腾讯会议+荷塘雨课堂

腾讯会议

会议 ID：462 7969 8867

会议密码：520520

注：前三周密码会更新

Chapter 1 Introduction to Probabilistic Graphical Models

2021 Fall

Jin Gu (古槿)

What will You Learn?

- **Change your minds**
- Solve your problems by **probability** & **graph**
  - General theories & methods
  - Advantages & disadvantages for different models
  - Go deep to either theory or application in future
- ***NOT*** *to learn*
  - *Programming skills*
  - *Deterministic models*

Model the World with ***Probability***

- Heisenberg's uncertainty principle
  - Position and momentum of a particle can not be known simultaneously.
  - So, even if you have an extreme accurate machine, you cannot determine all the initial values of a system. That is to say, the world cannot be accurately predicted.
- Life is full of uncertainty. Everybody needs to think and learn how to survive with it.
- *Probability*: a mathematical model of *uncertainty*

Decision making based on probabilities

Complexity with ***Decision***

- When the variables you need to consider are very large, human brain will struggle to get an "optimal" decision.
- Why?

The parameters increase **exponentially** with the number of variables!

For binary variables: $\sim 2^n$ parameters

Complexity with ***Decision***

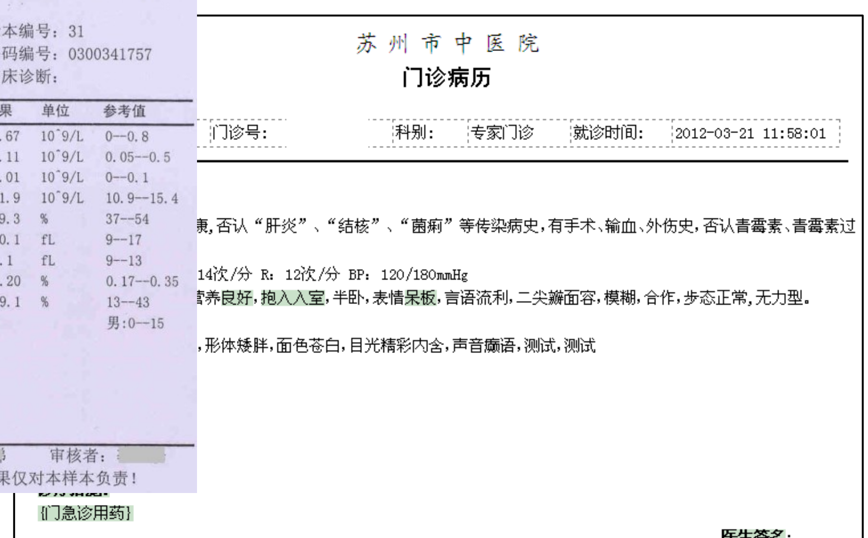
- Example: disease diagnosis and treatment
  - Doctors predict diseases based on many clinical features and diagnostic indications
  - Doctors infer the possible causes of the diseases
  - Doctors choose the most applicable treatments

市人民医院血常规报告单 [质评合格 省内参考]

姓 名: 病 案: 费 别: 标本编号: 31
性 别: 申请科室: 门诊抽血室 送检医师: 许沛然 条码编号: 0300341757
年 龄: 床 号: 标本种类: 临床诊断:

序号	代码	项目名称	结果	单位	参考值
1	WBC	白细胞	7.33	10ˆ9/L	4--10
2	RBC	红细胞	4.76	10ˆ12/L	3.5--5.5
3	HGB	血红蛋白	151	g/L	110--160
4	HCT	红细胞压积	44.1	%	36--50
5	MCV	红细胞平均体积	92.6	fL	82--100
6	MCH	平均血红蛋白量	31.7	pg	26--32
7	MCHC	平均血红蛋白浓度	342	g/L	320--360
8	PLT	血小板	215	10ˆ9/L	100--300
9	LYMPHP	淋巴细胞比率	32.10	%	20--40
10	NEUTP	中性细胞比率	57.20	%	50--70
11	MONOP	单核细胞比率	9.10	%	↑ 3--8
12	EOP	嗜酸性粒细胞比率	1.50	%	0.5--5
13	BASOP	嗜碱性粒细胞比率	0.10	%	0--1
14	LYMPHN	淋巴细胞数	2.35	10ˆ9/L	0.8--4
15	NEUT	中性细胞数	4.19		
16	MONON	单核细胞	0.67	10ˆ9/L	0--0.8
17	EON	嗜酸性粒细胞	0.11	10ˆ9/L	0.05--0.5
18	BASON	嗜碱性粒细胞	0.01	10ˆ9/L	0--0.1
19	RDW-CV	红细胞分布宽度-CV	11.9	10ˆ9/L	10.9--15.4
20	RDW-SD	红细胞分布宽度-SD	39.3	%	37--54
21	PDW	血小板分布宽度	10.1	fL	9--17
22	MPV	平均血小板体积	9.1	fL	9--13
23	PCT	血小板压积	0.20	%	0.17--0.35
24	P-LCR	大型血小板比率	19.1	%	13--43
25	ESR	血沉	8		男:0--15

核收时间: 2009-03-21 08:45 报告时间: 2009-03-21 09:15:32 检验者: 审核者:
备注: 此结果仅对本样本负责!



