Conditional Random Fields (CRFs): Learning

$f_k(x, y) \overset{\text{def}}{=}$ feature function (for feature k)

$$P(y|x,w) = \frac{P(y,x|w)}{\sum_{y'} P(y',x|w)} = \frac{\exp\left\{\sum_k w_k f_k(y_t, y_{t-1}, x_t)\right\}}{\underbrace{\sum_{y'} \exp\left\{\sum_k w_k f_k(y'_t, y'_{t-1}, x_t)\right\}}_{Z(x)}}$$

$$LL(w) = \sum_{n=1}^{N} P(y_n | x_n, w)$$

$$= \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} w_k f_k(y_t, y_{t-1}, x_t) - \log Z(x_n)$$

$$\frac{\partial}{\partial w_k} LL(w) = \sum_{n=1}^{N} \sum_{t=1}^{T} \left(f_k(y^n_t, y^n_{t-1}, x^n_t) - \frac{\partial}{\partial w_k} \log Z(x^n) \right)$$

$\frac{1}{Z(x^n)}$

$$= \frac{1}{Z(x^n)} \frac{\partial}{\partial w_k} \sum_{y'} \exp \left\{ \sum_j w_j f_j (y'_t, y'_{t-1}, X^n_t) \right\}$$

$$= \sum_{y'} \exp \left\{ \sum_j w_j f_j (y'_t, y'_{t-1}, X^n_t) \right\} \cdot \frac{\partial}{\partial w_k} \sum_j w_j f_j (y'_t, y'_{t-1}, X^n_t)$$

$$= \sum_{y'} \exp \left\{ \sum_j w_j f_j (y'_t, y'_{t-1}, X^n_t) \right\} \cdot f_k (y'_t, y'_{t-1}, X^n_t)$$

$$\sum_{n=1}^{N} \sum_{t=1}^{T} \Big(f_k (y^n_t, y^n_{t-1}, X^n_t) - \frac{1}{Z(x_n)} \sum_{y'} \exp \left\{ \sum_j w_j f_j (y'_t, y'_{t-1}, X^n_t) \right\} \cdot f_k (y'_t, y'_{t-1}, x_{nt})$$

$$= \sum_{n=1}^{N} \sum_{\tau=1}^{T} \Big(f_k(y^n_\tau, y^n_{\tau-1}, x^n_\tau) -$$

$$\sum_{y'} f_k(y'_\tau, y'_{\tau-1}, x^n_t) \cdot \underbrace{\boxed{\frac{\exp\left\{ \sum_j w_j f_j (y'_t, y'_{\tau-1}, x^n_t) \right\}}{Z(x^n)}}}_{P(y' \mid x)}$$

Putting back together...

$$\frac{\partial}{\partial w_k} LL(w) =$$

$$\sum_{n=1}^{N} \sum_{\tau=1}^{T} \Big(f_k(y^n_\tau, y^n_{\tau-1}, x^n_\tau) -$$

$$\sum_{y'} f_k(y'_\tau, y'_{\tau-1}, x^n_\tau) \cdot P(y' \mid x) \Big)$$

Expectation of f_k under
$p(y \mid x, w)$ -- can compute
via Viterbi, although still
expensive!