Probabilistic Latent Semantic Analysis (PLSA)

Percentage of background words (known)

Background LM (known)

Coverage of topic θ $_j$ in doc d

Prob. of word w in topic θ $_j$

$$p_d(w) = \lambda_B p(w \mid \theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} \pi_{d,j} p(w \mid \theta_j)$$

$$\log p(d) = \sum_{w \in V} c(w,d) \log[\lambda_B p(w \mid \theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} \pi_{d,j} p(w \mid \theta_j)]$$

$$\log p(C \mid \Lambda) = \sum_{d \in C} \sum_{w \in V} c(w,d) \log[\lambda_B p(w \mid \theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} \pi_{d,j} p(w \mid \theta_j)]$$

Unknown Parameters: $\Lambda=(\{\pi_{d,j}\}, \{\theta_j\})$, j=1, …, k

How many unknown parameters are there in total?