苏 州 市 中 医 院

门诊病历

门诊号: 科别: 专家门诊 就诊时间: 2012-03-21 11:58:01

康,否认“肝炎”、“结核”、“菌痢”等传染病史,有手术、输血、外伤史,否认青霉素、青霉素过

14次/分 R: 12次/分 BP: 120/180mmHg
营养良好,抱入入室,半卧,表情呆板,言语流利,二尖瓣面容,模糊,合作,步态正常,无力型。

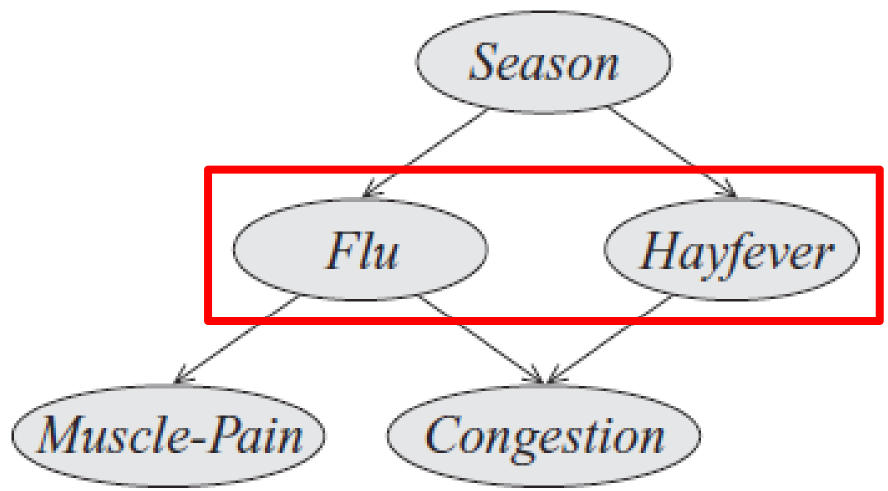
,形体矮胖,面色苍白,目光精彩内含,声音癫语,测试,测试

{门急诊用药}

医生签名:

Intuitive Representation with ***Graph***

- Example: infer whether a patient get *Flu* (流感) or *Hayfever* (花粉热)
  - Symptoms: Muscle-Pain, Congestion
  - Factors: Season



Intelligent Health Systems

IBM Watson Health

扁鹊 (秦越人)

Intelligent systems to help disease diagnosis and therapy by learning from literatures, EMRs, medical images, genetic data, etc.

Basics in with ***Probability***

- Classical probability
- Frequentist probability
  - Defined by random experiments

$$P(x) \approx \frac{n_x}{n_t}. \qquad P(x) = \lim_{n_t \to \infty} \frac{n_x}{n_t}.$$

- Bayesian probability
  - specifies some prior probability, which is then updated in the light of new, relevant data
  - Posterior probability $\mathrm{P}(X|Y) = \frac{\mathrm{P}(Y|X)\mathrm{P}(X)}{\mathrm{P}(Y)} = \frac{\mathrm{P}(Y|X)\mathrm{P}(X)}{\sum_X \mathrm{P}(Y|X)\mathrm{P}(X)}$

Basics in ***Probability***

- Random variable
- Probability, joint probability, marginal probability and conditional probability
- Probability function and probability density function

