EM Algorithm for PLSA: M-Step

Hidden Variable (=topic indicator): $z_{d,w} \in \{B, 1, 2, \ldots, k\}$

ML Estimate based on "allocated" word counts to topic θ_j

Re-estimated **probability** of **doc d** covering **topic** θ_j

$$\pi_{d,j}^{(n+1)} = \frac{\sum_{w \in V} c(w,d)(1 - p(z_{d,w} = B))p(z_{d,w} = j)}{\sum_{j'} \sum_{w \in V} c(w,d)(1 - p(z_{d,w} = B))p(z_{d,w} = j')}$$

$$p^{(n+1)}(w \mid \theta_j) = \frac{\sum_{d \in C} c(w,d)(1 - p(z_{d,w} = B))p(z_{d,w} = j)}{\sum_{w' \in V} \sum_{d \in C} c(w',d)(1 - p(z_{d,w'} = B))p(z_{d,w'} = j)}$$

Re-estimated **probability** of **word w** for **topic** θ_j