https://pollev.com/haroldsohsoo986

CS5340:
Tutorial 3

Asst. Prof. Harold Soh

TA: Eugene Lim

SCAN ME

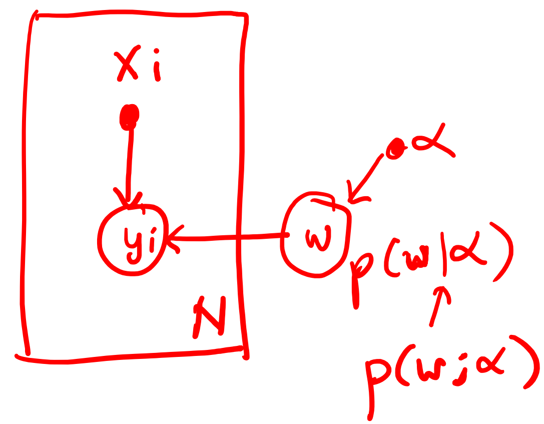
Announcements

- Group formations due: **7 Feb 2024, 11:59PM**
  - **Put your group on Canvas**
- Project Abstracts due: **4 March 2024, 11:59PM**
  - **There will be NO FURTHER EXTENSION.**
  - NeurIPS LaTeX Template
  - Up to 4 pages excl. references
    - MAX pages, don't need 4 pages
  - PDF submission on Canvas
  - Project Discussion Meetings (Calendly link)
- Quiz: **20 Feb 2024**
  - Covers everything up to Variable Elimination and Belief Propagation.

$$y = \underline{\underline{w}}^T x$$

$$\underset{w}{\text{minimize}}\ \mathcal{L}(w, D).$$

Bayesian. $p(w|D)$



Course Schedule (Tentative)

Week	Date	Lecture Topic	Tutorial
1	16 Jan	Introduction to Uncertainty Modeling + Probability Basics	~~Introduction~~
2	23 Jan	Simple Probabilistic Models	Introduction and Probability Basics
3	30 Jan	Bayesian networks (Directed graphical models)	More Basic Probability
4	6 Feb	Markov random Fields (Undirected graphical models)	DGM modelling and d-separation
5	13 Feb	Variable elimination and belief propagation	MRF + Sum/Max Product
6	20 Feb	Factor graphs	**Quiz 1**
-	-	**RECESS WEEK**	
7	5 Mar	Mixture Models and Expectation Maximization (EM)	Linear Gaussian Models
8	12 Mar	Hidden Markov Models (HMM)	Probabilistic PCA
9	19 Mar	Monte-Carlo Inference (Sampling)	Linear Gaussian Dynamical Systems
10	26 Mar	Variational Inference	MCMC + Langevin Dynamics
11	2 Apr	Inference and Decision-Making	Diffusion Models + Sequential VAEs
12	9 Apr	Gaussian Processes (optional)	**Quiz 2**
13	16 Apr	Project Presentations	Closing Lecture

CS5340:
Tutorial 3

Asst. Prof. Harold Soh

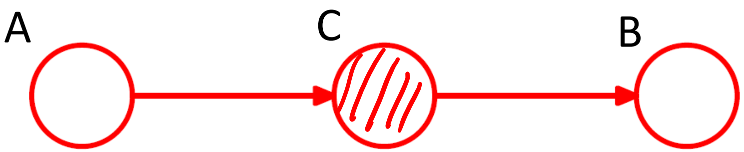
TA: Eugene Lim

Bayes Nets (BN)

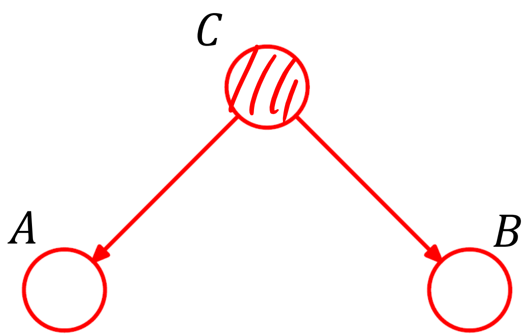
- **Definition** (*Bayesian Network*) A Bayesian network is a tuple $B = (G, P)$ where P factorizes according to G and where P is specified as a set of conditional probability distributions (CPDs) associated with G's nodes.

