Neural networks

Conditional random fields - factors, sufficient statistics and linear CRF

LINEAR CHAIN CRF

Topics: factor, sufficient statistic

- CRFs are often written in the standard form:

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

factors or compatibility functions

- A parametrization of the factor that is often used is:

$$\Psi_f(\mathbf{y}, \mathbf{X}) = \exp\left(\sum_s \theta_{f,s} \; \underbrace{t_{f,s}(\mathbf{y}, \mathbf{X})}\right)$$

sufficient statistics

LINEAR CHAIN CRF

Topics: factor, sufficient statistic

- With hidden units, the CRF factors could be:

$$\Psi_f(\mathbf{y}, \mathbf{X}) = \begin{cases} \phi_f(y_k, \mathbf{x}_{k-1}) = \exp\left(a^{(L+1,-1)}(\mathbf{x}_{k-1})_{y_k}\right) \\ \phi_f(y_k, \mathbf{x}_k) = \exp\left(a^{(L+1,0)}(\mathbf{x}_k)_{y_k}\right) \\ \phi_f(y_k, \mathbf{x}_{k+1}) = \exp\left(a^{(L+1,+1)}(\mathbf{x}_{k+1})_{y_k}\right) \\ \phi_f(y_k, y_{k+1}) = \exp\left(V_{y_k, y_{k+1}}\right) \end{cases}$$

- There is no simple form for the sufficient statistics

LINEAR CHAIN CRF

Topics: factor, sufficient statistic

- With hidden units, the CRF factors could be:

unary factors

$$\Psi_f(\mathbf{y}, \mathbf{X}) = \begin{cases} \boxed{\phi_f(y_k, \mathbf{x}_{k-1})} = \exp\left(a^{(L+1,-1)}(\mathbf{x}_{k-1})_{y_k}\right) \\ \boxed{\phi_f(y_k, \mathbf{x}_k)} = \exp\left(a^{(L+1,0)}(\mathbf{x}_k)_{y_k}\right) \\ \boxed{\phi_f(y_k, \mathbf{x}_{k+1})} = \exp\left(a^{(L+1,+1)}(\mathbf{x}_{k+1})_{y_k}\right) \\ \phi_f(y_k, y_{k+1}) = \exp\left(V_{y_k, y_{k+1}}\right) \end{cases}$$

- There is no simple form for the sufficient statistics

LINEAR CHAIN CRF

Topics: factor, sufficient statistic

- With hidden units, the CRF factors could be:

unary factors

$$\Psi_f(\mathbf{y}, \mathbf{X}) = \begin{cases} \boxed{\phi_f(y_k, \mathbf{x}_{k-1})} = \exp\left(a^{(L+1,-1)}(\mathbf{x}_{k-1})_{y_k}\right) \\ \boxed{\phi_f(y_k, \mathbf{x}_k)} = \exp\left(a^{(L+1,0)}(\mathbf{x}_k)_{y_k}\right) \\ \boxed{\phi_f(y_k, \mathbf{x}_{k+1})} = \exp\left(a^{(L+1,+1)}(\mathbf{x}_{k+1})_{y_k}\right) \\ \boxed{\phi_f(y_k, y_{k+1})} = \exp\left(V_{y_k, y_{k+1}}\right) \end{cases}$$

pairwise factors

- There is no simple form for the sufficient statistics

LINEAR CHAIN CRF

Topics: linear CRF

- In a linear (linear chain) CRF (i.e. without hidden units), the CRF factors can be written as:

$$\Psi_f(\mathbf{y}, \mathbf{X}) = \begin{cases} \phi_f(y_k) = \exp\left(b_c^{(L+1)} 1_{y_k=c}\right) \\ \phi_f(y_k, x_{k-1,i}) = \exp\left(W_{c,i}^{(L+1,-1)} x_{k-1,i} 1_{y_k=c}\right) \\ \phi_f(y_k, x_{k,i}) = \exp\left(W_{c,i}^{(L+1,0)} x_{k,i} 1_{y_k=c}\right) \\ \phi_f(y_k, x_{k+1,i}) = \exp\left(W_{c,i}^{(L+1,+1)} x_{k+1,i} 1_{y_k=c}\right) \\ \phi_f(y_k, y_{k+1}) = \exp\left(V_{c,c'} 1_{y_k=c} 1_{y_{k+1}=c'}\right) \end{cases}$$

LINEAR CHAIN CRF

Topics: linear CRF

- In a linear (linear chain) CRF (i.e. without hidden units), the CRF factors can be written as:

$$\Psi_f(\mathbf{y},\mathbf{X}) = \begin{cases} \phi_f(y_k) = \exp\left(b_c^{(L+1)} \boxed{1_{y_k=c}}\right) \\ \phi_f(y_k, x_{k-1,i}) = \exp\left(W_{c,i}^{(L+1,-1)} \boxed{x_{k-1,i} 1_{y_k=c}}\right) \\ \phi_f(y_k, x_{k,i}) = \exp\left(W_{c,i}^{(L+1,0)} \boxed{x_{k,i} 1_{y_k=c}}\right) \\ \phi_f(y_k, x_{k+1,i}) = \exp\left(W_{c,i}^{(L+1,+1)} \boxed{x_{k+1,i} 1_{y_k=c}}\right) \\ \phi_f(y_k, y_{k+1}) = \exp\left(V_{c,c'} \boxed{1_{y_k=c} 1_{y_{k+1}=c'}}\right) \end{cases}$$

sufficient statistics

LINEAR CHAIN CRF

Topics: linear CRF

- In a linear (linear chain) CRF (i.e. without hidden units), the CRF factors can be written as:

$$\Psi_f(\mathbf{y}, \mathbf{X}) = \begin{cases} \phi_f(y_k) = \exp\left(\boxed{b_c^{(L+1)}}\,\boxed{1_{y_k=c}}\right) \\ \phi_f(y_k, x_{k-1,i}) = \exp\left(\boxed{W_{c,i}^{(L+1,-1)}}\,\boxed{x_{k-1,i} 1_{y_k=c}}\right) \\ \phi_f(y_k, x_{k,i}) = \exp\left(\boxed{W_{c,i}^{(L+1,0)}}\,\boxed{x_{k,i} 1_{y_k=c}}\right) \\ \phi_f(y_k, x_{k+1,i}) = \exp\left(\boxed{W_{c,i}^{(L+1,+1)}}\,\boxed{x_{k+1,i} 1_{y_k=c}}\right) \\ \phi_f(y_k, y_{k+1}) = \exp\left(\boxed{V_{c,c'}}\,\boxed{1_{y_k=c} 1_{y_{k+1}=c'}}\right) \end{cases}$$

parameters $\theta_{f,s}$ | sufficient statistics