Max Likelihood Estimation for CRFs

Parameter Estimation

$$P_\theta(Y|x) = \frac{1}{Z_x(\theta)} \tilde{P}_\theta(x, Y)$$

$$Z_x(\theta) = \sum_Y \tilde{P}_\theta(x, Y)$$

M instances

$$D = \{(x[m], y[m])\}_{m=1}^{M}$$

Data is now a set of pairs

log conditional likelihood

$$\ell_{Y|X}(\theta : D) = \sum_{m=1}^{M} \ln P_\theta(y[m] \mid x[m], \theta)$$

feature of that instance

$$\ell_{Y|X}(\theta : (x[m], y[m])) = \left(\sum_i \theta_i f_i(x[m], y[m])\right) - \ln Z_{x[m]}(\theta)$$

parameter weight w_i

summation of two expectations

$$\frac{\partial}{\partial \theta_i} \frac{1}{M} \ell_{Y|X}(\theta : D) = \frac{1}{M} \sum_{m=1}^{M} \left(f_i(x[m], y[m]) - E_\theta[f_i(x[m], Y)] \right)$$

Expectation of that feature relative to $x[m]$ and Y

sum over data instances M

$$E_\theta[f_i(\underbrace{x[m]}_{\text{fixed}}, Y)]$$

Expectation is only over variable Y

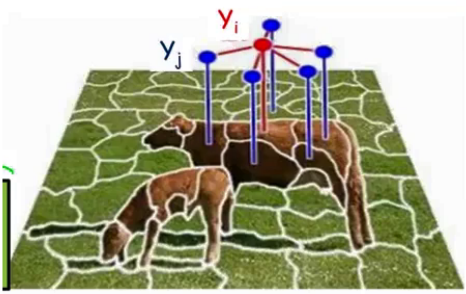
grass

Average green value for superpixel s

$$f_1(Y_s, X_s) = \mathbb{1}(Y_s = g) \times G_s$$

$$f_2(Y_s, Y_T) = \mathbb{1}(Y_s = Y_T)$$

1 if two adjacent superpixels are labeled the same value



Gradient for the FIRST parameter

$$\frac{\partial}{\partial \theta_1} = \overbrace{\sum_s \mathbb{1}\{y_s[m]=g\}\, G_s[m]}^{\text{empirical expectation}} - \overbrace{\sum_s P_\theta\left(Y_s = g \mid x[m]\right) G_s[m]}^{\text{model based expectation}}$$

$$f_1 = \mathbb{1}(Y_s = g) \times G_s$$

sum over All superpixels

single image m

Average greeness in that superpixel of that instance

$$\frac{\partial}{\partial \theta_2} = \sum_{(s,t)\in \mathcal{N}} \mathbb{1}\{y_s[m] = y_t[m]\} - \sum_{(s,t)\in \mathcal{N}} P_\theta\left(Y_s = Y_t \mid x[m]\right)$$

gradient for the second parameter

$$f_2 = \mathbb{1}(Y_s = Y_t)$$