Learning & Inference in Graphical Models
Monday, March 28, 2011

Instructors: Raquel Urtasun (rurtasun@ttic.edu)
Tamir Hazan (tamir@ttic.edu)

- Monday, Wednesday, Friday 1:30-2:20

http://ttic.uchicago.edu/~rurtasun/courses/GraphicalModels/graphical_models.html

Course Outline

- Book: Probabilistic Graphical Models, Daphne Koller, Nir Friedman (in library)

- Part I: Models

  Chapter 2: Basic Notions.

  Chapter 3: Bayesian Networks.

  Chapter 4: Undirected Graphical Models.

Course Outline

- Part II: Inference

  Chapters 9&10&11: Exact Inference

  Chapter 12: Sampling methods for Inference.

  Chapter 13: MAP inference

Course Outline

- Part III: Learning

  Chapter 17: Parameters Estimation

  Chapter 18: Learning Structure

  Chapter 19: Partially Observed Data

- Causality: Chapter 21 (if time permits)

Course load

- Homework: 50% of the grade.
  6-7 exercises.
  2 programming exercises.

- Exam: 50% of the grade.
- No mid-term exam.

Background - Probability

- The confidence that an event will occur

"there is a 30% chance of rain"

"Tossing coin, there is a 50% probability for 'head' "

- Probability Space:

1) What are the possible **events**?

2) How we **measure** each event?

Probability

- What are the possible outcomes ?

Coin toss: $\Omega = \{\text{“head”}, \text{“tail”}\}$

Die: $\Omega = \{1,2,3,4,5,6\}$

- **Event** is subset of outcomes $S \subset \Omega$:

Examples for die: {1,2,3}, {2,4,6}, ...

- How we **measure** each event?

Probability function.

Probability Function

- Assign non-negative weight for atomic events
- Probability of event $S \subset \Omega$

$$P(S) = \sum_{\omega \in S} P(\omega)$$

Examples for die: P({2,4,6}) = P(2)+P(4)+P(6)

- Claim: $P(S_1 \cup S_2) = P(S_1) + P(S_2) - P(S_1 \cap S_2)$

Probability Function

- Overall weight is one $\sum_{\omega \in \Omega} P(\omega) = 1$

Coin: P("head") + P("tail")=1

Die: P(1)+P(2)+P(3)+P(4)+P(5)+P(6)=1

Conditional Probability

- S1,S2 are independent if

$$P(S_1 \cap S_2) = P(S_1)P(S_2)$$

- Conditional Probability: $S \subset \Omega$

$$P(S_1|S) = P(S_1 \cap S)/P(S)$$

- Claims: $\sum_{\omega \in S} P(\omega|S) = 1$

If S1,S are independent then $P(S_1|S) = P(S_1)$

Conditional Probability

- Claim (Chain Rule):

$$P(S_1 \cap S_2 \cap \cdots \cap S_n) = P(S_1)P(S_2|S_1) \cdots P(S_n|S_1, ..., S_{n-1})$$

Joint distribution

- Given two spaces: Ω_1, Ω_2 (e.g. coin, die, two dice)
- Joint probability function

$$P(\omega_1, \omega_2) \geq 0, \qquad \sum_{\omega_1 \in \Omega_1, \omega_2 \in \Omega_2} P(\omega_1, \omega_2) = 1$$

- Induces marginal probability functions

$$P(\omega_1) = \sum_{\omega_2 \in \Omega_2} P(\omega_1, \omega_2)$$

Random Variable

- A random variable is a function, which maps events or outcomes (e.g., the possible results of rolling two dice: (1, 1), (1, 2), etc.) to real numbers (e.g. their sum)

- A discrete random variable have a discrete set of values

$$X(\omega) \in \{r_1, ..., r_n\}$$

- A discrete random with n value induces a probability space with n elements.

$$P(r) = P(X = r) = P(\{\omega : X(\omega) = r\})$$

Joint Distribution

- Two random variables induce a joint distribution

$$P(r_1, r_2) = P(X_1 = r_1, X_2 = r_2) = P(X_1 = r_1 \text{ and } X_2 = r_2)$$

- Joint distribution induces a marginal distribution

$$P(X_1 = r_1) = \sum_{r_2} P(X_1 = r_1, X_2 = r_2)$$

- Two random variables are independent $X_1 \perp X_2$ if

$$P(X_1 = r_1, X_2 = r_2) = P(X_1 = r_1)P(X_2 = r_2)$$

Conditional Distribution

- Conditional distribution:

$$P(X_1|X_2 = r_2) = P(X_1, X_2 = r_2)/P(X_2 = r_2)$$

- Claim: If two random variables are independent then

$$P(X_1|X_2) = P(X_1)$$

- Three random variables are conditionally independent $X_1 \perp X_2 | X_3$ if

$$P(X_1 = r_1, X_2 = r_2 | X_3 = r_3) = P(X_1 = r_1 | X_3 = r_3) P(X_2 = r_2 | X_3 = r_3)$$