- **Independence** and **conditional independence**
- ……

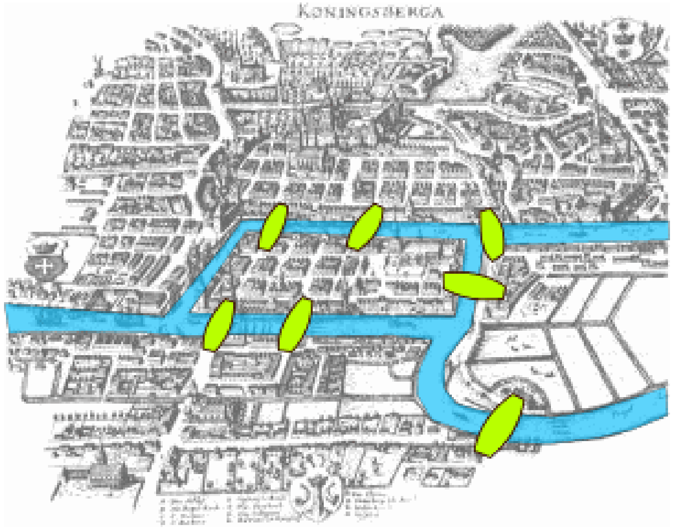
Basics in ***Probability***

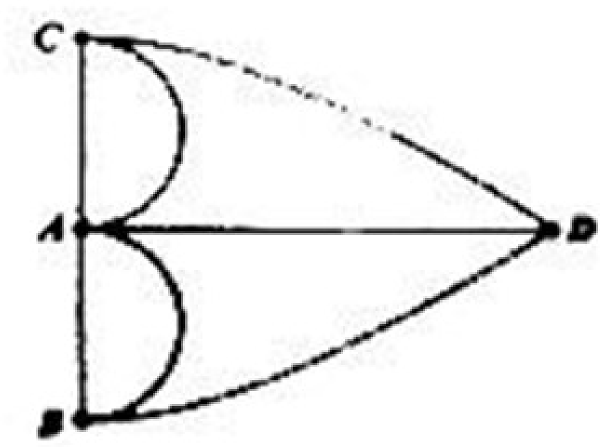
- Chain rule
  - $\mathrm{P}(X_1, X_2, \cdots, X_n) = \mathrm{P}(X_1\ )\mathrm{P}(X_2|X_1) \cdots \mathrm{P}(X_n|X_1, \cdots, X_{n-1})$
- Formula of total probability
  - $\mathrm{P}(X) = \sum_Y \mathrm{P}(XY) = \sum_Y \mathrm{P}(X|Y)\mathrm{P}(Y)$

- **Bayes' theorem** (revise your prior knowledge based on observed data)
  - $\mathrm{P}(X|Y) = \frac{\mathrm{P}(Y|X)\mathrm{P}(X)}{\mathrm{P}(Y)} = \frac{\mathrm{P}(Y|X)\mathrm{P}(X)}{\sum_X \mathrm{P}(Y|X)\mathrm{P}(X)}$

Basics in ***Graph***

- The Seven Bridges of Königsberg is a historically notable problem in mathematics. Its negative resolution by Leonhard Euler in 1735 laid the foundations of graph theory and prefigured the idea of topology.





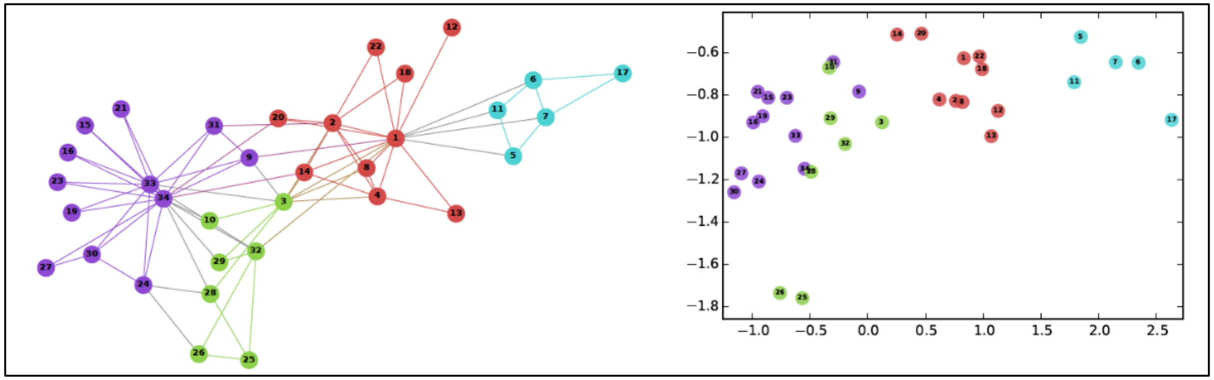
- Graph models
- Complex networks

Basics in ***Graph***

- The representations of graphs
  - G(V, E): nodes/vertex, edges
  - Adjacent matrix: $A=\{a_{ij}\}_{n*n}$
- The basic features of graphs
  - Path, trail, connectivity, …
  - Polytree, tree, chordal graph, …
  - Centrality: degree, betweenness, closeness
  - Sub-graph, cut, clique

Basics in ***Graph***

- Shortest path searching
- Random walk & diffusion on graph
- Spectral analysis
  - Comparable to principal analysis in vector space
- **Graph embedding**
  - Map nodes in graph to variables in vector space by preserving their “similarities



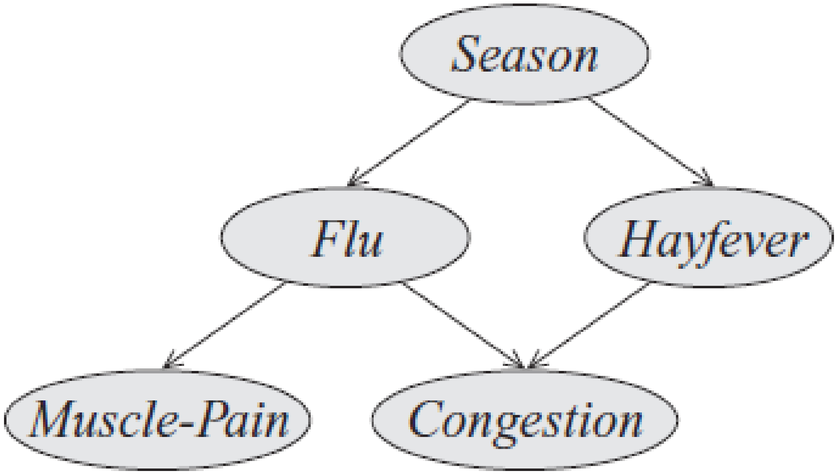
Outlines

- *Model the world with probability*
- *Model the world with graph*
- Probabilistic graphical models (PGMs)
- Some applications of different PGMs