$$\boxed{p(x_1, \dots, x_N) = \prod_{i=1}^{N} p(x_i | x_{\pi_i})}$$

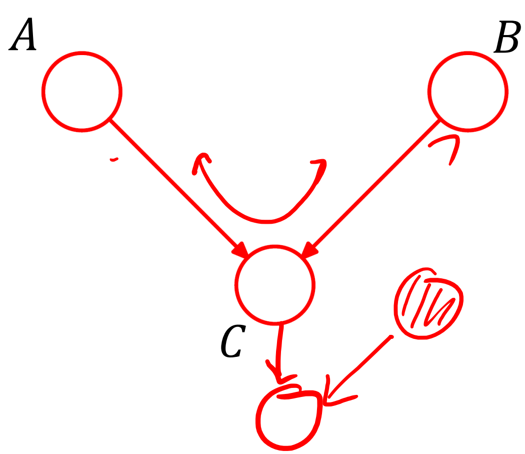
The Canonical 3-node graphs



Head-Tail (wrt C)
(Chain/Causal-trail)



Tail-Tail (wrt C)
(Tent/Common cause)



Head-Head (wrt C)
(V-structure/Collider/Common Effect)

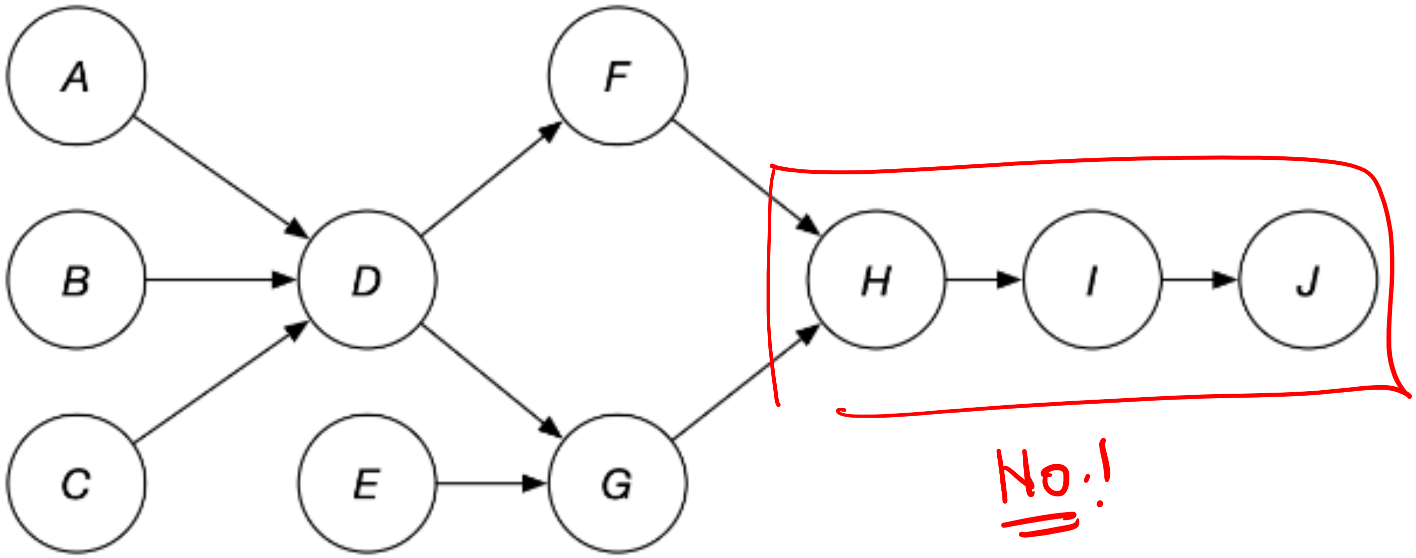
Graph Separation

- $A \perp B \mid C$ if all trails from nodes in set A are "blocked" from nodes in set B when all nodes from set C are observed.

- *A* is said to be d-separated from *B* by *C*, and the joint distribution over all of the variables in the graph will satisfy $A \perp B \mid C$.

1. D-separation test

$(H \perp J \mid \phi)$?



