Agenda and Resources

Sunday, November 9, 2025 8:28 PM

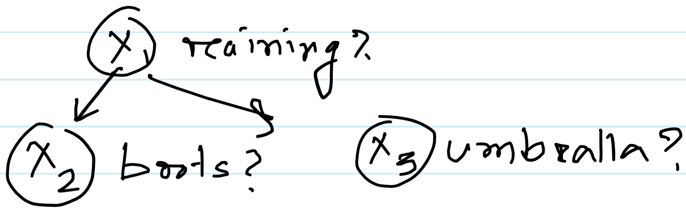
Agenda
Review
Parameter learning
Structure learning

1. Scoring
2. K2
3. Local search
4. Mar kov equivalence

Algorithms for decision making
Probabilistic Graphical Model (Kellar and Friedman)
Jack's course nores jemoka.com
Project 1 video AA228 website project 1 page

Review

Sunday, November 9, 2025 8:38 PM



Conditional independence

$(X \perp Y \mid Z) \leftrightarrow P(X, Y \mid Z) = P(X \mid Z)\, P(Y \mid Z)$

X is conditionally independence of y given z if and only if probability of x and y given z is equal to probability of x given z times probability of y given z.

Once we know z Knowing x doesn't affect y.

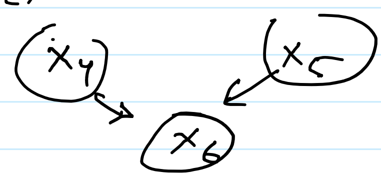
$X_{1:n}$ discrete Random variables

r_i # of instantiation of X_i

q_i # of parantal instantiation of X_i (1 if no parents)

π_{ij} j^{th} parantal instantiation of X_i

ex —

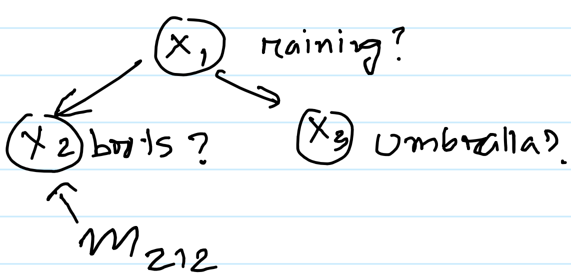


$q_4 = q_5 = 1$

$q_6 = r_4 \times r_5 = 2 \times 2 = 4$

	X_4	X_5
π_{63}	1	1
	1	2
	$\boxed{2}$	$\boxed{1}$
	2	2

Parameter learning



$\theta_{ijk} = P(X_i = k \mid \pi_{ij})$

$\sum_{i=1}^{n} r_i q_i$ total parameters

$\sum_{i=1}^{n} (r_i - 1) q_i$ independent parameters

$\theta_{322} = P(X_3 = 2 \mid X_1 = 2)$

node 3

2nd parantal instantiation

node

2nd parental instantiation

2nd node instantiation

4.1.3. Maximum Likelihood Estimation

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \; P(D|\theta, G)$$ → Prob of observed data given θ, G graph structure

m_{ijk} # times $X_i = k$ given parental instantiation j

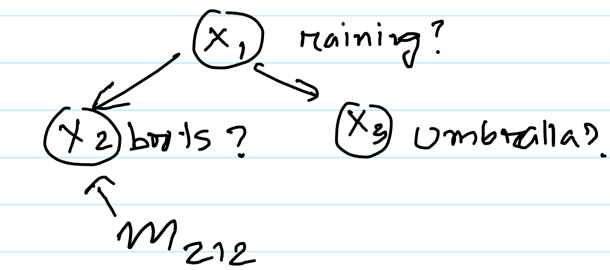
$$P(D|\theta, G) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{m_{ijk}}$$

$$\hat{\theta}_{ijk} = \frac{m_{ijk}}{\sum_{k'} m_{ijk'}}$$

Parameter Learning

Sunday, November 9, 2025 9:20 PM

Parameter learning



$$\theta_{ijk} = P(x_i = k \mid \pi_{ij})$$

$\sum_{i=1}^{n} r_i q_i$ total parameters

$\sum_{i=1}^{n} (r_i - 1) q_i$ independent parameters

$$\theta_{322} = P(x_3 = 2 \mid x_1 = 2)$$

node 3
2nd parental instantiation
2nd node instantiation

4.1.3. Maximum Likelihood Estimation

$$\hat{\theta} = \underset{\theta}{\arg\max}\, P(D \mid \theta, G)$$ → Prob of observed data given θ, G graph structure