- Course schedule (homework, quiz, …)
- Teachers and TAs

Probabilistic Graphical Models

- Intuitive
- Concise
- Integrative



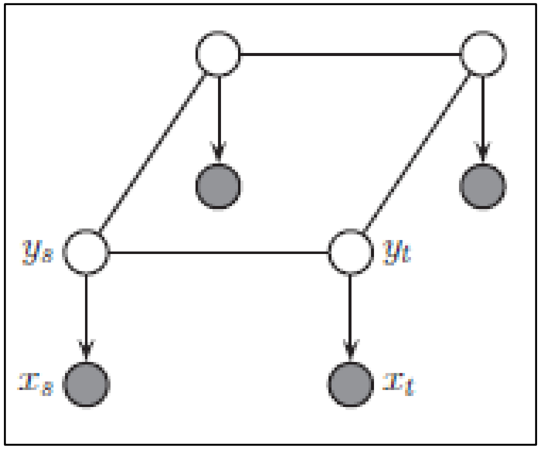
- <u>Uniform framework</u> for reasoning
  - Representation
  - Reasoning

$$(F \perp H \mid S)$$
$$(C \perp S \mid F, H)$$
$$(M \perp H, C \mid F)$$
$$(M \perp C \mid F)$$

$$P(S, F, H, C, M) = P(S)P(F \mid S)$$
$$P(H \mid S)P(C \mid F, H)P(M \mid F)$$

Probabilistic Graphical Models

- Intuitive
- Concise
- Integrative



- <u>Uniform framework</u> for reasoning
  - Representation
  - Reasoning

Not only a method by a view about *modeling* & problem solving

History: Bayesian Networks (Directed)

BAYESIAN NETWORKS: A MODEL OF SELF-ACTIVATED
MEMORY FOR EVIDENTIAL REASONING*

Judea Pearl
Computer Science Department
University of California
Los Angeles, CA 90024
(judea@UCLA-locus)
(213) 825-3243

Topics: Memory Models
Belief Systems
Inference Mechanisms
Knowledge Representation

Submitted to the
Seventh Annual Conference of
the Cognitive Science Society
15-17 August 1985

Paper length: 5666

**BAYESIAN NETWORKS: A MODEL OF SELF-ACTIVATED
MEMORY FOR EVIDENTIAL REASONING**

ABSTRACT

Bayesian networks are directed acyclic graphs in which the nodes represent propositions (or variables), the arcs signify the existence of direct causal dependencies between the linked propositions, and the strengths of these dependencies are quantified by conditional probabilities. A network of this sort can be used to represent the deep causal knowledge of an agent or a domain expert and turns into a computational architecture if the links are used not merely for storing factual knowledge but also for directing and activating the data flow in the computations which manipulate this knowledge.

The first part of the paper defines the properties of Bayes networks which are necessary to guarantee completeness and consistency, and shows how dependencies and conditional-independence relationships can be tested using simple link-tracing operations.

The second part of the paper deals with the task of fusing and propagating the impacts of new evidence and beliefs through Bayesian networks in such a way that, when equilibrium is reached, each proposition will be assigned a belief measure consistent with the observed data. We first argue that any viable model of human reasoning should be able to perform this task by a self-activated propagation mechanism, i.e., by an array of simple and autonomous processors, communicating locally via the links provided by the Bayes network itself. We then quote results which show that these objectives can be fully realized only in singly-connected networks, where there exists only one (undirected) path between any pair of nodes. Finally, the paper discusses several approaches to achieving belief propagation in more general networks, and argues for the feasibility of turning a Bayes network into a tree by introducing dummy variables, mimicking the way people develop causal models.

- Definition: directed acyclic graphs, nodes represent variables, edges represent direct causal dependences

History: Markov Networks (Undirected)

CONTEMPORARY MATHEMATICS

1

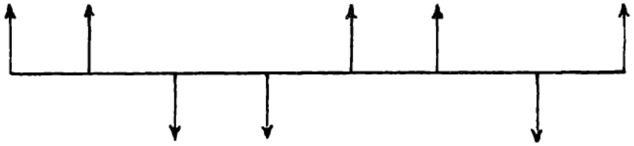
Markov Random Fields and Their Applications

Ross Kindermann
J. Laurie Snell

1980

American Mathematical Society
Providence, Rhode Island

The concept of a Markov random field came from attempts to put into a general probabilistic setting a very specific model named after the German physicist Ernst Ising. Ising was a student of Lenz and wrote his doctoral thesis on a model now called the Ising model. He tried to explain, using this model, certain empirically observed facts about ferromagnetic materials. When Ising (1925) published a summary of his results, he stated that the model was suggested by Lenz. A paper written by Lenz (1920) gives a very sketchy idea of the model.

