Modeling Document Collections

David Mareček

Ocrober 14, 2021

EUROPEAN UNION
European Structural and Investment Fund
Operational Programme Research,
Development and Education

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

Lagrange Multipliers

We want to find the maximum or minimum of a function $f(x)$ subjected to some equality constraint $g(x) = 0$.

We form the Lagrangian function $L(x, \lambda) = f(x) - \lambda g(x)$.

The solution corresponding to the original constrained optimization is always a saddle point of the Lagrangian function.

In our case, the equality constraint is:

$$g(\beta) = \sum_{m=1}^{M} \beta_m - 1 = 0$$

We form the following Lagrangian function:

$$L(\beta, \lambda) = \sum_{m=1}^{M} c_m \log \beta_m + \lambda(1 - \sum_{m=1}^{M} \beta_m)$$

Maximum Likelihood for Multinomial Distribution

We want to maximize the (log) likelihood

$$\log p(w|\beta) = \sum_{m=1}^{M} c_m \log \beta_m$$

We take derivatives of the Lagrangian function

$$L(\beta, \lambda) = \sum_{m=1}^{M} c_m \log \beta_m + \lambda(1 - \sum_{m=1}^{M} \beta_m)$$

By setting them to zero, we obtain

$$\frac{\partial L}{\partial \beta_m} = \frac{c_m}{\beta_m} - \lambda = 0 \quad \Rightarrow \quad \beta_m = \frac{c_m}{\lambda}$$

$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{m=1}^{M} \beta_m = 0 \quad \Rightarrow \quad \sum_{m=1}^{M} \frac{c_m}{\lambda} = \frac{n}{\lambda} = 1 \quad \Rightarrow \quad n = \lambda \quad \Rightarrow \quad \beta_m = \frac{c_m}{n}$$

Expectation Maximization and Mixture of Categoricals

We want to maximize the likelihood of the data:

$$p(w|\theta, \beta) = \prod_{d=1}^{D} \sum_{k=1}^{K} p(z_d = k|\theta) \prod_{n=1}^{N_d} p(w_{nd}|z_d = k, \beta_k)$$

However, the latent variables (document categories) are unknown.

Expectation-Maximization algorithm:

1. Initialize θ and β randomly.
2. *E-step*: For each d and k, compute responsibilities r_{kd} as probabilities $q(z_d = k|\theta, \beta)$
3. *M-step*: Maximize the likelihood of the model with weighted by the responsibilities r_{kd} from step 2 and update the parameters β and θ.
4. Repeat steps 2 and 3 unil convergence.

Expectation Maximization and Mixture of Categoricals

E-step: For each document, compute posterior distribution over categories:

$$r_{kd} = q(z_d = k) \propto p(z_d = k|\theta) \prod_{n=1}^{N_d} p(w_{nd}|z_d = k, \beta_k) = \theta_k \prod_{m=1}^{M} \beta_{km}^{c_{md}}$$

M-step: Maximize the log-likelihood weighted by the responsibilities r_{kd}:

$$\sum_{d=1}^{D} \sum_{k=1}^{K} r_{kd} \log p(w, z_d) = \sum_{k,d} r_{kd} \log[p(z_d = k|\theta) \prod_{n=1}^{N_d} p(w_{nd}|z_d = k, \beta_k)]$$

$$= \sum_{k,d} r_{kd} (\log \theta_k + \log \prod_{m=1}^{M} \beta_{km}^{c_{md}})$$

$$= \sum_{k,d} r_{kd} (\log \theta_k + \sum_{m=1}^{M} c_{md} \log \beta_{km})$$

Expectation Maximization and Mixture of Categoricals

M-step (continued): We need Lagrange multipliers to constrain the maximization of the function ensure proper distributions.

