Assignment 1

Ethan Holleman

October 24, 2021

1 Part 2

$$\ln P(X, C|\boldsymbol{\theta}) = \sum_i \sum_j C_{i,j} \ln \lambda_j + \sum_i \sum_j \sum_p \sum_k X_{i,j,p} [C_{i,j} \ln \psi^{(1)}_{p,k} + (1 - C_{i,j}) \ln \psi^{(0)}_{p,k}] \quad (1)$$

1.1 E step

$$P(\boldsymbol{C}|\boldsymbol{X}, \boldsymbol{\theta}) = \frac{P(\boldsymbol{X}, \boldsymbol{C}|\boldsymbol{\theta})}{P(\boldsymbol{X}|\boldsymbol{\theta})} \quad (2)$$

Product individual posteriors is proportional to posterior over all latent variables.

$$P(\boldsymbol{C}|\boldsymbol{X}, \boldsymbol{\theta}) \propto \prod_i \prod_j \prod_p P(C_{i,j}|X_{i,j,p}, \boldsymbol{\theta}) \quad (3)$$

Compute individual posteriors.

$$P(C_{i,j}|X_{i,j,p}) = \frac{P(X_{i,j,p}|C_{i,j}, \boldsymbol{\theta}) P(X_{i,j}|\boldsymbol{\theta})}{P(X_{i,j,p}|\boldsymbol{\theta})} \quad (4)$$

$P(E, F|G) = P(E|F, G)P(F|G)$

$$P(C_{i,j}|X_{i,j,p}) = \frac{P(X_{i,j,p}, C_{i,j}|\boldsymbol{\theta})}{P(X_{i,j,p}|\boldsymbol{\theta})} \quad (5)$$

$$P(C_{i,j}|X_{i,j,p}) = \frac{\sum_p \sum_k P(C_{i,j} = 1|\boldsymbol{\theta}) P(X_{i,j,p} = k|C_{i,j}, \boldsymbol{\theta})}{P(X_{i,j,p}|\boldsymbol{\theta})} \quad (6)$$

How to rewrite $P(X_{i,j,p}|\boldsymbol{\theta})$ and does the numerator make any sense at all?

1.2 M step

$$\boldsymbol{E}_q[\log P(\boldsymbol{X},\boldsymbol{C}|\boldsymbol{\theta})] = \sum_i \sum_j C_{i,j} \ln \lambda_j + \sum_i \sum_j \sum_p \sum_k X_{i,j,p}[C_{i,j} \ln \psi^{(1)}_{p,k} + (1-C_{i,j}) \ln \psi^{(0)}_{p,k}] \quad (7)$$

$$= \sum_i \sum_j \boldsymbol{E}_q[C_{i,j} \ln \lambda_j] + \sum_i \sum_j \sum_p \sum_k \boldsymbol{E}_q[X_{i,j,p}[C_{i,j} \ln \psi^{(1)}_{p,k} + (1-C_{i,j}) \ln \psi^{(0)}_{p,k}]] \quad (8)$$

$$= \sum_i \sum_j \boldsymbol{E}_q[C_{i,j}] \ln \lambda_j + \sum_i \sum_j \sum_p \sum_k \boldsymbol{E}_q X_{i,j,p}[C_{i,j} \ln \psi^{(1)}_{p,k} + (1-C_{i,j}) \ln \psi^{(0)}_{p,k}] \quad (9)$$

1.2.1 Derivitive with respect to λ_j

Constraint equation defines bounds of λ_j that we must optimize $\boldsymbol{E}_q[\log P(\boldsymbol{X},\boldsymbol{C}|\boldsymbol{\theta})]$ within. λ_j is probibility that a particular motif is the TFBS. There is only one TFBS per sequence so probibility over all possible motifs must sum to 1.

$$g(\lambda_j) = \sum_j \lambda_j - 1 \quad (10)$$

$$h(\lambda_j) = \mathrm{ELBO}(q_t(C),\boldsymbol{\theta}) + \phi g(\lambda_j) \quad (11)$$

$$\frac{\partial h}{\partial \lambda_j} = \frac{\sum_i E_q[C_{i,j}]}{\lambda_j} - \phi \quad (12)$$

$$\lambda_j = \frac{\sum_i E_q[C_{i,j}]}{\phi} \quad (13)$$

Derivitive with respect to ϕ (lagrangian multiplier) is just the constraint equation.

$$\frac{\partial h}{\partial \phi} = g(\lambda_j) \quad (14)$$

Set equal to zero and plug in value for λ_j.

$$0 = \sum_j \frac{\sum_i E_q[C_{i,j}]}{\phi} - 1 \tag{15}$$

$$1 = \sum_j \frac{\sum_i E_q[C_{i,j}]}{\phi} \tag{16}$$

$$\phi = \sum_j \sum_i E_q[C_{i,j}] \tag{17}$$

Plug back into $\lambda_j = \frac{\sum_i E_q[C_{i,j}]}{\phi}$.

$$\lambda_j = \frac{\sum_i E_q[C_{i,j}]}{\sum_j \sum_i E_q[C_{i,j}]} \tag{18}$$

1.3 Derivitive with respect to $\psi_{p,k}^{(C_{i,j})}$

Define constraint equation to optimize within. $\psi_{p,k}^{(C_{i,j})}$

$$g(\psi_{p,k}^{(0)}, \psi_{p,k}^{(1}) = (1 - \sum_p \psi_{p,k}^{0}) + (1 - \sum_p \psi_{p,k}^{1}) \tag{19}$$

$$h(\psi_{p,k}^{(0)}, \psi_{p,k}^{(1)}) = \mathrm{ELBO}(q_t(C), \boldsymbol{\theta}) + \phi g(\psi_{p,k}^{(0)}, \psi_{p,k}^{(1)}) \tag{20}$$

$$\frac{\partial h}{\partial \psi_{p,k}^{(C_{i,j})}} = \sum_i \sum_j \frac{\boldsymbol{E}_q X_{i,j,p}[C_{i,j}}{\psi_{p,k}^{(1)}} + \frac{\boldsymbol{E}_q X_{i,j,p}[1 - C_{i,j}]}{\psi_{p,k}^{(0)}} \tag{21}$$