MRFs w/ tabular potentials
(i.e. discret vars)

① Log-likelihood (one instance):

$$p(\vec{x}|\theta) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c)$$

$$\log p(\vec{x}|\theta) = \sum_{c \in C} \log \psi_c(x_c) - \log Z$$

② Derivatives (one instance):

$$\frac{d \log p(\vec{x}|\theta)}{d \psi_c(\vec{x}_c')} = \frac{d}{d\psi_c(x_c')} \left(\sum_{d \in C} \log \psi_d(x_d) - \log Z \right)$$

$$= \frac{\delta(x_c, x_c')}{\psi_c(x_c)} - \frac{d}{d\psi_c(x_c} \log Z$$

$$= \boxed{}$$

$$\frac{d \log Z}{d\psi_c(\vec{x}_c)} = \frac{1}{Z} \frac{dZ}{d\psi_c(x_c)}$$ (chain rule)

$$= \frac{1}{Z} \frac{d}{d\psi_c(x_c)} \left(\sum_{x'} \prod_d \psi_d(x_d') \right)$$

$$= \frac{1}{Z} \sum_{x' : x_c' = x_c} \frac{d}{d\psi_c(x_c)} \left(\prod_d \psi_d(x_d') \right)$$

$$= \frac{1}{Z} \sum_{x' : x_c' = x_c} \frac{1}{\psi_c(x_c)} \left(\prod_d \psi_d(x_d') \right)$$ mult by $1 = \frac{\psi_c(x_c)}{\psi_c(x_c)}$

$$= \frac{1}{\psi_c(x_c)} \sum_{x' : x_c' = x_c} \underbrace{\left(\frac{1}{Z} \prod_d \psi_d(x_d') \right)}_{p(x'|\theta)}$$

$$= \frac{p(x_c)}{\psi_c(x_c)}$$

③ Log-likelihood (data):

$$\ell(\theta|D) = \sum_{n=1}^{N} \log p(x^{(n)}|\theta)$$

$$= \sum_{\vec{x}} m(\vec{x}) \log p(\vec{x}|\theta)$$

where $m(\vec{x})$ = # occurrences of $\vec{x}$ in D
$m(x_c)$ = # occurrences of x_c in D

$$= \left(\sum_{c \in C} \sum_{x_c} m(x_c) \log \psi_c(x_c) \right) - N \log Z$$

④ Derivatives (data):

$$\frac{d\ell(\theta|D)}{d\psi_c(x_c)} = \frac{m(x_c)}{\psi_c(X_c)} - N \frac{p(x_c)}{\psi_c(x_c)}$$

MRF w/ Feature-based Potentials

Suppose $\psi_c(\vec{x}_c) = \exp[\vec{\theta} \cdot \vec{f}_c(\vec{x}_c)]$

$= \exp[\sum_{k=1}^{K} \theta_k f_{c,k}(\vec{x}_c)]$

Avg. Log-likelihood:

Let $\tilde{\ell}(\theta|D) \triangleq \frac{1}{N} \sum_{n=1}^{N} \log p(x^{(n)}|\theta)$

$= \sum_{\vec{x}} \tilde{p}(\vec{x}) \log p(x^{(n)}|\theta)$

$\tilde{p}(\vec{x}) \triangleq \frac{\tilde{m}(\vec{x})}{N}$

$= \left(\sum_c \sum_{x_c} \tilde{p}(\vec{x}_c) \log \psi_c(\vec{x}_c)\right) - \log Z$

Scribes

counting

inference

Derivatives:

$$\frac{d\tilde{\ell}(\theta|D)}{d\theta_k} = \left(\sum_c \sum_{x_c} \tilde{p}(\vec{x}_c) f_{c,k}(\vec{x}_c)\right) - \left(\sum_c \sum_{x_c} p(\vec{x}_c) f_{c,k}(\vec{x}_c)\right)$$

$$= \mathbb{E}_{\vec{x} \sim \tilde{p}(\cdot|D)}[f_{\cdot,k}(\vec{x})] - \mathbb{E}_{\vec{x} \sim p(\cdot|\theta)}[f_{\cdot,k}(\vec{x})]$$

data dist. model dist.

where $f_{\cdot,k}(\vec{x}) = \sum_c f_{c,k}(\vec{x}_c)$

Regularization

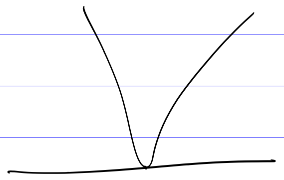
$\log p(D|\theta) + \log p(\theta)$

$$J(\vec{\theta}) = \ell(\vec{\theta}|D) + r(\vec{\theta})$$

L2: $r(\vec{\theta}) = \frac{\lambda}{2} (\|\theta\|_2)^2 = \frac{\lambda}{2} \sum_{k=1}^{K} \theta_k^2$

equivalent to MAP estimation w/ $p(\theta) = \mathcal{N}(0, \frac{1}{\lambda})$

L1: $r(\vec{\theta}) = \lambda \|\theta\|_1 = \lambda \sum_{k=1}^{K} |\theta_k|$



w/ $p(\theta)$ = Laplace

MLE by Guessing

Condition: $p^{*}_{MLE}(x_c) = \dfrac{m(x_c)}{N} \triangleq \tilde{p}(x_c)$

Example 1

- $X_1 - X_2 - X_3$

$$p(\vec{x}) = \psi_{12}(x_1, x_2)\, \psi_{23}(x_2, x_3)$$

- Random guess

$$\hat{p}(x_1, x_2, x_3) = \underbrace{\tilde{p}(x_1, x_2)}_{\psi_{12}(x_1,x_2)} \underbrace{\frac{\tilde{p}(x_2, x_3)}{\tilde{p}(x_2)}}_{\psi_{23}(x_2,x_3) = p(x_3|x_2)}$$

- $\hat{p}(x_1, x_2) = \sum_{x_3} \hat{p}(\vec{x})$

$$= \frac{\tilde{p}(x_1, x_2)}{\cancel{\tilde{p}(x_2)}} \cancel{\sum_{x_3} \tilde{p}(x_2, x_3)}$$

$$= \tilde{p}(x_1, x_2)$$

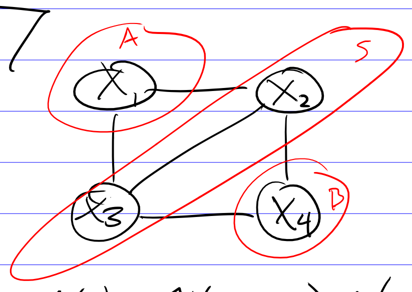
$$\hat{p}(x_2, x_3) = \sum_{x_1} \hat{p}(\vec{x})$$

$$= \frac{\tilde{p}(x_2, x_3)}{\cancel{\tilde{p}(x_2)}} \cancel{\sum_{x_1} \tilde{p}(x_1, x_2)}$$

$$= \tilde{p}(x_2, x_3)$$

$\Longrightarrow \hat{p} = p^{*}_{ME}$

Example 2



Guess: $\hat{p}(\vec{x}) = \dfrac{\tilde{p}(x_1, x_2, x_3)\, \tilde{p}(x_2, x_3, x_4)}{\tilde{p}(x_2, x_3)}$

Verify condition