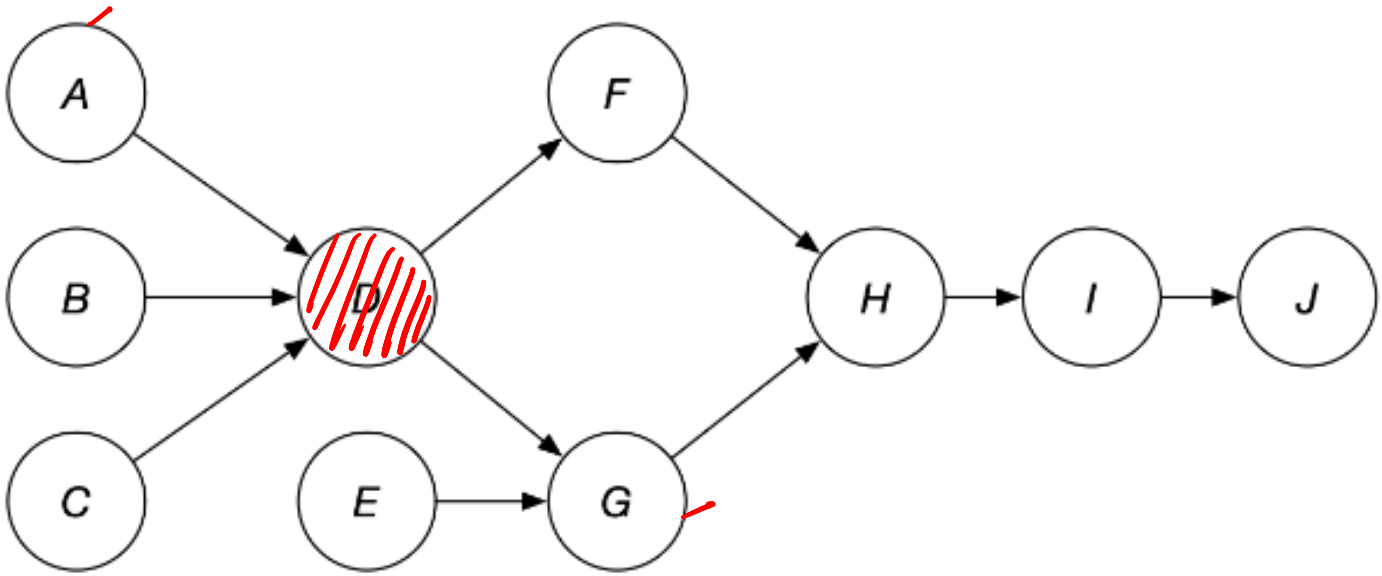
No!

$(H \perp J \mid I)$? True. ✓

1. D-separation test

$(A \perp G \mid D)$?

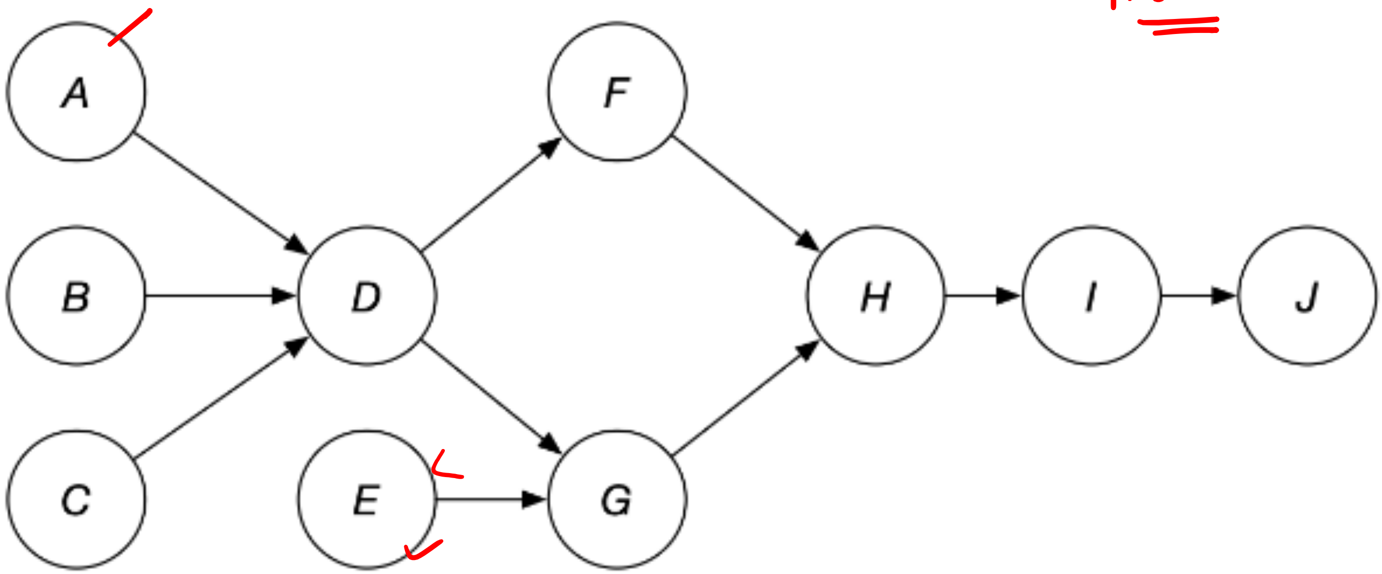
True. ✓



1. D-separation test

$(A \perp E \mid \phi)$?

