PS2

Aina López, Nick Halliwell, Yaroslav Marchuk

2.a In this model, what are the parameters, what are the latent variables, and what is the observed data?

Parameters: ρ_k, $\beta^1_{z_i}$, $\beta^2_{z_i}$

Latent variables: z_i

Observed data: x^1_{iv} and x^2_{iv}

2.b Write down the complete data log-likelihood function.

$l_{comp}(X, z|\rho, B) = \sum_d \sum_k 1_{z_d=k}\Big[\sum_{v_1} x^1_{dv} log\beta^1_{k,v_1} + \sum_{v_2} x^2_{dv} log\beta^2_{k,v_2}\Big]$

2.c Compute the expected value of the above log-likelihood function given fixed values for the parameters.

$Q(\rho, B_1, B_2, \rho^i, B^i_1, B^i_2) = \sum_d \sum_k \hat{z_{d,k}}\Big[+\sum_{v_1} x^1_{dv_1} log\beta^1_{k,v_1} + \sum_{v_2} x^2_{dv_2} log\beta^2_{k,v_2}\Big]$

where

$\hat{z_{d,k}} = \rho_k \prod_{v_1} {\beta^1_{k,v_1}}^{x_{d,v_1}} \prod_{v_2} {\beta^2_{k,v_2}}^{x_{d,v_2}}$

2.d Maximize Q with respect to the parameter values.

The Lagrangian of the complete log-likelihood is:

$L = \sum_d \sum_k \hat{z_{d,k}}\Big[+\sum_{v_1} x^1_{dv_1} log\beta^1_{k,v_1} + \sum_{v_2} x^2_{dv_2} log\beta^2_{k,v_2}\Big] + v(1-\sum_k \rho_k) + \sum_k \lambda^1_k(1-\sum_{v_1}) + \sum_k \lambda^2_k(1-\sum_{v_2})$

- Derivative w.r.t ρ_k

$\frac{dL}{d\rho_k} = \sum_d \frac{\hat{z_{d,k}}}{\rho_k} - v = 0$

$\frac{\sum_d \hat{z_{d,k}}}{\rho_k} = v$

$\sum_d \hat{z_{d,k}} = v\rho_k$

$v\rho_k = \sum_d \hat{z_{d,k}}$

$v\sum_k \rho_k = \sum_d \sum_k \hat{z_{d,k}}$

If we compute the derivative w.r.t v and set it to zero, we will get that $\sum_k \rho_k$.

$v = \sum_d \sum_k \hat{z_{d,k}}$

Using v in the first equation we get that:

$\rho_k = \frac{\sum_d \hat{z_{d,k}}}{\sum_d \sum_k \hat{z_{d,k}}}$

- Derivative w.r.t β^1_{k,v_1}

$\frac{dL}{d\beta^1_{k,v_1}} = \sum_{d=1}^{D} \hat{z_{d,k}} \frac{x^1_{dv_1}}{\beta^1_{k,v_1}} - \lambda^1_k = 0$

$\sum_d \hat{z_{d,k}} x^1_{dv_1} = \beta^1_{k,v_1} \lambda^1_k$

Taking the derivarive w.r.t λ we get that $\sum_{v_1} \beta^1_{k,v_1} = 1$

$\sum_{v_1} \beta^1_{k,v_1} \lambda^1_k = \sum_d \hat{z_{d,k}} \sum_{v_1} x^1_{dv_1}$

$\lambda^1_k = \sum_d \hat{z_{d,k}} \sum_{v_1} x^1_{dv_1}$

Plugging this result in the initial equation we get that:

$\beta^1_{k,v_1} = \frac{\sum_d \hat{z_{d,k}} x^1_{dv_1}}{\sum_d \hat{z_{d,k}} \sum_{v_1} x^1_{dv_1}}$

- Derivative w.r.t β^2_{k,v_2}

$\frac{dL}{d\beta^2_{k,v_2}} = \sum_d \hat{z_{d,k}} \frac{x^2_{dv_2}}{\beta^2_{k,v_2}} - \lambda^2_k = 0$

$\sum_d \hat{z_{d,k}} x^2_{dv_2} = \beta^2_{k,v_2} \lambda^2_k$

Taking the derivarive w.r.t λ we get that $\sum_{v_2} \beta^2_{k,v_2} = 1$

$\sum_{v_2} \beta^2_{k,v_2} \lambda^2_k = \sum_d \hat{z_{d,k}} \sum_{v_2} x^2_{dv_2}$

$\lambda^2_k = \sum_d \hat{z_{d,k}} \sum_{v_2} x^2_{dv_2}$

Plugging this result in the initial equation we get that:

$\beta^2_{k,v_2} = \frac{\sum_d \hat{z_{d,k}} x^2_{dv_2}}{\sum_d \hat{z_{d,k}} \sum_{v_2} x^2_{dv_2}}$

2.e Write pseudo-code for implementing the EM algorithm for this model.

Initialize ρ^0, ${\beta^1_{k,v_1}}^0$ and ${\beta^2_{k,v_2}}^0$

for i in 1:Number of steps:

1. Expectation step

$Q(\rho, B_1, B_2, \rho^i, B_1^i, B_2^i) = \sum_d \sum_k \hat{z_{d,k}} \left[+ \sum_{v_1} x^1_{dv_1} log \beta^1_{k,v_1} + \sum_{v_2} x^2_{dv_2} log \beta^2_{k,v_2} \right]$

where

$\hat{z_{d,k}} = \rho_k \prod_{v_1} {\beta^1_{k,v_1}}^{x_{d,v_1}} \prod_{v_2} {\beta^2_{k,v_2}}^{x_{d,v_2}}$

2. Minimization step

$\rho^i_k = \frac{\sum_d \hat{z_{d,k}}}{\sum_d \sum_k \hat{z_{d,k}}}$

${\beta^1_{k,v_1}}^i = \frac{\sum_d \hat{z_{d,k}} x^1_{dv_1}}{\sum_d \hat{z_{d,k}} \sum_{v_1} x^1_{dv_1}}$

${\beta^2_{k,v_2}}^i = \frac{\sum_d \hat{z_{d,k}} x^2_{dv_2}}{\sum_d \hat{z_{d,k}} \sum_{v_2} x^2_{dv_2}}$

3. Convergence criterion

If $Q(\rho, B_1, B_2\rho^i, B_1^i, B_2^i) - Q(\rho, B_1, B_2\rho^i, B_1^{i-1}, B_2^{i-1}) < \epsilon$

`Stop`

Else

`Continue`