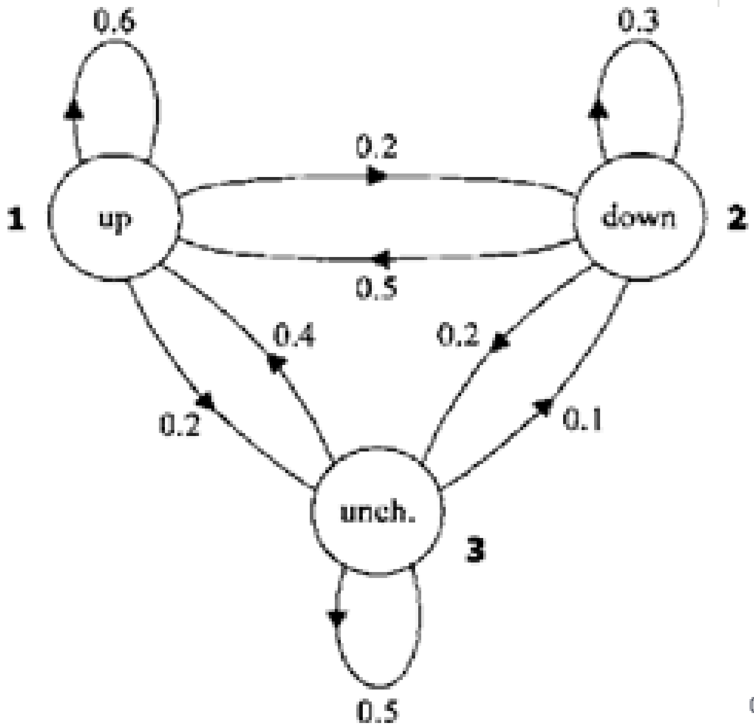


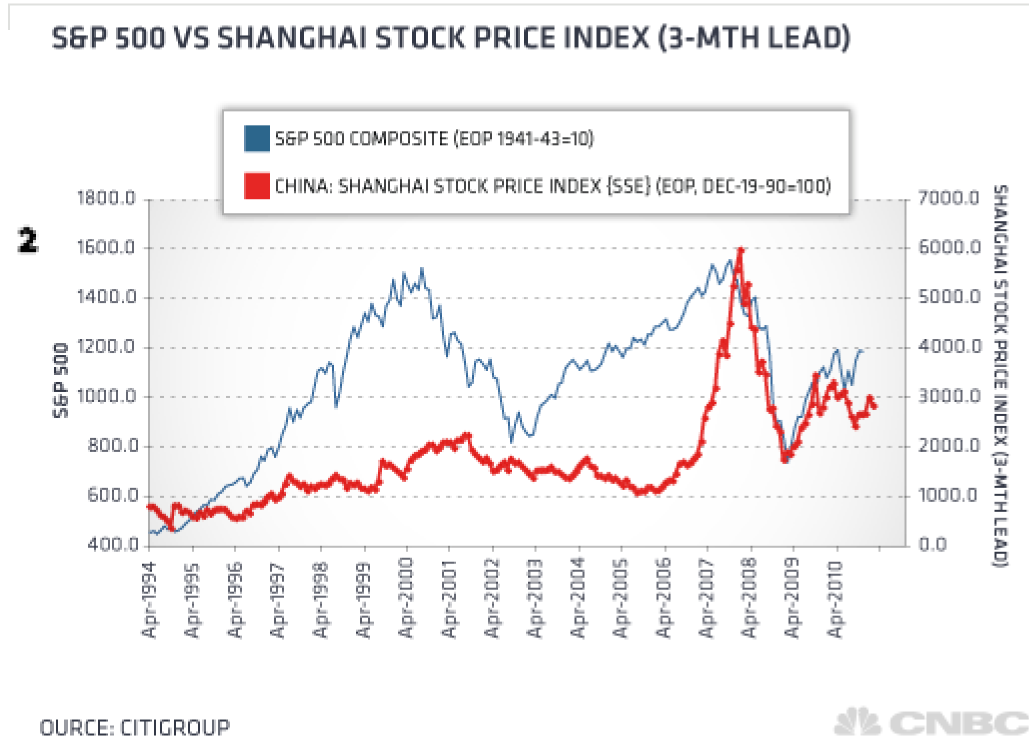
One dimension Ising model
(Proposed by Ernst Ising 1925)

+	+	-	+	-	+
+	+	-	+	-	+
+	+	-	+	-	+
+	+	-	+	-	+
+	+	-	+	-	+
+	+	-	+	-	+

The foundations of the theory of Markov random fields may be found in Preston (1974) or Spitzer (1971).