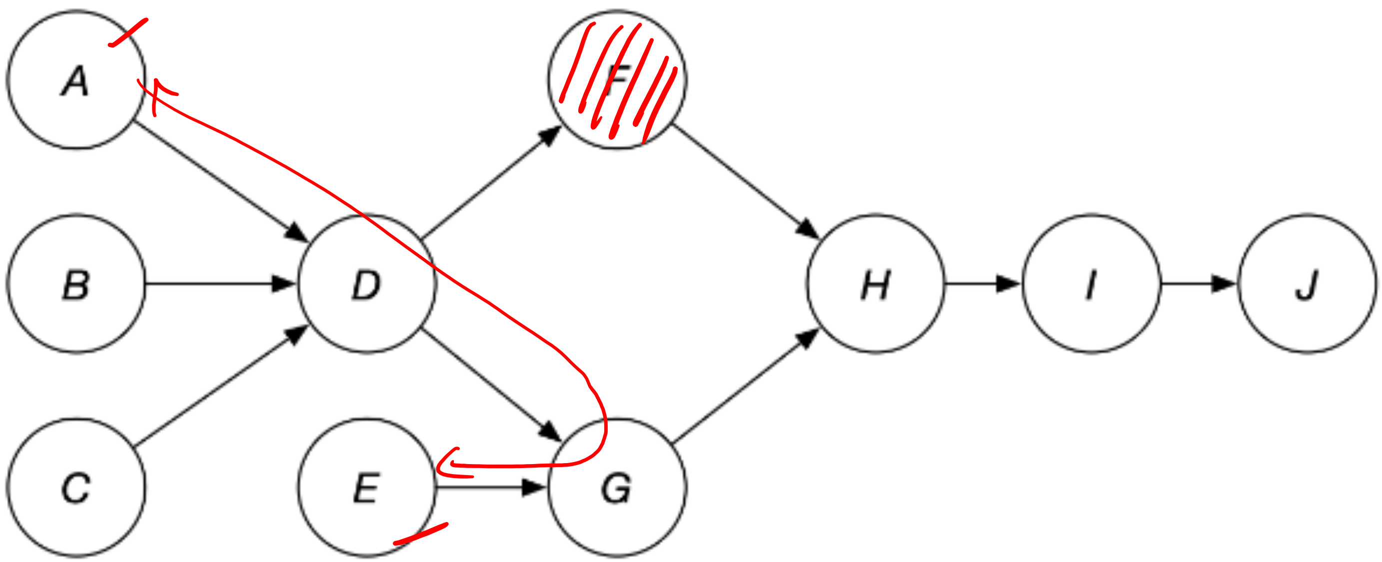
True



1. D-separation test

$(A \perp E \mid F, J)$?

False.



Questions?

https://pollev.com/haroldsohsoo986

SCAN ME

2. Healthy Students

- HealthyStudents will build an AI system that will monitor students to predict the occurrence of mental health issues.
- Inform the university of such occurrences so that interventions can be taken and support be given to at-risk students.
- Access available sensors (cameras on campus, student grades, and participation in extra-curricular activities).

Disclaimer: This is a fictional name and scenario. Any similarity to an actual company is purely coincidental.

