CS 6347

Lecture 12

Maximum Likelihood Learning

Maximum Likelihood Estimation

- Given samples $x^1, \dots, x^M$ from some unknown distribution with parameters θ…
  - The log-likelihood of the evidence is defined to be

$$\log l(\theta) = \sum_m \log p(x|\theta)$$

  - Goal: maximize the log-likelihood

MLE for Bayesian Networks

- Given samples $x^1, \dots, x^M$ from some unknown Bayesian network that factors over the directed acyclic graph G
  - The parameters of a Bayesian model are simply the conditional probabilities that define the factorization
  - For each $i \in G$ we need to learn $p(x_i | x_{parents(i)})$, create a variable $\theta_{x_i | x_{parents(i)}}$

$$\log l(\theta) = \sum_m \sum_{i \in V} \log \theta_{x_i^m | x_{parents(i)}^m}$$

MLE for Bayesian Networks

$$\log l(\theta) = \sum_{m} \sum_{i \in V} \log \theta_{x_i^m | x_{parents(i)}^m}$$

$$= \sum_{i \in V} \sum_{m} \log \theta_{x_i^m | x_{parents(i)}^m}$$

$$= \sum_{i \in V} \sum_{x_{parents(i)}} \sum_{x_i} \mathrm{N}_{x_{\mathrm{i}}, x_{\mathrm{parents(i)}}} \log \theta_{x_i | x_{parents(i)}}$$

MLE for Bayesian Networks

$$\log l(\theta) = \sum_{m} \sum_{i \in V} \log \theta_{x_i^m | x_{parents(i)}^m}$$

$$= \sum_{i \in V} \sum_{m} \log \theta_{x_i^m | x_{parents(i)}^m}$$

$$= \sum_{i \in V} \sum_{x_{parents(i)}} \sum_{x_i} \mathrm{N}_{x_\mathrm{i}, x_{\mathrm{parents(i)}}} \log \theta_{x_i | x_{parents(i)}}$$

$N_{x_i, x_{parents(i)}}$ **is the number of times** $(x_i, x_{parents(i)})$ **was observed in the samples**

MLE for Bayesian Networks

$$\log l(\theta) = \sum_{m} \sum_{i \in V} \log \theta_{x_i^m | x_{parents(i)}^m}$$

$$= \sum_{i \in V} \sum_{m} \log \theta_{x_i^m | x_{parents(i)}^m}$$

$$= \sum_{i \in V} \sum_{x_{parents(i)}} \sum_{x_i} \mathrm{N}_{x_{\mathrm{i}}, x_{\mathrm{parents(i)}}} \log \theta_{x_i | x_{parents(i)}}$$

Fix $x_{parents(i)}$ solve for $\theta_{x_i | x_{parents(i)}}$ for all x_i

(on the board)

MLE for Bayesian Networks

$$\theta_{x_i|x_{parents(i)}} = \frac{\mathrm{N}_{x_i,x_{\text{parents}(i)}}}{\sum_{x_i'} \mathrm{N}_{x_i',x_{\text{parents}(i)}}} = \frac{\mathrm{N}_{x_i,x_{\text{parents}(i)}}}{\mathrm{N}_{x_{\text{parents}(i)}}}$$

- This is just the empirical conditional probability distribution
  - Worked out nicely because of the factorization of the joint distribution
- Similar to the coin flips result from last time