HMMs for Sequential Data





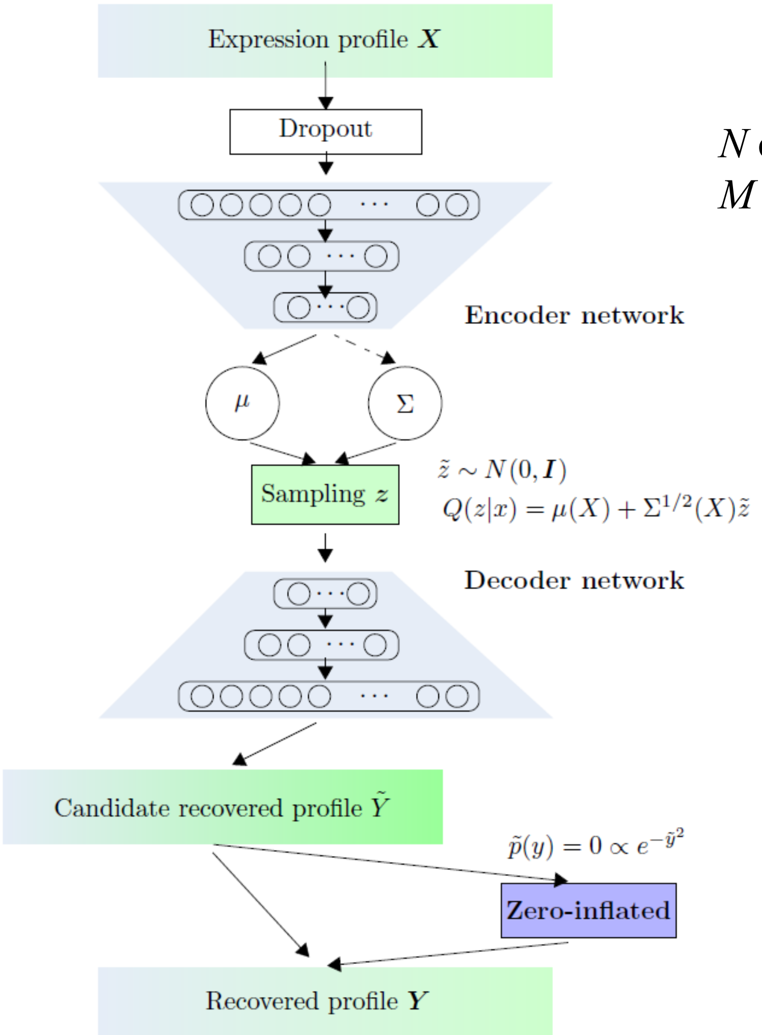
Initial state probability matrix

$$\boldsymbol{\pi} = (\pi_i) = \begin{pmatrix} 0.5 \\ 0.2 \\ 0.3 \end{pmatrix}$$

State-transition probability matrix

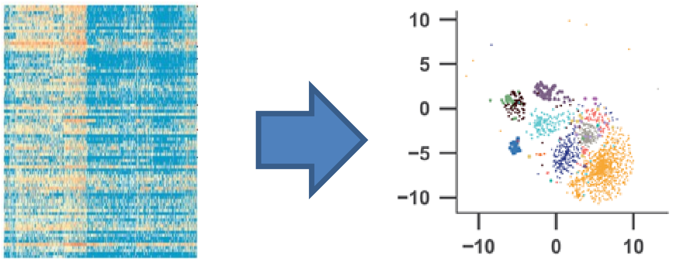
$$\mathbf{A} = \{a_{ij}\} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

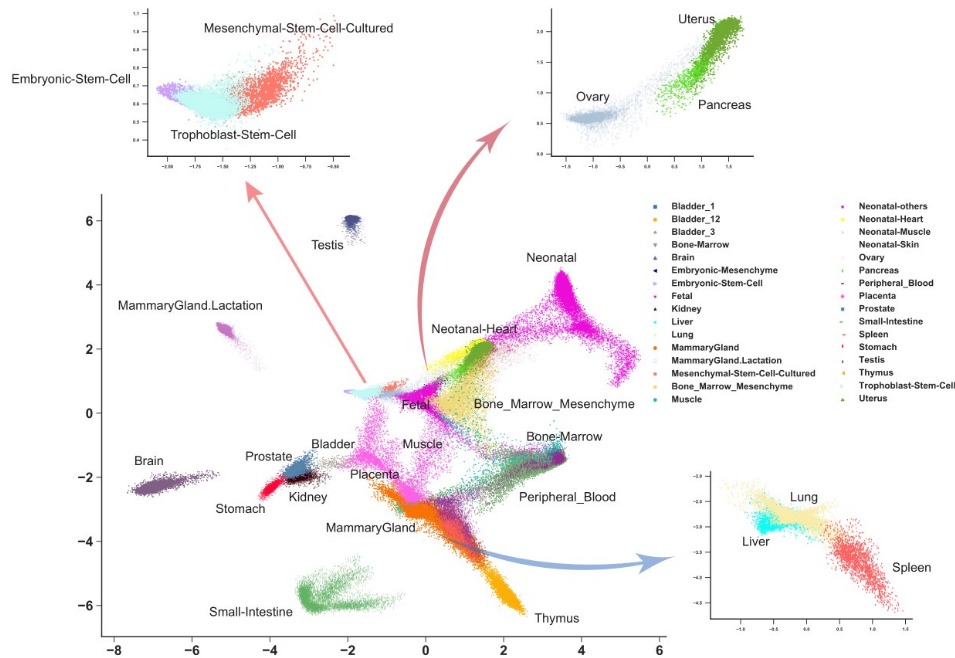
VAE for Dimension Reduction



Loss(X, Y)=binary_entropy(X, Y)+KL($Q(z|X)||P(z)$)

