COMP90051

Workshop Week 11

About the Workshops

- 7 sessions in total
  - Tue 12:00-13:00 AH211
  - Tue 12:00-13:00 AH108 *
  - Tue 13:00-14:00 AH210
  - Tue 16:15-17:15 AH109
  - Tue 17:15-18:15 AH236 *
  - Tue 18:15-19:15 AH236 *
  - Fri 14:15-15:15 AH211

About the Workshops

- Homepage
  - https://trevorcohn.github.io/comp90051-2017/workshops

- Solutions will be released on next Friday (a week later).

Syllabus

1	Introduction; Probability theory	Probabilistic models; Parameter fitting	
2	Linear regression; Intro to regularization	Logistic regression; Basis expansion	
3	Optimization; Regularization	Perceptron	
4	Backpropagation	CNNs; Auto-encoders	
5	Hard-margin SVMs	Soft-margin SVMs	
6	Kernel methods	Ensemble Learning	
7	Clustering	EM algorithm	
8	Principal component analysis; Multidimensional Scaling	Manifold Learning; Spectral clustering	
9	Bayesian inference (uncertainty, updating)	Bayesian inference (conjugate priors)	
10	PGMs, fundamentals	PGMs, independence	←
11	Guest lecture (TBC)	PGMs, inference	
12	PGMs, statistical inference	Subject review	

Outline

- Review the lecture, background knowledge, etc.
  - Formal notations for probability
  - Joint probability for probabilistic graphical models (PGMs)
    - Directed PGMs
    - Undirected PGMs
  - Independence in PGMs
    - Directed PGMs
    - Undirected PGMs
  - Worksheet

Formal notations for probability

- Upper-case letters for random variables (r.v.'s)
- A, B, C
- Lower-case letters for specific values
- a, b, c
- P for the operator to calculate the probability
- $P(A = a)$, $P(A = a|B = b)$, $P(A < 10)$, $P(A = 1|B \geq 5)$

Formal notations for probability

- Suppose A is a binary random variable

- $P(A = 0)$ is a number
- $P(A = 1)$ is a number

- $P(A)$ is the distribution for the random variable A
- $P(A)$ represents $P(A = 0)$ and $P(A = 1)$

$P(A, B)$ and $P(A = a, B = b)$

- Suppose A and B are both binary random variables

- $P(A, B) = P(A)P(B)$ is equivalent to …
- $P(A = a, B = b) = P(A = a)P(B = b) \quad \forall a, b \in \{0, 1\}$

- $P(A = 0, B = 0) = P(A = 0)P(B = 0)$
- $P(A = 1, B = 0) = P(A = 1)P(B = 0)$
- $P(A = 0, B = 1) = P(A = 0)P(B = 1)$
- $P(A = 1, B = 1) = P(A = 1)P(B = 1)$

Sometimes we write partially $P(A=1,B)$

- Given the probability table:
- $P(A=1,B)$ can be
- interpreted as an unnormalized distribution for B

$P(A,B)$	$B=0$	$B=1$
$A=0$	0.1	0.2
$A=1$	0.3	0.4

$P(A = 1, B)$ is unnormalized $P(B|A = 1)$

- Given the probability table:

$P(A, B)$	$B = 0$	$B = 1$
$A = 0$	0.1	0.2
$A = 1$	0.3	0.4

- $P(A = 1, B)$ can be
- interpreted as an unnormalized distribution for B $(A = 1)$
- $P(B|A = 1) \propto P(A = 1, B)$
- $P(A = 1, B = 0) = 0.3 \quad \rightarrow \quad P(B = 0|A = 1) = 3/7$
- $P(A = 1, B = 1) = 0.4 \quad \rightarrow \quad P(B = 1|A = 1) = 4/7$

$P(A)$

- Given the probability table:

$P(A,B)$	$B=0$	$B=1$
$A=0$	0.1	0.2
$A=1$	0.3	0.4

- $P(A)$ is
- the marginal distribution for A
- $P(A)=\sum_B P(A,B)$
- $$P(A=0)=\sum_B P(A=0,B)=\sum_{b\in\{0,1\}} P(A=0,B=b)$$
  $$=P(A=0,B=0)+P(A=0,B=1)=0.3$$
- $P(A=1)=1-P(A=0)=0.7$

What about $P(A|B)$?