m_{ijk} # times $X_i = k$ given parental instantiation j

$$P(D \mid \theta, G) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{m_{ijk}}$$

$$\hat{\theta}_{ijk} = \frac{m_{ijk}}{\sum_{k'} m_{ijk'}}$$

Bayesian Parameter Learning

Sunday, November 9, 2025 9:22 PM

$$\hat{\theta} \approx \underset{\theta}{\arg\max}\ P(D \mid \theta) \quad \text{MLE}$$

$$\left.\begin{aligned} &P(\theta \mid D) \\ &\hat{\theta} = \underset{\theta}{\arg\max}\ P(\theta \mid D) \end{aligned}\right\} \begin{array}{l}\text{Bayesian} \\ \text{Maximum a posteriori (MAP)}\end{array}$$

$$P(\theta \mid G) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} P(\theta_{ij})$$

all params in network

$$\theta_{ij} = (\theta_{ij1}, \theta_{ij2}, \ldots\ldots, \theta_{ijn})$$

$$P(\theta_{ij}) = \mathrm{Dir}(\theta_{ij} \mid \alpha_{ij1}, \ldots\ldots, \alpha_{ijr_i})$$

Dirichlete Distributions

Dirichlete Distributions generalizes the beta distributions to arbitrary number of categories.

$\mathrm{Dir}(\alpha_{ij1}, \ldots \alpha_{ijr_i})$ after observations $\rightarrow \mathrm{Dir}(\alpha_{ij1} + m_{ij1}, \ldots \alpha_{ijr_i} + m_{ijr_i})$

$$P(\theta_{1:n} \mid \alpha_{1:n}) = \mathrm{Dir}(\theta_{1:n} \mid \alpha_{1:n})$$

$$\propto \prod_{i=1}^{n} \theta_i^{\alpha_i - 1}$$

$$= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \prod_{i=1}^{n} \theta_i^{\alpha_i - 1}$$

* Higher the pseudo counts are more confident we are

$$P(\theta \mid G)$$

$$P(\theta_{ij} \mid \alpha_{ij}, m_{ij}) = \mathrm{Dir}(\theta_{ij} \mid \alpha_{ij1} + m_{ij1}, \ldots\ldots, \alpha_{ijr_i} + m_{ijr_i})$$

ex.

$r_3 = 2$ $\quad$ $r_4 = 2$ $\quad$ $r_5 = 2$

ex.

$$\begin{tikzcd} X_3 \arrow[dr] & & X_4 \arrow[dl] \\ & X_5 & \end{tikzcd}$$

$$q_3 = 1 \left\{ \overbrace{\begin{bmatrix} 2, & 1 \end{bmatrix}}^{r_3 = 2} \right. \qquad q_4 = 1 \left\{ \overbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}^{r_4 = 2} \right. \qquad q_5 = 4 \left\{ \overbrace{\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}}^{r_5 = 2} \right.$$