Expectation, Variance

- Expectation $$\mathbb{E}[X] = \sum_r P(X = r) \cdot r$$

- Variance

$$V[X] = \sum_r P(X = r)(r - \mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Continuous Random Var

- A continuous random variable have has a density function f(r)

$$P(X(\omega) \in [r_1, r_2]) = \int_{r_1}^{r_2} f(r)$$

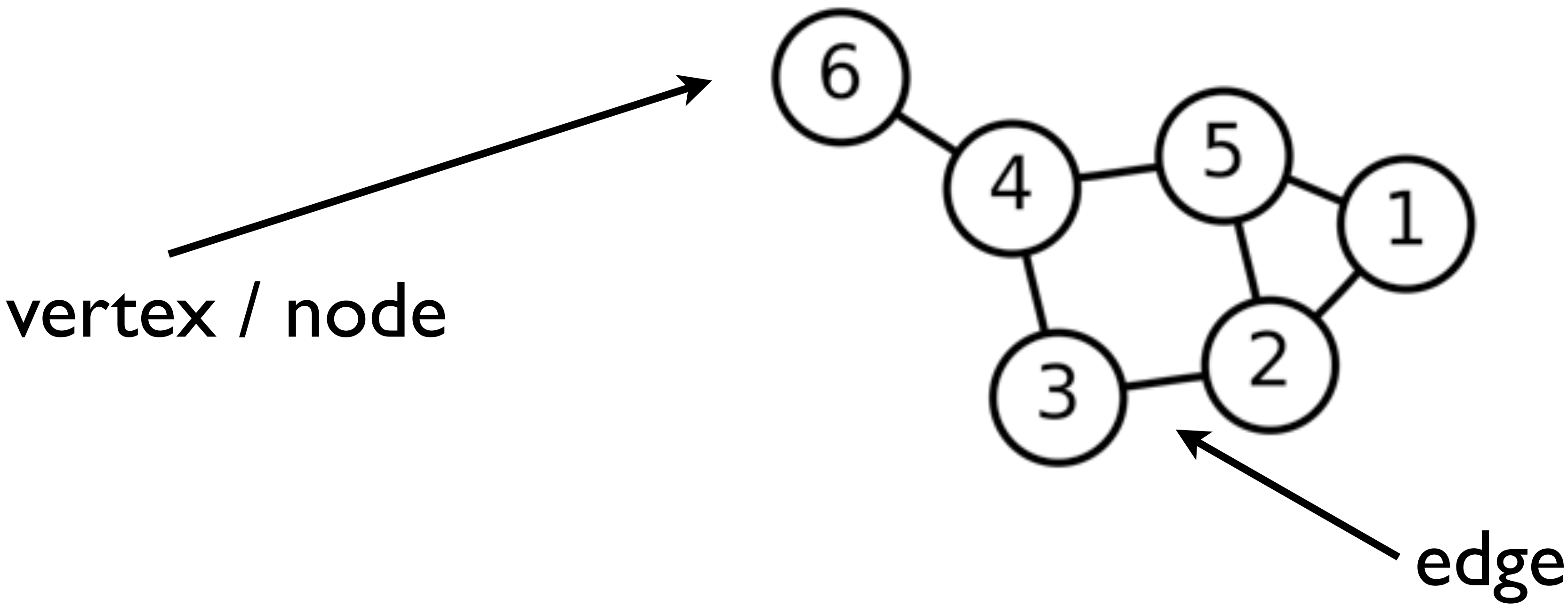
- Expectation

$$\mathbb{E}[X] = \int_r f(r) \cdot r$$

- Variance

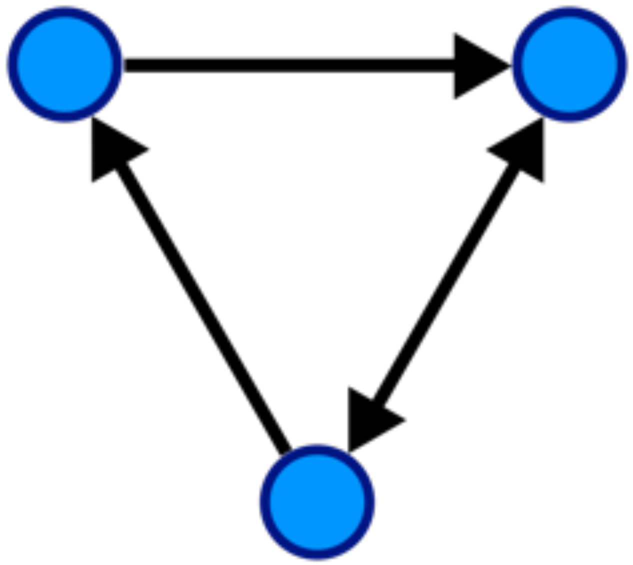
$$V[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Graphs



- G=(V,E)
- The edges are not directed (called undirected graph)
- undirected graph without cycles is called tree

Direct Graphs



- Directed graph without cycles is direct a cyclic graph (DAG)