- Given the probability table:

$P(A,B)$	$B = 0$	$B = 1$
$A = 0$	0.1	0.2
$A = 1$	0.3	0.4

- $P(A|B)$ can be
- interpreted as two distributions for A given $B = 0$ and 1
- $P(A|B = 0)$ is a distribution
- $P(A = 0|B = 0) = 1/4$, $P(A = 1|B = 0) = 3/4$
- $P(A|B = 1)$ is another one
- $P(A = 0|B = 1) = 1/3$, $P(A = 1|B = 1) = 2/3$

And $P(A = 1|B) \ldots ?$

- Given the probability table:

$P(A,B)$	$B = 0$	$B = 1$
$A = 0$	0.1	0.2
$A = 1$	0.3	0.4

- $P(A = 1|B)$ is
- the likelihood of observing $A = 1$ under different values of B

- We have calculated that
- $P(A = 0|B = 0) = 1/4$, $P(A = 1|B = 0) = 3/4$
- $P(A = 0|B = 1) = 1/3$, $P(A = 1|B = 1) = 2/3$

If we know the joint distribution $P(A, B)$

- We can calculate everything, such as
- Marginal distributions
  - $P(A)$ and $P(B)$
  - Summation or integration over other random variables
  - $P(A) = \sum_B P(A, B) = \sum_{b \in \{0,1\}} P(A, B = b)$
- Conditional distributions
  - $P(A|B)$ and $P(B|A)$
  - Division of two unconditional distributions
  - $P(A|B) = P(A, B)/P(B)$

In general, if we know $P(A, B, C, D, E)$

- $P(A) = \sum_{B,C,D,E} P(A, B, C, D, E)$
- $P(A, B) = \sum_{C,D,E} P(A, B, C, D, E)$
- $P(A, B, C) = \sum_{D,E} P(A, B, C, D, E)$
- $P(A, B, C, D) = \sum_{E} P(A, B, C, D, E)$

- $P(B|A) = P(A, B)/P(A)$
- $P(B, C|A) = P(A, B, C)/P(A)$
- $P(C, D|A, B) = P(A, B, C, D)/P(A, B)$

Outline

- Review the lecture, background knowledge, etc.
  - Formal notations for probability
  - Joint probability for probabilistic graphical models (PGMs)
    - Directed PGMs
    - Undirected PGMs
  - Independence in PGMs
    - Directed PGMs
    - Undirected PGMs
  - Worksheet

How to calculate the joint distribution?

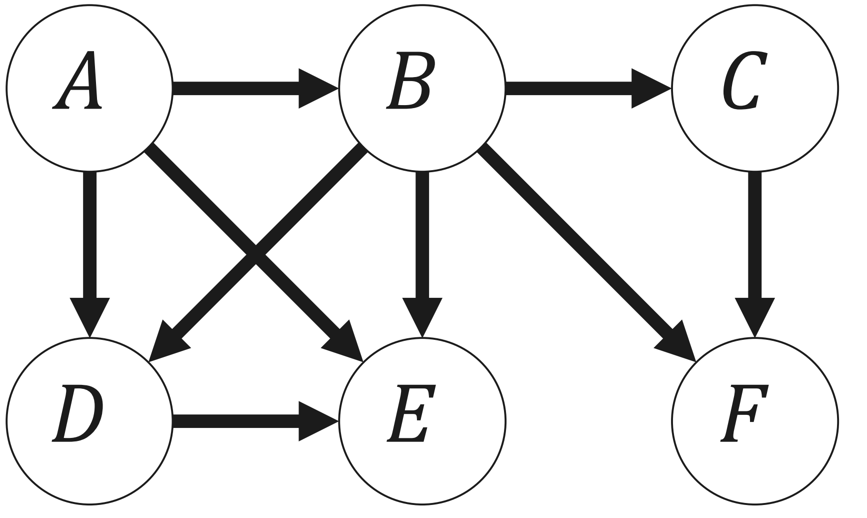
- Directed PGMs

$$P(\text{all } r.v.) = \prod_{\text{every } r.v.} P(r.v.|\text{parents of } r.v.)$$

