$$p(y|x) = \prod_t \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(x, y_{t-1}, y_t)\right)$$

$$= \frac{1}{Z(x)} \exp \lambda^T F(x,y)$$

$$\lambda = \begin{bmatrix} \vdots \\ \lambda_i \\ \vdots \end{bmatrix} \qquad F(x,y) = \begin{bmatrix} \vdots \\ \sum_t f_i(x, y_{t-1}, y_t) \\ \vdots \end{bmatrix}$$

$$Z(x) = \sum_{y'} \exp \lambda^T F(x,y')$$

$$\mathcal{L} = \log p(y|x) = \lambda^T F(x,y) - \log Z(x)$$

$$\nabla \mathcal{L} = \nabla_\lambda \log p(y|x) = \begin{bmatrix} \vdots \\ \frac{\partial}{\partial \lambda_i} \log p(y|x) \\ \vdots \end{bmatrix}$$

$$\frac{\partial}{\partial \lambda_i} \log p(y|x) = F_i(x,y) - \frac{1}{Z(x)} \sum_{y'} \left(\exp \lambda^T F(x,y')\right) F_i(x,y')$$

$$= F_i(x,y) - \frac{1}{Z(x)} \sum_{y'} p(y'|x) F_i(x,y')$$

$$\nabla \mathcal{L} = F(x,y) - \sum_{y'} p(y'|x) F(x,y)$$

$$\alpha_t = \begin{cases} \alpha_{t-1} M_t & 0 < t \leq n \\ 1 & t = 0 \end{cases} \qquad M_t = \exp\left(\sum_i \lambda_i f_i(x, y_{t-1}, y_t)\right)$$

$$\beta_t = \begin{cases} M_{t+1} \beta_{t+1}^T & 1 \leq t < n \\ 1 & t = n \end{cases} \qquad \sum_{y'} p(y'|x) F(x,y') = \sum_t \frac{\alpha_{t-1} f_t * M_t \beta_t}{Z(x)}$$

$$Z(x) = \alpha_n$$