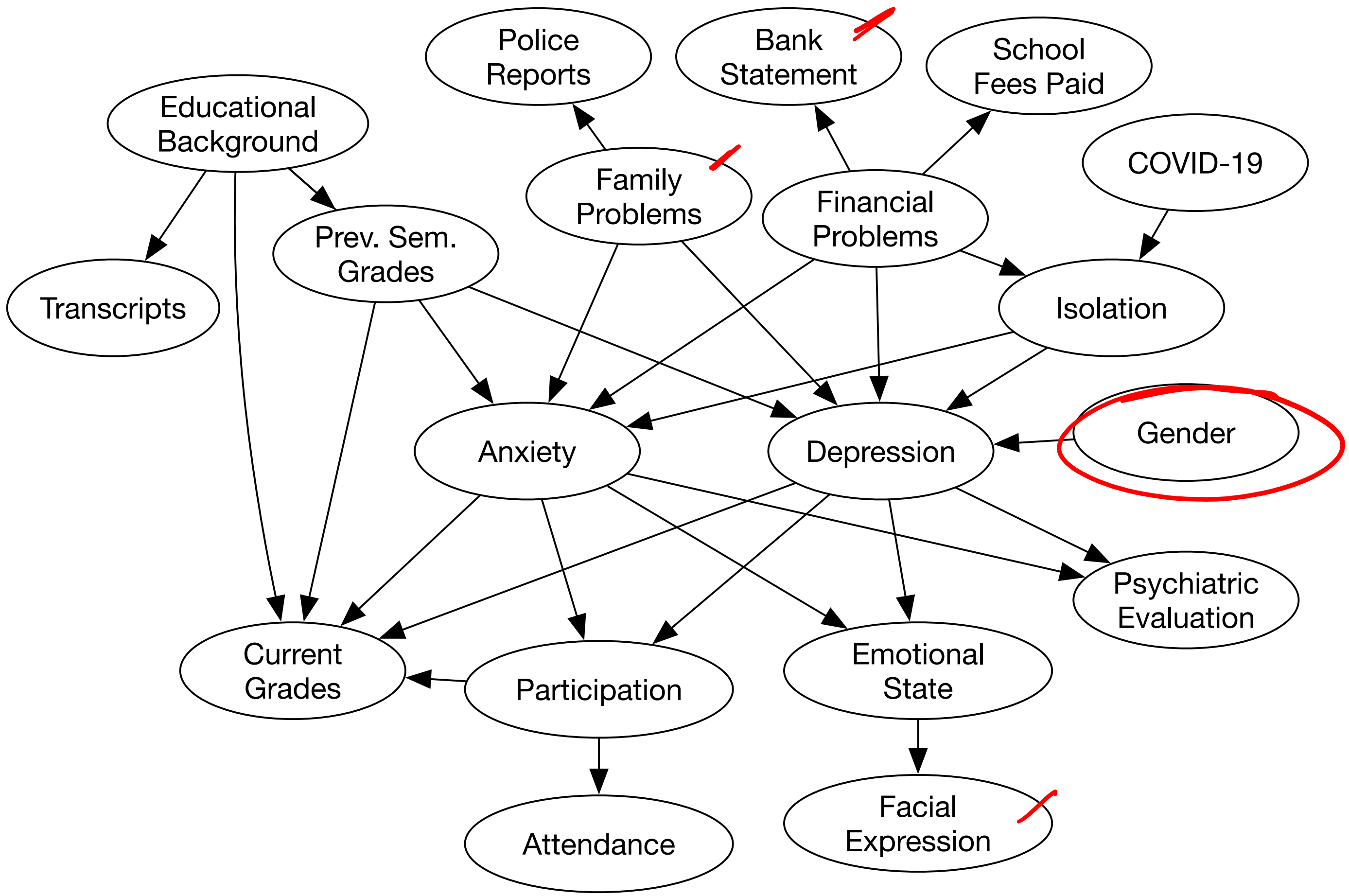
2. Healthy Students

- Design a Bayesian Network for this problem.
- Is it Ethical to build such a system?

