Learning

- Scalability: Perceptron-style additive updates
- Requires two approximations:
  1. Online learning
     For example i (entity pair), define

$$\phi(\mathbf{x}, \mathbf{z}) = \sum_j \phi(x_j, z_j)$$

     Use gradient of local log likelihood for example i:

$$\frac{\partial \log O_i(\theta)}{\partial \theta_j} = E_{p(\mathbf{z}|\mathbf{x}_i, \mathbf{y}_i; \theta)}[\phi_j(\mathbf{x}_i, \mathbf{z})] - E_{p(\mathbf{y}, \mathbf{z}|\mathbf{x}_i; \theta)}[\phi_j(\mathbf{x}_i, \mathbf{z})]$$

  2. Replace expectations with maximizations