- Undirected PGMs

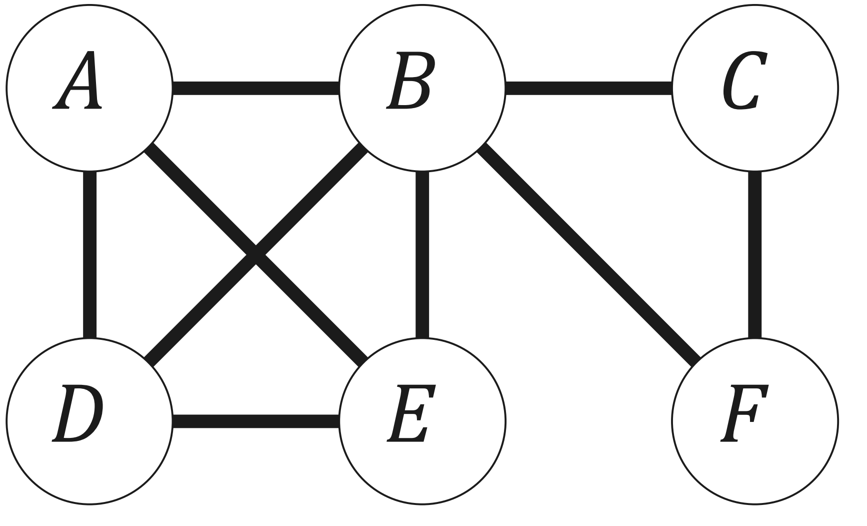
$$P(\text{all } r.v.) \propto \prod_{\text{every } clique} f_{clique}(r.v. \text{ in } clique)$$

$$
\begin{aligned}
&P(A,B,C,D,E,F) \\
&= P(A)P(B|A)P(C|B) \\
&\cdot P(D|A,B)P(E|A,B,D)P(F|B,C)
\end{aligned}
$$



$$
\begin{aligned}
&P(A,B,C,D,E,F) \\
&\propto f_1(A,B,D,E)f_2(B,C,F) \\
&= \frac{1}{Z} f_1(A,B,D,E)f_2(B,C,F)
\end{aligned}
$$

