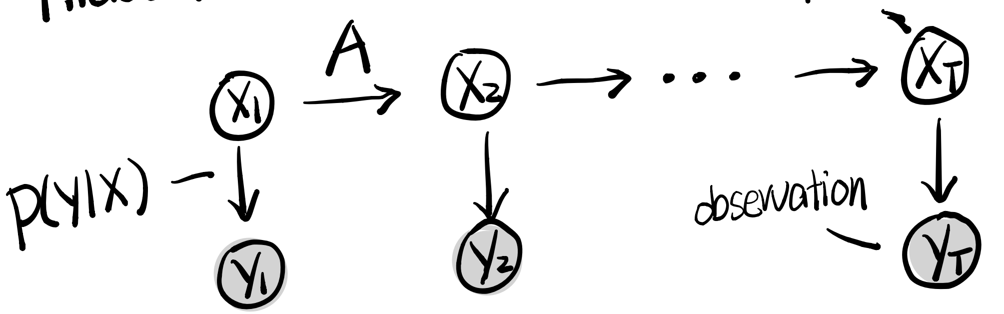
Hidden Markov Model

W2D3



parameter θ $\begin{cases} A: \text{state transition} \\ p(y|x) \text{ output probability} \end{cases}$

EM algorithm:

our goal: $\underset{\theta}{\operatorname{argmax}} \log p(y|\theta) = \int \log p(y, x|\theta)\, dx$

observed data — y; parameter — θ; latent (unobserved) data — x

- we would like to compute the MLE without the expensive marginalization over x

T2, T3

$$\log p(y|\theta) = \int \log p(y, x|\theta)\, dx$$

$$= \int \log \left[\frac{p(y, x|\theta)}{q(x)} q(x) \right] dx$$

arbitary distribution $q(x)$

Jensen's inequality

$$\geq \int \log \left[p(y, x|\theta) / q(x) \right] q(x)\, dz$$

$$= \int \log [p(y, x|\theta)]\, q(x)\, dx$$

$$- \int \log [q(x)]\, q(x)\, dx$$

①

②

$$= E_{q(x)} [\log p(y, x|\theta)] + H[q(x)]$$

entropy

①

expected likelihood under $q(x)$

$$= \int \log[\underbrace{p(x|y,\theta)}\,\underbrace{p(y|\theta)}]\, q(x)\, dx - \log[q(x)]\, q(x)\, dx$$

$$= \underbrace{\int \log[p(x|y,\theta)]\, q(x)\, dx} + \log p(y|\theta) - \underbrace{\log[q(x)]\, q(x)\, dx}$$

$$= \log p(y|\theta) - \int \log\left[\frac{p(x|y,\theta)}{q(x)}\right] q(x)\, dx$$

$$= \log p(y|\theta) - KL[p(x|y,\theta) \,\|\, q(x)] \quad ②$$

---

marginal likelihood

lower bound on the likelihood

Thus $\log p(y|\theta) \geq \underline{F[\theta, q(x)]}$

$$= ①\ E_{q(x)}[\log p(y, x|\theta] + H[q(x)]$$

$$= ②\ \log p(y|\theta) - KL[p(x|y,\theta) \,\|\, q(x)]$$

instead of max $\log p(y|\theta)$ directly, we maximize $F[\theta, q(x)]$, which is a lower bound of $\log p(y|\theta)$

EM algorithm:

repeat { E-step: $q(x) \leftarrow p(x|y,\theta)$

(increase F & making the bound tight by setting the KL term to 0) ②

M-step: $\theta \leftarrow \underset{\theta}{\operatorname{argmax}} \; E_{q(x)}[\log p(y, x|\theta)] + H[q(x)]$ ①

(maximize ①, given a fixed $q(x)$)

EM can be viewed as a coordinate ascent on $F[\theta, q(x)]$