x_i, x_h)) \log(\theta_{ih}(x_i, x_h)) \\
& \frac{1}{M} \sum_m \sum_{x_h} (1 + \log b_h^m(x_h))(b_h^m(x_h) - b_h^{m^k}(x_h)) + \sum_{x_h} \lambda_h^k(x_h)(\bar{b}_h(x_h) - \bar{b}_h^k(x_h)) \\
& - \sum_{x_h} \frac{1}{M} \lambda_h^k(x_h)(b_h^m(x_h) - b_h^{m^k}(x_h)) + \sum_i \sum_{x_i, x_h} \lambda_{ih}^k(x_i, x_h)(\bar{b}_{ih}(x_i, x_h) - \bar{b}_{ih}^k(x_i, x_h))
\end{aligned}$$

$$- \sum_i \sum_{x_h} \frac{1}{M} \sum_m \lambda_{ih}^k (x_i^m, x_h)(b_h^m(x_h) - b_h^{m^k}(x_h)) + \sum_m \lambda_m^k \sum_{x_h} (b_h^k(x_h) - b_h^{m^k}(x_h))$$

Canceling out some terms,

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