$$\theta_w = \frac{\exp \boldsymbol{\lambda}^\top \mathbf{f}(w)}{\sum_{w'} \exp \boldsymbol{\lambda}^\top \mathbf{f}(w')}$$

$$\mathcal{L}(\boldsymbol{\lambda}, \mathbf{x}) = \sum_{i=1}^{|\mathbf{x}|} \boldsymbol{\lambda}^\top \mathbf{f}(w) - \log \sum_{w'} \exp \boldsymbol{\lambda}^\top \mathbf{f}(w')$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_k}(\boldsymbol{\lambda}, \mathbf{x}) = \sum_{i=1}^{|\mathbf{x}|} f_k(w) - \frac{\sum_{w'} f_k(w') \exp \boldsymbol{\lambda}^\top \mathbf{f}(w')}{\sum_{w''} \exp \boldsymbol{\lambda}^\top \mathbf{f}(w'')}$$

$$= \sum_{i=1}^{|\mathbf{x}|} f_k(w) - \mathbb{E}_{p(w';\boldsymbol{\lambda})} f_k(w')$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_k}(\boldsymbol{\lambda}, \mathbf{x}) = 0$$