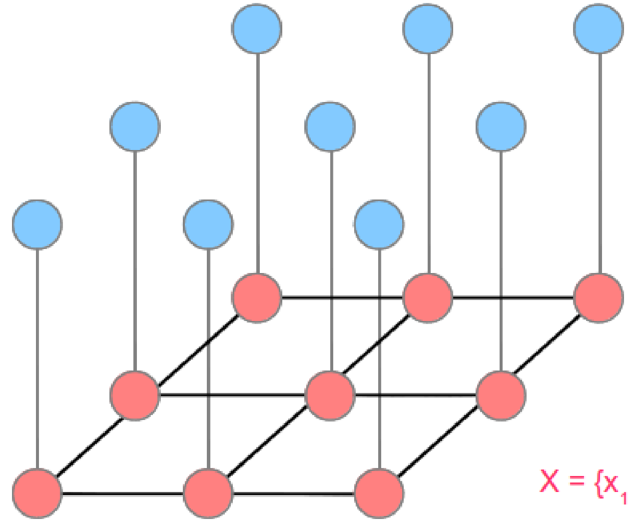
N cells, M genes
$M >> N$





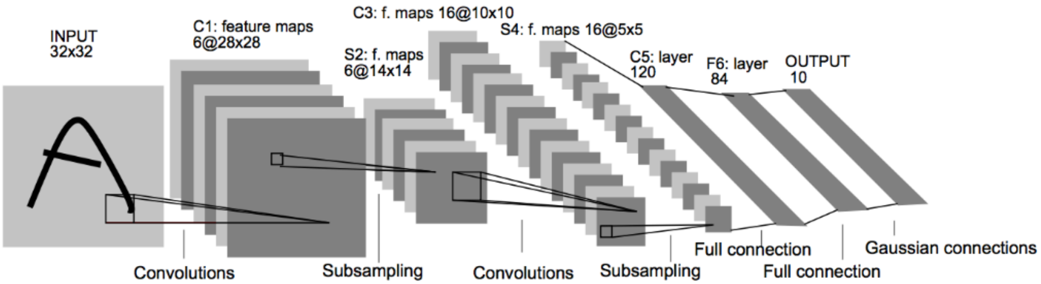
Project mouse cells to 2D (~10,000 features)

CRFs for Image Segmentation

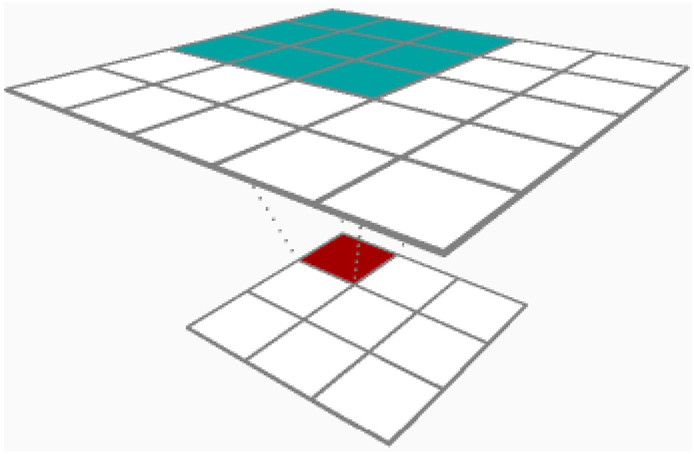


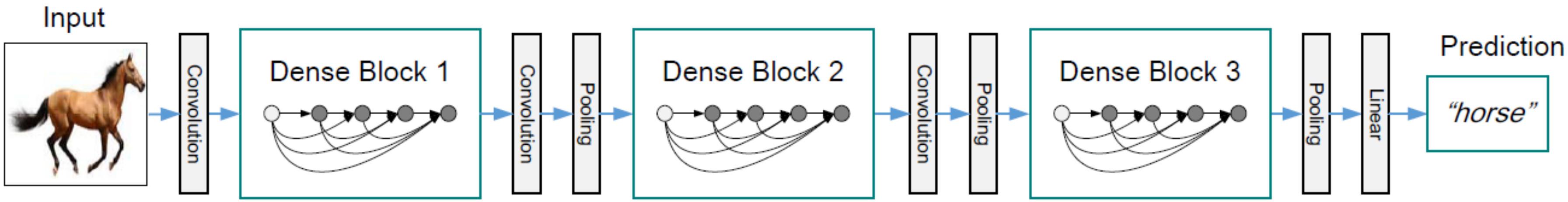
The hidden nodes should have the similar labels of their neighbors. The local label density can be described as "potential".

CNNs for computer vision



CNN called LeNet by Yann LeCun (1998)





Huang et al. Densely Connected Convolutional Networks. CVPR 2017 (Best paper)

CNNs can be treated as stacked filters of the input images.
You can easily design your models as PGMs.

Course Content

- ***Representation***
  - (W1) Chapter 1. Introduction
  - (W2) Chapter 2. Bayesian Network
  - (W3) Chapter 3. Local Probabilistic Models
  - (W4) Chapter 4. Dynamic Bayesian Networks
  - (W5) Chapter 5. Markov Random Fields
  - (W6) Chapter 6. Advanced Models
- ***Course Project & Quiz I* (W7)**
- ***Inference***
  - (W8-W9) Chapter 7. Particle Based Approximate Inference
  - (W9) Chapter 8. Cluster Graph & Belief Propagation
  - (W10) Chapter 9. Variational Inference
- ***Learning***
  - (W11) Chapter 10. Parameter Learning
  - (W12) Chapter 11. Learning with Incomplete Data
  - (W13) Chapter 12. Structure Learning
  - (W14) Chapter 13. Causal Models
- ***Final Review & Quiz II* (W15)**