$$q_5 = r_3 \times r_4 = 4$$

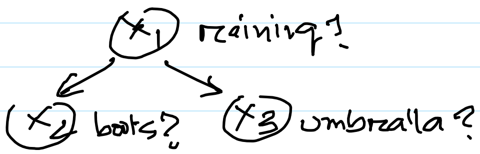
X_3	X_4	X_5
1	2	1
2	2	2

X_3	X_4
1	1
1	2
2	1
2	2

Structure Learning

Sunday, November 9, 2025 10:12 PM

5.1 Scoring



$$P(G|D) \propto P(G)\,P(D|G)$$ Bayes's rule

$$= P(G)\int P(D,\theta|G)\,d\theta$$ law of total probability

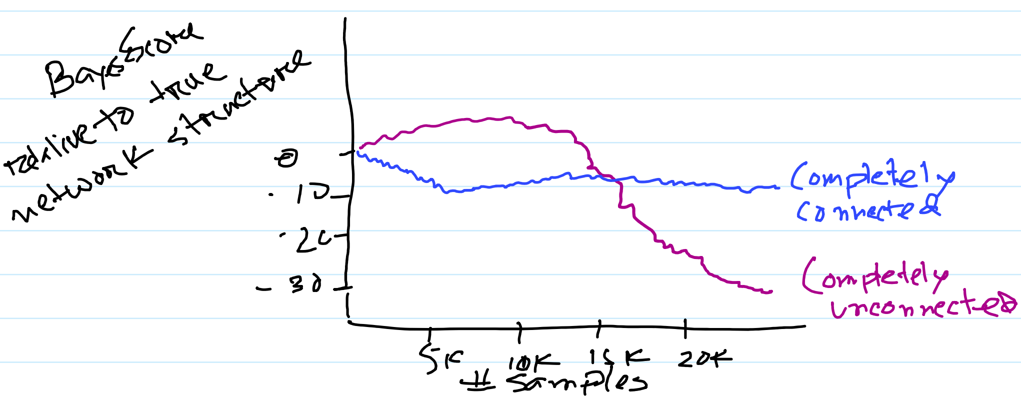
$$= P(G)\int P(D|\theta,G)\,p(\theta|G)\,d\theta$$ chain rule

$$P(G|D) = P(G)\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij0})}{\Gamma(\alpha_{ij0}+m_{ij0})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+m_{ijk})}{\Gamma(\alpha_{ijk})}$$

$$\alpha_{ij0} = \sum_{k=1}^{r_i}\alpha_{ijk} \qquad m_{ij0} = \sum_{k=1}^{r_i} m_{ijk}$$

$$\log P(G|D) = \log P(G) + \sum_{i=1}^{n}\sum_{j=1}^{q_i}\left(\log\left(\frac{\Gamma(\alpha_{ij0})}{\Gamma(\alpha_{ij0}+m_{ij0})}\right) + \sum_{k=1}^{r_i}\log\left(\frac{\Gamma(\alpha_{ijk}+m_{ijk})}{\Gamma(\alpha_{ijk})}\right)\right)$$

Fig 5.3



5.2 Structure learning.

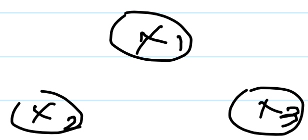
K2 Algorithm

poly time → not guaranteed to find global maxima.

K2 Algorithm

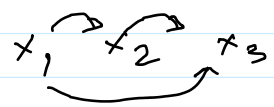
poly time → not guarantied to find global maxima.

① Begins with no edges.



② in some order, iterate over the nodes $x_1\ x_2\ x_3$

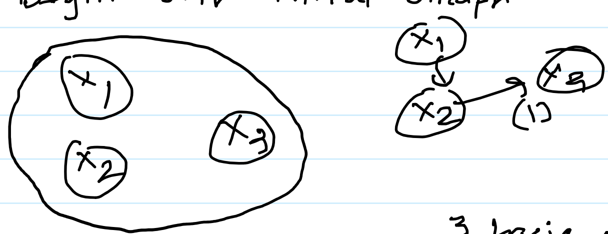
③ greedily add parents to maximally increase score
best so far



Local Search.

① begin with intial Graph

neighborhood: set of all graphs a single basic graph operation away from the current graph.



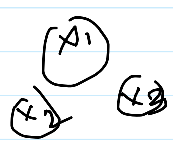
3 basic graph operation

(i) add an edge

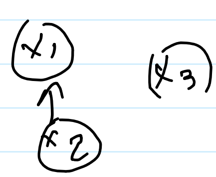
(ii) Remove all edges

(iii) Reverse an edge

II



III



② Score all neighbors and move to highest scoring graph

③ Repeat untill current graph has no higher scoring neighbors.

randomized restarts

simulated annealing

simulated annealing

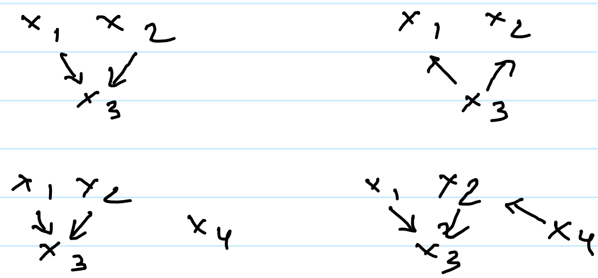
genetic algorithms

Markov Equivalence

Sunday, November 9, 2025 11:00 PM

Markov Equivalence -> encoding same conditional independence assumption

1. Same undirected edges
2. Same immoral v-structures



Can have != scores within group
Unless

$K = \sum_j \sum_k \alpha_{ijk}$ is constant for all i