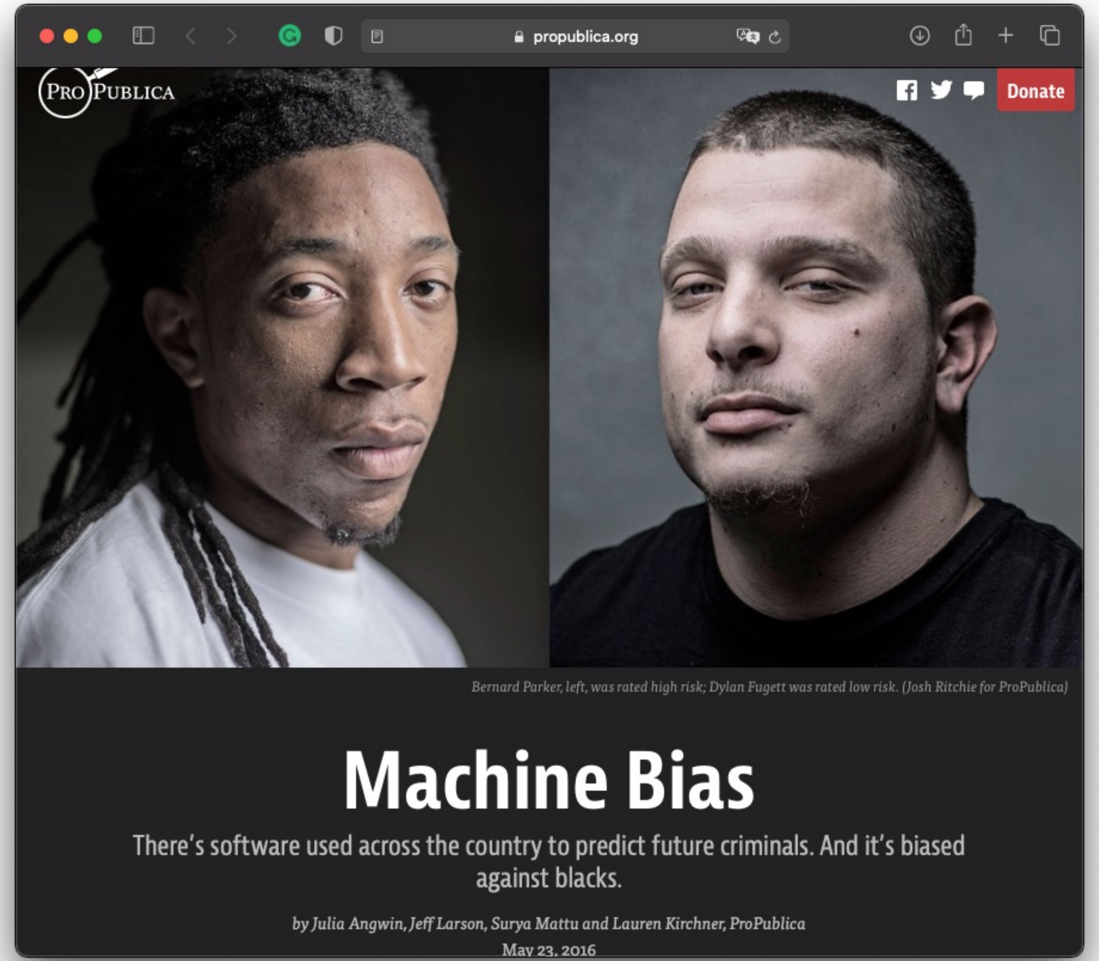
SCAN ME

https://pollev.com/haroldsohsoo986

2. Healthy Students



2. Healthy Students



2. Healthy Students

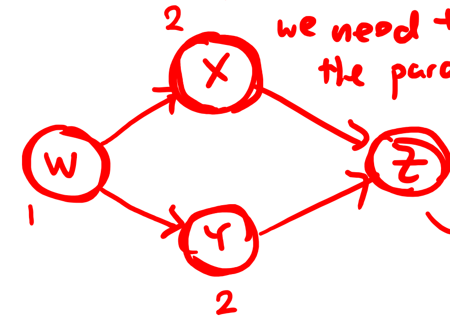
2. Healthy Students

1. Who are all the people affected by the action?
2. What option benefits me the most?
3. What option does my social group support?
4. What option is legal?
5. What option is the greatest good for the greatest number of people?
6. What option is based on truthfulness and respect/integrity towards each stakeholder?
7. What option would a virtuous person of high moral character do?

https://us.sagepub.com/sites/default/files/upm-assets/90084_book_item_90084.pdf

3. Your CS5340 Grade

For each value of W
we need to learn
the params of
X separately



4×3=12

Model how well students perform.

final grades (Z): either A, B, C, and D.
0 1 2 3

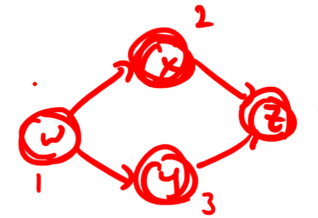
Only two components affect a student's final grade: the student's project (X) and the ~~final exam~~ quiz (Y).

X and Y have two possible outcomes each: Pass (1) or Fail (0)

Assume that whether a student does well for the project and final ~~exam~~ quiz depends only on how hard they work (W, which is binary).

Draw a Bayesian Network that models the scenario above.

$$p_\theta(w,x,y,z) = \underline{p_{\theta_w}(w)\, p_{\theta_x}(x|w)\, p_{\theta_y}(y|w)\, p_{\theta_z}(z|x,y)}$$



$$\theta = \{\theta_w, \theta_x, \theta_y, \theta_z\}$$
1 2 2 12

$$\log p_\theta(w,x,y,z)$$

$$= \log p_{\theta_w}(w) + \log p_{\theta_x}(x|w) + \log p_{\theta_y}(y|w) + \log p_{\theta_z}(z|x,y)$$

$$\log p_{\theta_w}(R_1) + \log p_{\theta_x}(R_2|R_1) \qquad R_3|R_1 \qquad R_4|R_2,R_3$$