$$L_1 = \sum_{k=1}^{K} \sum_{d=1}^{D} r_{kd} (\log \theta_k + \sum_{m=1}^{M} c_{md} \log \beta_{km}) + \lambda (1 - \sum_{k'=1}^{K} \theta_{k'})$$

$$\frac{\partial L_1}{\partial \theta_k} = \sum_{d=1}^{D} r_{kd} \frac{1}{\theta_k} - \lambda = 0 \quad \Rightarrow \quad \theta_k = \frac{\sum_{d=1}^{D} r_{kd}}{\lambda} = \frac{\sum_{d=1}^{D} r_{kd}}{\sum_{k'=1}^{K} \sum_{d=1}^{D} r_{k'd}} = \frac{\sum_{d=1}^{D} r_{kd}}{D}$$

$$L_2 = \sum_{k=1}^{K} \sum_{d=1}^{D} r_{kd} (\log \theta_k + \sum_{m=1}^{M} c_{md} \log \beta_{km}) + \sum_{k'=1}^{K} \lambda_{k'} (1 - \sum_{m'=1}^{M} \beta_{k'm'})$$

$$\frac{\partial L_2}{\partial \beta_{km}} = \sum_{d=1}^{D} r_{kd} \frac{c_{md}}{\beta_{km}} - \lambda_k = 0 \quad \Rightarrow \quad \beta_{km} = \frac{\sum_{d=1}^{D} r_{kd} c_{md}}{\lambda_k} = \frac{\sum_{d=1}^{D} r_{kd} c_{md}}{\sum_{m'=1}^{M} \sum_{d=1}^{D} r_{kd} c_{m'd}}$$

Exercises

1. Let's have $K = 2$, $M = \{a, b, c\}$ and observe the following set of documents

   $$D_1 = \{a, b, b\}, \quad D_2 = \{a, c, c\}, \quad D_3 = \{a, b\}.$$

   Assume the random initialization of parameters

   $$\theta = [1/4, 3/4], \quad \beta_1 = [1/4, 1/4, 1/2], \quad \beta_2 = [1/2, 1/4, 1/4].$$

   Compute the responsibilites and the first update of parameters.

2. What happens if we initialize the parameters θ and β uniformly?

Exercises

$$r_{11} = \frac{1}{4} \cdot \frac{1}{4} \cdot \left(\frac{1}{4}\right)^2 = \frac{1}{256} \qquad r_{21} = \frac{3}{4} \cdot \frac{1}{2} \cdot \left(\frac{1}{4}\right)^2 = \frac{6}{256}$$

$$r_{12} = \frac{1}{4} \cdot \frac{1}{4} \cdot \left(\frac{1}{2}\right)^2 = \frac{4}{256} \qquad r_{22} = \frac{3}{4} \cdot \frac{1}{2} \cdot \left(\frac{1}{4}\right)^2 = \frac{6}{256}$$

$$r_{13} = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{4}{256} \qquad r_{23} = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{24}{256}$$

$$\theta_1 \sim r_{11} + r_{12} + r_{13} = \frac{9}{256} \sim \frac{1}{5} \qquad \theta_2 \sim r_{21} + r_{22} + r_{23} = \frac{36}{256} \sim \frac{4}{5}$$

$$\beta_{1a} \sim \frac{1 \cdot 1}{256} + \frac{1 \cdot 4}{256} + \frac{1 \cdot 4}{256} = \frac{9}{256} \sim \frac{9}{23} \qquad \beta_{2a} \sim \frac{1 \cdot 6}{256} + \frac{1 \cdot 6}{256} + \frac{1 \cdot 24}{256} = \frac{36}{256} \sim \frac{3}{7}$$

$$\beta_{1b} \sim \frac{2 \cdot 1}{256} + \frac{0 \cdot 4}{256} + \frac{1 \cdot 4}{256} = \frac{6}{256} \sim \frac{6}{23} \qquad \beta_{2b} \sim \frac{2 \cdot 6}{256} + \frac{0 \cdot 6}{256} + \frac{1 \cdot 24}{256} = \frac{36}{256} \sim \frac{3}{7}$$

$$\beta_{1c} \sim \frac{0 \cdot 1}{256} + \frac{2 \cdot 4}{256} + \frac{0 \cdot 4}{256} = \frac{8}{256} \sim \frac{8}{23} \qquad \beta_{2c} \sim \frac{0 \cdot 6}{256} + \frac{2 \cdot 6}{256} + \frac{0 \cdot 24}{256} = \frac{12}{256} \sim \frac{1}{7}$$