Maximum Likelihood Estimator

Likelihood Functions

Single: $P(z|\theta) = \prod_{k=1}^{K} \theta_k^{z_k}$ Dataset: $P(D|\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} \theta_k^{z_{nk}}$

Log-Likelihood

$$\log P(D|\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log \theta_k = \sum_{k=1}^{K} N_k \log \theta_k$$

$N_k = \sum_{n=1}^{N} z_{nk}$ count samples in state K

Optimization with Constraint

Constraint: $\sum_{k=1}^{K} \theta_k = 1$

$$\mathcal{L} = \sum_{k=1}^{K} N_k \log \theta_k + \lambda \left(1 - \sum_{k=1}^{K} \theta_k\right)$$

Derivative $\frac{N_k}{\theta_k} - \lambda = 0$ $\theta_k = \frac{N_k}{\lambda}$

Use constraint: $\sum_k \theta_k = 1$

$$\sum_{k=1}^{K} \frac{N_k}{\lambda} = 1 \Rightarrow \lambda = \sum_{k=1}^{K} N_k = N$$

ML Estimator: $\theta_k^{ML} = \frac{N_k}{N}$

Maximum A Posteriori Estimator

Posterior Distribution

$$p(\theta \mid D, \alpha) \propto p(D \mid \theta) \cdot p(\theta \mid \alpha)$$

$$\propto \left[\prod_{k=1}^{K} \theta_k^{N_k} \right] \cdot \left[\prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \right]$$

$$= \prod_{k=1}^{K} \theta_k^{N_k + \alpha_k - 1}$$

Log Posterior

$$\log p(\theta \mid D, \alpha) = \sum_{k=1}^{K} (N_k + \alpha_k - 1) \log \theta_k + \text{const}$$

Optimization with Constraint

Lagrangian $\quad \mathcal{L} = \sum_{k=1}^{K} (N_k + \alpha_k - 1) \log \theta_k + \lambda \left(1 - \sum_{k=1}^{K} \theta_k\right)$

Derivative: $\quad \dfrac{N_k + \alpha_k - 1}{\theta_k} - \lambda = 0$

$$\theta_k = \frac{N_k + \alpha_k - 1}{\lambda}$$

Using Constraint

$$\lambda = \sum_{k=1}^{K} (N_k + \alpha_k - 1) = N + \sum_{k=1}^{K} \alpha_k - K$$

Result

$$\theta_k^{MAP} = \frac{N_k + \alpha_k - 1}{N + \alpha_0 - K}$$

Special Cases

Case 1: Uniform Prior ($\alpha_k = 1$ for all k)

$$\theta_k^{MAP} = \frac{N_k}{N} = \theta_k^{ML}$$

MAP reduces to ML

Case 2: General $\alpha_k > 1$

$$\theta_k^{MAP} = \frac{N_k + (\alpha_k - 1)}{N + (\alpha_0 - K)}$$