CRF w/ feature-based potentials

Data:

CRF: $\mathcal{D} = \{x^{(n)}, y^{(n)}\}_{n=1}^{N}$

MRF: $\mathcal{D} = \{y^{(n)}\}_{n=1}^{N}$

Model:

CRF: $p(\vec{y} | \vec{x}, \vec{\theta}) = \frac{1}{Z(\vec{x}, \vec{\theta})} \prod_{c \in \mathcal{C}} \psi_c(\vec{y}_c, \vec{x})$ where $\psi_c(\vec{y}_c, \vec{x}) = \exp[\vec{\theta} \cdot \vec{f}_c(\vec{y}_c, \vec{x})]$

MRF: $p(\vec{y} | \vec{\theta}) = \frac{1}{Z(\vec{\theta})} \prod_{c} \psi_c(\vec{y}_c)$ where $\psi_c(\vec{y}_c) = \exp[\vec{\theta} \cdot f_c(\vec{y}_c)]$

Avg. Log-likelihood:

CRF: $\tilde{\ell}(\vec{\theta}; \mathcal{D}) = \frac{1}{N} \sum_{n=1}^{N} \log p(\vec{y}^{(n)} | \vec{x}^{(n)}, \vec{\theta})$

MRF: $\tilde{\ell}(\vec{\theta}; \mathcal{D}) = \frac{1}{N} \sum_{n=1}^{N} \log p(\vec{y}^{(n)} | \vec{\theta})$

inference

Derivative:

CRF $\frac{d\tilde{\ell}(\vec{\theta}; \mathcal{D})}{d\theta_k} = \frac{1}{N} \left(\sum_{n=1}^{N} \sum_{c} f_{c,k}(\vec{y}_c^{(n)}, \vec{x}^{(n)}) \right) - \frac{1}{N} \left(\sum_{n=1}^{N} \sum_{c} \sum_{y_c} p(\vec{y}_c | \vec{x}^{(n)}) f_{c,k}(\vec{y}_c, \vec{x}^{(n)}) \right)$

MRF: $\frac{d\tilde{\ell}(\vec{\theta}; \mathcal{D})}{d\theta_k} = \frac{1}{N} \left(\sum_{n} \sum_{c} f_{c,k}(\vec{y}_c^{(n)}) \right) - \frac{1}{N} \left(\sum_{n} \sum_{c} \sum_{y_c} p(\vec{y}_c) f_{c,k}(\vec{y}_c) \right)$