Conditional Random Field (CRF)

Jacob Goldberger

logistic regression:

$$p(y=i|x) = \frac{\exp(a_i^\top x + b_i)}{\sum_j \exp(a_j^\top x + b_j)}, \qquad \log p(y|x) \propto y^\top A x + y^\top b$$

$$L(A,b) = \sum_t \log p(y_t|x_t;A,b), \qquad \frac{\partial L}{\partial a_i} = \sum_t (1_{\{y_t=i\}} - p(y_t=i|x_t))x_t$$

Log-linear linear-time CRF:

$$p(y_1=i_1,...,y_n=i_n|x) = \frac{1}{Z(x)}\exp(\sum_t (a_{i_t}^\top x_t + b_{i_t} + c_{i_t,i_{t-1}}))$$

$$p(y_1,...,y_n|x_1,...,x_n) = \frac{1}{Z(x)}\exp(\sum_t (y_t^\top A x_t + y_t^\top b + y_t^\top C y_{t-1}))$$

Linear time CRF:

$$p(y_1,..,y_n|x) = \frac{1}{Z(x)}\psi_1(y_1,x)\prod_{t=2}^n \psi_t(y_{t-1},y_t,x)$$

Forward-Backward Algorithm: $p(y_t=i|x) \propto \alpha_t(i)\beta_t(i)$

$$\begin{aligned}
\alpha_1(j) &= \psi_1(j,x) \\
\alpha_t(j) &= \sum_{i=1}^k \alpha_{t-1}(i)\psi_t(i,j,x) && t=2,...,n \\
\beta_n(j) &= 1 \\
\beta_t(j) &= \sum_{i=1}^k \beta_{t+1}(i)\psi_{t+1}(j,i,x) && t=n-1,..,1 \\
Z(x) &= \sum_i \alpha_n(i)
\end{aligned}$$

Viterbi Algorithm: $\hat{y}_1,...,\hat{y}_n = \arg\max_y p(y_1,...,y_n|x)$

$$\begin{aligned}
v_1(j) &= \log\psi_1(j,x) \\
v_t(j) &= \max_{i=1}^k (v_{t-1}(i) + \log\psi_t(i,j,x)) && t=2,...,n \\
\delta_t(j) &= \arg\max_{i=1}^k (v_{t-1}(i) + \log\psi_t(i,j,x)) \\
\hat{x}_n &= \arg\max_{j=1}^k v_n(j) \\
\hat{x}_t &= \delta_{t+1}(\hat{x}_{t+1}) && t=n-1,...,1
\end{aligned}$$

General CRF model:

$$p(y_1, .., y_n|x) = \frac{1}{Z(x)} \prod_i \psi_i(y_i, x) \prod_{ij \in E} \psi_{ij}(y_i, y_j, x)$$

Belief Propagation (Sum Product):

$$\begin{aligned} m_{j\to i}(y_i) &= \sum_{y_j} (\psi_j(y_j, x)\psi_{ij}(y_i, y_j, x) \prod_{k \in N(j)\setminus i} m_{k\to j}(y_j)) \\ b_i(y_i) &= \psi_i(y_i, x) \prod_{k \in N(i)} m_{k\to i}(y_i) \end{aligned}$$

Belief Revision (Max Product):

$$\begin{aligned} m_{j\to i}(y_i) &= \max_{y_j} (\psi_j(y_j, x)\psi_{ij}(y_i, y_j, x) \prod_{k \in N(j)\setminus i} m_{k\to j}(y_j)) \\ \hat{y}_i &= \arg\max_{y_i} (\psi_i(y_i, x) \prod_{k \in N(i)} m_{k\to i}(y_i)) \end{aligned}$$

Iterated conditional mode (ICM) algorithm:

$$\hat{y}_i = \arg\max_{y_i} (\psi_i(y_i, x) \prod_{j \in N(i)} \psi_{ij}(y_i, y_j, x))$$

Gibbs Sampling:

$$p(y_i|y\setminus y_i, x) \propto \psi_i(y_i, x) \prod_{j \in N(i)} \psi_{ij}(y_i, y_j, x)$$

$$\hat{p}(y_i = a) = \frac{1}{n} |\{1 \leq t \leq n | y_i(t) = a\}|$$

Mean Field Algorithm:

$$q_i(y_i) \propto \exp\left(\log \psi_i(y_i, x) + \sum_{j \in N(i)} \sum_{y_j} q_j(y_j) \log \psi_{ij}(y_i, y_j, x)\right)$$

Maximum-Entropx Markov model (MEMM):

$$p(y_1, ..., y_n|x_1, ..., x_n) = p(y_1|x_1) \prod_t p(y_t|x_t, y_{t-1})$$

$$\log p(y_t|x_t, y_{t-1}) \propto y_t^\top A x_t + y_t^\top b + y_t^\top C y_{t-1}$$

$$p(y_t = i|x_t, y_{t-i} = k) = \frac{\exp(a_i x_t + b_i + c_{ik})}{\sum_j \exp(a_j x_t + b_j + c_{jk})}$$