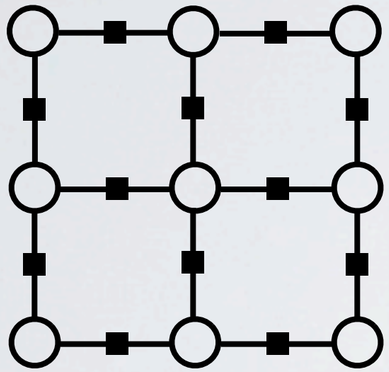
Neural networks

Training CRFs - general conditional random field

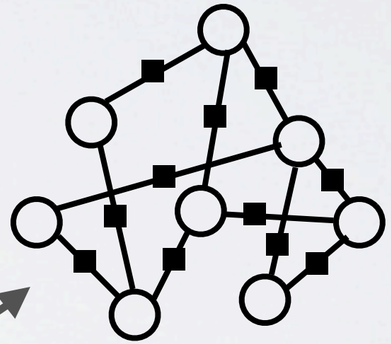
GENERAL CRF

Topics: CRFs in general

- We don't have to restrict the CRF structure to linear chains



$$p(\mathbf{y}|\mathbf{X}) = \frac{1}{Z(\mathbf{X})} \prod_f \Psi_f(\mathbf{y}, \mathbf{X})$$



Grid structure
(pixels in image)

General pair-wise structure
(webpages sharing a link)

- We could also have n-ary factors, with $n>2$

GENERAL CRF

Topics: CRFs in general

- Gradients in general CRFs always take the form:

make $y^{(t)}$ more likely

$$\frac{\partial -\log p(\mathbf{y}^{(t)}|\mathbf{X}^{(t)})}{\partial \theta} = -\Big(\overbrace{\textstyle\sum_f \frac{\partial}{\partial \theta} \log \Psi_f(\mathbf{y}^{(t)}, \mathbf{X}^{(t)})}$$

$$\underbrace{- \mathrm{E}_{\mathbf{y}} \left[\textstyle\sum_f \frac{\partial}{\partial \theta} \log \Psi_f(\mathbf{y}, \mathbf{X}^{(t)}) \,\middle|\, \mathbf{X}^{(t)} \right]}\Big)$$

make everything less likely

- The expectation over $\mathbf{y}$ will often need to be approximated, using loopy belief propagation
  - it will often involve only a few of the y_k variables

(LOOPY) BELIEF PROPAGATION

Topics: CRFs in general

- Marginals can be approximated with:

$$p(y_k|\mathbf{X}) = \frac{\exp(\log \phi_f(y_k) + \sum_{f' \in \mathrm{Ne}(k) \setminus f} \log \mu_{f' \to k}(y_k))}{\sum_{y'_k} \exp(\log \phi_f(y'_k) + \sum_{f' \in \mathrm{Ne}(k) \setminus f} \log \mu_{f' \to k}(y'_k))}$$

- In general, an approximated marginal is computed by

1. summing all the log-factors that involve only the y_k variables of interest
2. summing all the log-messages coming into the y_k variables from other factors
3. exponentiating
4. renormalizing