Denote:
- N : # observations $(N=8)$
- U : # random vars. $(U=4)$
- $R_{u,n}$: the u^{th} r.v. (n^{th} sample)
- $R_{\pi_{u,n}}$: the parent of u^{th} r.v.
- D : dataset

$$\log \underline{p(D|\theta)} = \log \prod_{n=1}^{N} \prod_{u=1}^{U} p(R_{u,n} | R_{\pi_{u,n}}; \theta_u)$$

$$= \sum_{n=1}^{N} \sum_{u=1}^{U} \underbrace{\log p(R_{u,n} | R_{\pi_{u,n}}; \theta_u)}$$

$$\theta_u = \{\lambda_{u11}, \dots, \lambda_{uck}, \lambda_{uCK}\}$$

value of the node

value of the parent

$$\sum_u \boxed{\sum_n \log p \cdots}$$

$$p(R_{u,n} = \underline{2} \mid R_{\pi_{u,n}} = \underline{(0,1)})$$

z_n $\qquad$ x_n, y_n $\qquad$ $\lambda_{u,(0,1),2}$

Fix u. Optimize:

$\Theta_u = \{\lambda_{u11}, \ldots \lambda_{uck}, \ldots, \lambda_{uCK}\}$

$$\sum_{n=1}^{N} \log \underbrace{p(R_{u,n} \mid R_{\pi_{u,n}} ; \Theta_u)}$$

$$= \sum_{n=1}^{N} \sum_{c=1}^{C} \sum_{k=1}^{K} \overbrace{\log p(R_{u,n}=k \mid R_{\pi_{u,n}}=c ; \Theta_u)}^{I_{uck,n}}$$

$$= \sum_{c} \sum_{k} \sum_{n} I_{uck,n} \underbrace{\log p(R_{u,n}=k \mid R_{\pi_{u,n}}=c ; \Theta_u)}$$

$$= \sum_{c} \sum_{k} \sum_{n} I_{uck,n} \underbrace{\log \lambda_{uck}}$$

$$= \sum_{c} \sum_{k} \log \lambda_{uck} \underbrace{\sum_{n} I_{uck,n}}_{:= N_{uck}}$$

$$= \sum_{c} \sum_{k} N_{uck} \log \lambda_{uck} \quad \text{s.t.} \ \boxed{\sum_{k} \lambda_{uck} = 1}$$

$\mathbb{1}[\text{Statement}] = \begin{cases} 0 & \text{if FALSE} \\ 1 & \text{if TRUE} \end{cases}$

$X \in \{0,1\}$

$p(X) = p(X=0)^{\overset{=0}{\mathbb{1}[X=0]}} \, p(X=1)^{\overset{=1}{\mathbb{1}[X=1]}}$

Assume $X=0$

$$\Downarrow$$

$p(X=0)^1 \, \cancel{p(X=1)^0}$

$I_{uck,n} = \mathbb{1}[R_{u,n}=k, R_{\pi_{u,n}}=c]$

$$L = \sum_{k} N_{uck} \log \lambda_{uck} + \nu \left(\sum_{k} \lambda_{uck} - 1 \right)$$

$$\frac{d}{d\lambda_{uck}} L = \frac{N_{uck}}{\hat{\lambda}_{uck}} + \hat{\nu} \quad := 0$$

$$\Downarrow$$

$$\hat{\lambda}_{uck} = \frac{-N_{uck}}{\hat{\nu}}$$

$$\frac{dL}{d\nu} = \sum_{k} \hat{\lambda}_{uck} - 1 := 0$$

$$\Rightarrow \underbrace{\sum_{k} \hat{\lambda}_{uck}}_{=}\overset{\text{``}1}{} = \frac{1}{\hat{\nu}} \sum_{k} -N_{uck}$$

$$\hat{\nu} = \frac{-1}{\sum_{m=1}^{K} N_{ucm}}$$

$$\hat{\lambda}_{uck} = \frac{N_{uck}}{\sum_{m=1}^{K} N_{ucm}}$$

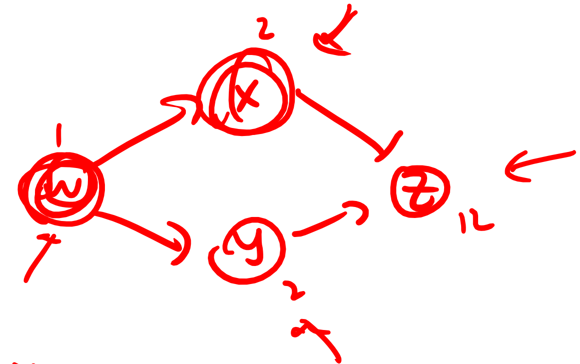
W	p(W)
0	1/2
1	1/2

1

$$\lambda_{W\phi 0} = \frac{N_{W\phi 0}}{N_{W\phi 0} + N_{W\phi 1}}$$

4

4 4



X	Y	Z	
0	0	0	
0	0	1	
0	0	2	
0	0	3	
0	1	0	
0	1	1	
0	1	2	1
0	1	3	
1	0		
	.		
	1		
	1		
	1		

3

3

3

3

W	X	
0	0	✓
0	1	✓
1	0	
1	1	

1

1

Questions?

