$P(X,Y)$ → generative model

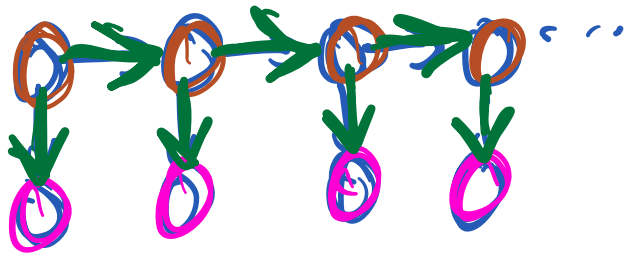
$$P(Y|X) = \frac{P(X,Y)}{P(X)}$$

$P(Y|X)$ → discriminative model

$P(X|Y)$

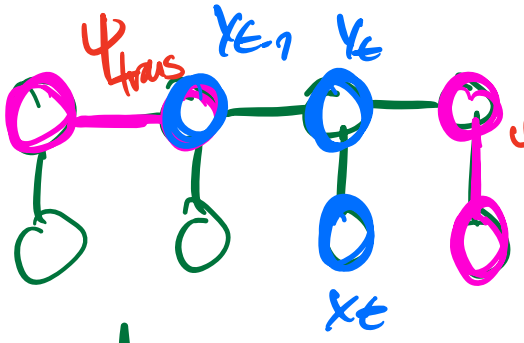
How:

Bayesian network



Idea:

→ MRF



Markov Random Field

Examples cliques of size 2.

Instead of conditional distributions, we'll use potential functions ψ → defined on cliques

$$P(x|y) = \frac{1}{Z} \prod_t \psi_{trans}(y_t, y_{t-1}) \cdot \prod \psi_{obs}(x_t, y_t)$$

potential functions

$\Rightarrow$ non-negative !!!

$$Z(x|y) = \sum_x \sum_y \prod_t \psi_{trans}(y_t, y_{t-1}) \cdot \prod \psi_{obs}(x_t, y_t)$$

compare Bayes Theorem:

$$P(y|x) = \frac{P(x|y)P(y)}{\sum_y P(x|y)P(y)}$$

How to implement the potential functions ψ?

$\Rightarrow$ log-linear combinations of basis functions

$$\psi_{obs}(x_t, y_t) = \exp\left\{ \sum_{d=1}^{D^o} w_d \, \phi_d^o(x_t, y_t) \right\}$$

$$\psi_{trans}(y_t, y_{t-1}) = \exp\left\{ \sum_{d=1}^{D^t} w_d \, \phi_d^t(y_t, y_{t-1}) \right\}$$

$$= \exp\left\{ w^T \phi(y_t, y_{t-1}) \right\}$$

inner product → $\begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}$ ← $\begin{pmatrix} \phi_1^t(\dots) \\ \phi_2^t(\dots) \\ \vdots \\ \phi_{|t|}^t(\dots) \end{pmatrix}$

⇒ Example:

trans. →
$$\phi_1^t(\underline{Y_t}, \underline{Y_{t-1}}) = \begin{cases} 1: & Y_t = N \wedge Y_{t-1} = A \\ 0: & \text{otherwise} \end{cases}$$

trans. →
$$\phi_2^t(\underline{Y_t}, \underline{Y_{t-1}}) = \begin{cases} 1: & Y_t = N \wedge Y_{t-1} = N \\ 0: & \text{otherwise} \end{cases}$$

obs →
$$\phi_{12\#}^o(X_t, Y_t) = \#\text{class in } X_t$$

$$\phi_{245}^o(X_t, Y_t) = \begin{cases} 1: & X_t \text{ contains a number} \\ 0: & \text{otherwise} \end{cases}$$

$$\phi_{1111}^o(X_t, Y_t) = \begin{cases} 1: & X_t \text{ is capitalized} \wedge \\ & Y_t \text{ is a } N \\ 0: & \text{otherwise} \end{cases}$$

⇒
$$P(X|Y) = \frac{1}{Z(X|Y)} \prod_t \exp\left(w_0^T \phi_0(X_t, Y_t)\right) \cdot$$

$$\prod_b \exp\left(w_+^T \phi_t(Y_b, Y_{b-1})\right)$$

$$= \frac{1}{Z(X|Y)} \exp\left(\sum_b w_0^T \phi_0(X_b, Y_b)\right) \times$$

↑ ↑

$$\exp\left(\sum_t w_t^T \phi_t(y_t, y_{t-1})\right)$$ ↑ up

Example: Det — N — V — Det — N

The cat chases the dog

$$\phi_0(N, \text{cat}) = \begin{pmatrix} x = \text{dog}\ ? \\ x = \text{cat}\ ? \\ x = \text{cat} \wedge y = V\ ? \\ x = \text{cat} \wedge y = N\ ? \\ \vdots \end{pmatrix} \Rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ \vdots \end{pmatrix}$$

y x

$$P(x,y) = \frac{1}{Z(x,y)} \cdot \exp\left(\begin{pmatrix} \sum_t w_0^T \phi_0(x_t, y_t) \\ \sum_t w_t^T \phi_t(y_t, y_{t-1}) \end{pmatrix}\right)$$

$D_0 \times D_t$

$$\begin{pmatrix} \phi_{01}(x_t, y_t) \\ \vdots \\ \phi_{0D_0}(x_t, y_t) \\ \phi_{t1}(y_t, y_{t-1}) \\ \vdots \\ \phi_{tD_t}(y_t, y_{t-1}) \end{pmatrix}$$

$$\Rightarrow P(x,y) = \frac{1}{Z(x,y)} \exp\left[w^T \begin{pmatrix} \sum_t \phi_0(x_t, y_t) \\ \sum_t \phi_t(y_t, y_{t-1}) \end{pmatrix}\right]$$

$w = \begin{pmatrix} w_0 \\ w_t \end{pmatrix}$ $\quad \Phi(x, y)$

joint input output space!!

feature representation of input X and output Y

$\Rightarrow$ move to conditional model: discard normalization in Z over inputs

$$\Rightarrow Z = \sum_y \exp\left(w^T \Phi(x, y)\right)$$

$\Rightarrow$ model:

$$P(y|x) = \frac{1}{Z(x)} \exp\left(w^T \Phi(x, y)\right)$$

$\Rightarrow$ following this approach leads to conditional random fields (CRFs)

Lafferty et. al

$\Rightarrow$ or simplify $P(y|x)$ by

- MAP (maximum a posteriori) approach

$\Rightarrow \operatorname{argmax}_y P(y|x)$

$$= \operatorname{argmax}_y \frac{1}{Z(x)} \exp\{w^T \phi(x,y)\}$$

$$= \operatorname{argmax}_y \; w^T \phi(x,y)$$

---

perceptron $\Leftrightarrow$ primal / dual

probably Collins & Duffy (2002)

$f(x) = w^T x$

$w \leftarrow 0$

$w \leftarrow w + yx$

$w \leftarrow 0$

$w \leftarrow 0 + y_1 x_1$

$w \leftarrow 0 + y_1 x_1 + y_2 x_2$

$w \leftarrow 0 + y_1 x_1 + y_2 x_2 + y_5 x_5$

dual perceptron:

$\alpha_1, \alpha_2, \ldots \alpha_N \leftarrow 0$

update: $\alpha_n \leftarrow \alpha_n + 1$

model: $w = \sum \alpha_n y_n x_n$ (!)

prediction: $f(x) = w^T x$

$$= \sum \alpha_n y_n x_n^T x$$

$$= \sum \alpha_n y_n K(x_n, x)$$