Divergence Minimization

$$\theta^* = argmin_\theta[\frac{1}{n}\sum_{i=1}^{n} D(\theta||\theta_{D_i})] - \lambda D(\theta||\theta_C)$$

Subject to $\sum_{w\in V} p(w|\theta) = 1$

$$\frac{\partial g(\theta)}{\partial \mu} = 0$$

$$\theta^* = argmin_\theta[\frac{1}{n}\sum_{i=1}^{n} H(\theta;\theta_{D_i}) - H(\theta)] - \lambda H(\theta;\theta_C) + \lambda H(\theta)$$

$$= argmin_\theta \frac{1}{n}\sum_{i=1}^{n}[H(\theta;\theta_{D_i}) - \lambda H(\theta;\theta_C)] - (1-\lambda)H(\theta)$$

$\theta_w = p(w|\ \theta)$

$$g(\theta) = -\frac{1}{n}\sum_{i=1}^{n}[\sum_{w\in V} p(w|\theta)[\log p(w|\theta_{D_i}) - \lambda \log p(w|\theta_C)\ ]] + (1-\lambda)\ \sum_{w\in V} p(w|\theta)\log p(w|\theta) + \mu[\sum_{w\in V} p(w|\theta) - 1]$$

$$\frac{\partial g(\theta)}{\partial \theta_w} = -\frac{1}{n}\sum_{i=1}^{n}[\log p(w|\theta_{D_i}) - \lambda \log p(w|\theta_C)]\ + (1-\lambda)\ [\log p(w|\theta) + 1]\ + \mu = 0$$

Solution:

$$\boldsymbol{\theta_w = p(w|\theta)} \quad \propto \exp[\frac{1}{1-\lambda}\frac{1}{n}\sum_{i=1}^{n}\log p(w|\theta_{D_i}) - \frac{\lambda}{1-\lambda}\log p(w|\theta_C)]$$