$$Z = \sum_{A,B,C,D,E,F} f_1(A,B,D,E)f_2(B,C,F)$$

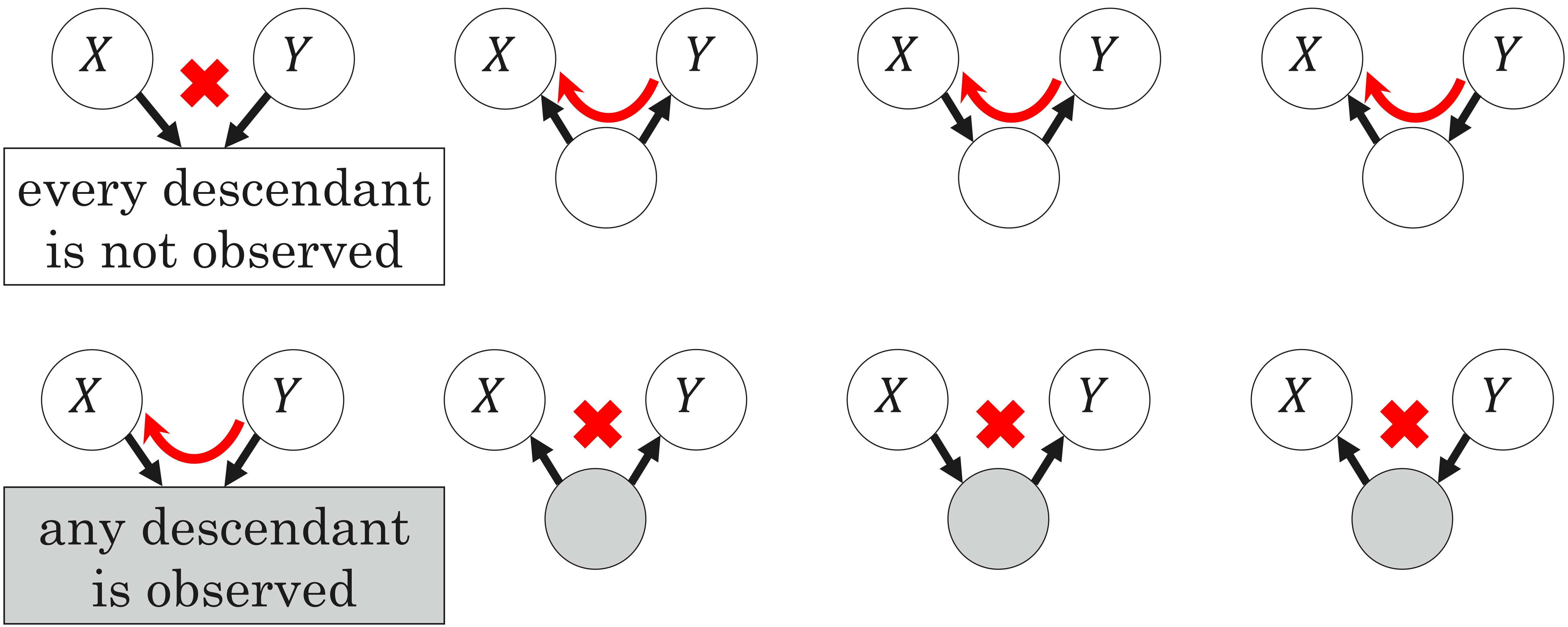


Outline

- Review the lecture, background knowledge, etc.
  - Formal notations for probability
  - Joint probability for probabilistic graphical models (PGMs)
    - Directed PGMs
    - Undirected PGMs
  - Independence in PGMs
    - Directed PGMs
    - Undirected PGMs
  - Worksheet

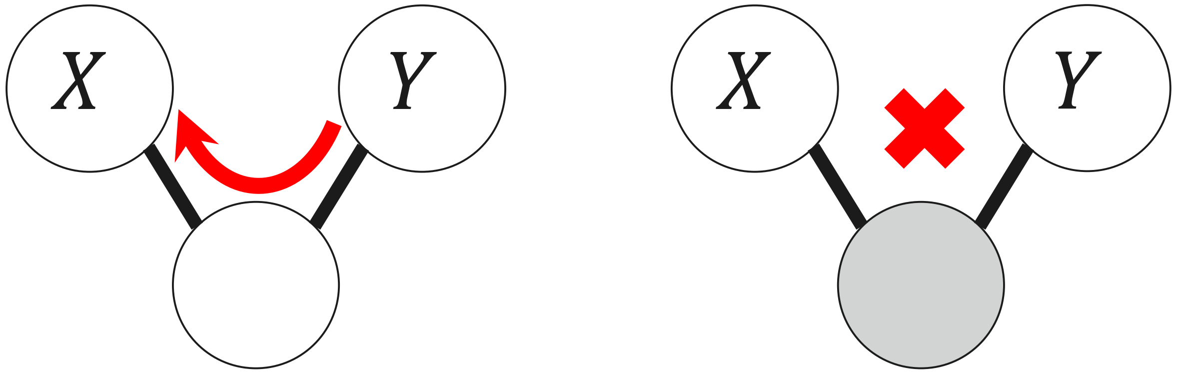
Independence in directed PGMs

- Paths from Y to X
- If a path exists, then X and Y are dependent*
  - * the PGM doesn't require they should be independent



Independence in undirected PGMs

- Paths from Y to X
- If the path exists, then X and Y are dependent



A practice for independence in PGMs

- http://web.mit.edu/jmn/www/6.034/d-separation.pdf

Outline

- Review the lecture, background knowledge, etc.
  - Formal notations for probability
  - Joint probability for probabilistic graphical models (PGMs)
    - Directed PGMs
    - Undirected PGMs
  - Independence in PGMs
    - Directed PGMs
    - Undirected PGMs
  - Worksheet