Modeling Document Collections

David Mareček

October 26, 2020
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:

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

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

$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{m=1}^{M} \beta_m = 0 \Rightarrow \sum_{m=1}^{M} \frac{c_m}{\lambda} = \frac{n}{\lambda} = 1 \Rightarrow n = \lambda \Rightarrow \beta_m = \frac{c_m}{n}$$
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 $\theta$ and $\beta$ uniformly.
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 $\beta$ and $\theta$.
4. Repeat steps 2 and 3 until 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 $r_{kd}$:

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

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

$$ = \sum_{k,d} r_{kd} (\log \theta_k + \sum_{m=1}^{M} c_{md} \log \beta_{km}) $$
**M-step (continued):** We need Lagrange multipliers to constrain the maximization of the function to ensure proper distributions.

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

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

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

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