https://pollev.com/haroldsohsoo986

SCAN ME

4. Label Errors (Code on this link)

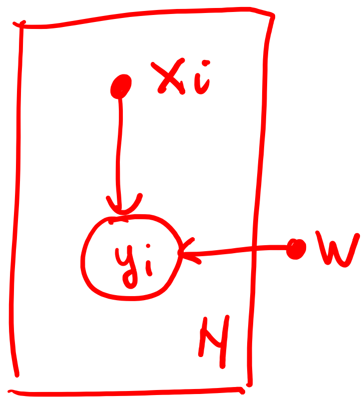
4. Label Errors

$$p(y|x) = \text{Bern}[p]$$

$$p = \sigma(w^T x)$$

$$= \frac{1}{1+\exp(-w^T x)}$$

4.a. Bayesian Network



$$p(y_i | x_i, w) = p_i^{y_i} (1-p_i)^{1-y_i}$$

4. Label Errors

4.b. MLE for Logistic Regression

$$\arg\max_{w} \log p(D \mid w)$$

$$\log \prod_{i}^{N} p(y_i \mid x_i, w) \quad \} \text{ (iid)}$$

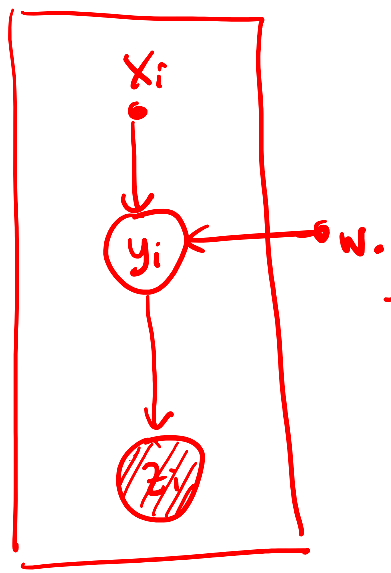
$$\Rightarrow \sum_{i}^{N} \lg \underbrace{p(y_i \mid x_i, w)}.$$

$$= \sum_{i}^{N} \lg p_i^{y_i} (1 - p_i)^{1-y_i}$$

$$= \sum_{i}^{N} \underbrace{y_i \lg p_i + (1 - y_i) \lg (1 - p_i)}_{\text{cross-entropy}}$$

4. Label Errors

4.c. Bayesian Network with Label Errors



$p(y_i \mid x_i, w)$

z	y	$p(z \mid y)$
0	0	0.75
0	1	0.05
1	0	0.25
1	1	0.95

4. Label Errors

4.d. New MLE that accounts for label errors.

$$\hat{w} = \arg\max_{w} \lg \prod_{i=1}^{N} \underbrace{p(z_i \mid x_i, w)}.$$

$$= \sum_{y_i} p(z_i, y_i \mid x_i, w).$$

$$= \sum_{y_i} p(z_i \mid y_i)\, p(y_i \mid x_i, w).$$

$$= \sum_{i=1}^{N} \lg\left(\sum_{y_i} p(z_i \mid y_i)\, p(y_i \mid x_i, w)\right).$$

$$= \sum_{i}^{N} z_i \lg\left[p(z_i=1 \mid y_i=1)\, p_i + p(z_i=1 \mid y_i=0)\,(1-p_i)\right] + (1-z_i) \lg\left[p(z_i=0 \mid y_i=1)\, p_i + p(z_i=0 \mid y_i=0)\,(1-p_i)\right].$$

Homework!

- Watch the videos:
- Do the tutorial
    - In addition to the sheet, there is a notebook on Image Denoising

Video Lectures

- L4 - Part 1 (Intro and Recap)
- L4 - Part 2 (MRF Intuition)
- L4 - Part 3 (Markov Properties)
- L4 - Part 4 (MRF Parameterization)
- L4 - Part 5 (MRF Examples)
- L4 - Part 6 (MRF Theory)
- L4 - Part 7 (MRF Parameter Learning)

Tutorial (for next week)

- Tut4_24.pdf
- T4-Image-Denoising.zip