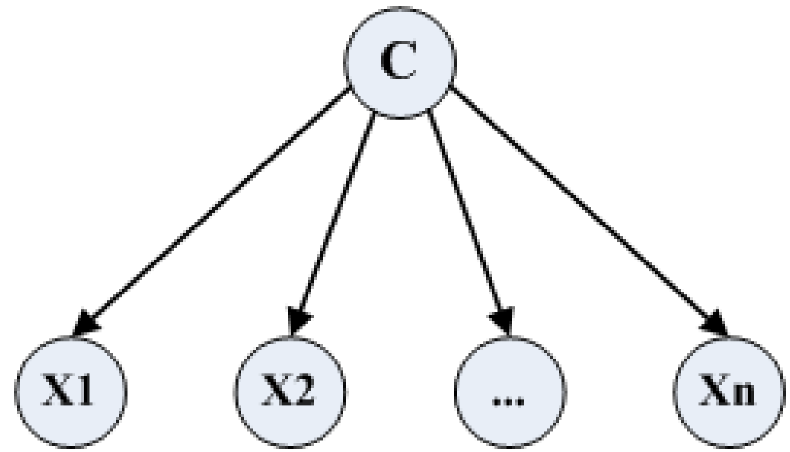
Naive Bayesian Model

- Description: if the class label is given, all the observation variables are independent

- Its joint probability should be
  - $P(X_1, X_2, X_3, C) = P(X_1|C)P(X_2|C)P(X_3|C)P(C)$

- For example, the doctors need to diagnose the disease (denoted by *C*) based on a set of symptoms (denoted by X_i)

Representation

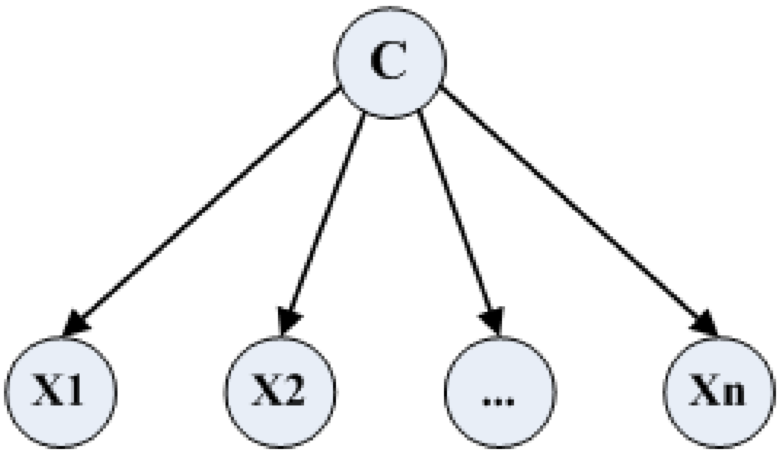


$$\mathbf{P} \Leftrightarrow \{\mathbf{P}, \mathbf{G}/\mathbf{H}\}$$

$$\mathbf{P} = \prod_i \mathbf{P}(X_i|C)\mathbf{P}(C)$$

C indicates the possible disease label
X_i indicates the existence/degree of the i-th symptom

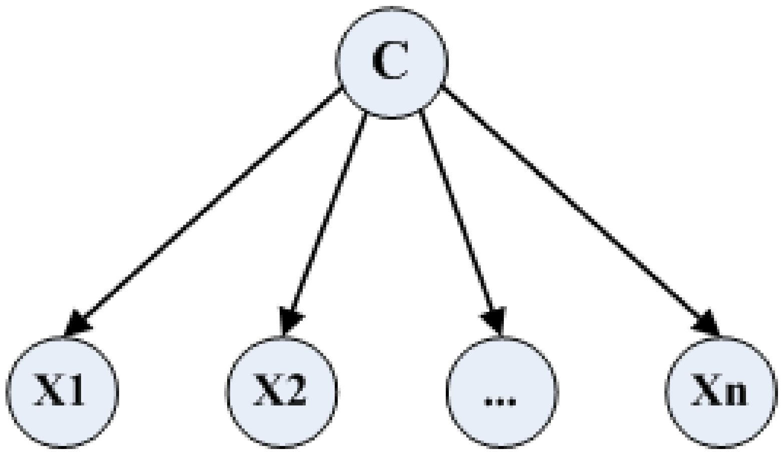
Inference



$$\mathbf{P}(C|X=x,\theta)$$

If the model is known, doctors can ***infer*** the possible disease ***C*** based on the observed symptoms ***X=x***!!

Learning



$$\mathbf{P}(\theta \mid \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots)$$

If the model is unknown, doctors need to ***learn*** a model by studying a set of patients!!

Course Assessment (Proposed)

- Assignments (40 points)
  - Paper assignments (20 points)
  - Practice (20 points)
- Course project (20 points)
  - Proposal (LDA & Random Fields)
  - Final report
- Two quiz (40 points)
  - Graph model representation (20 points)
  - Inference and learning (20 points)

Previous Projects

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

SocoTraveler: Travel-package recommendations leveraging social influence of different relationship types

Jiangning He[a,b], Hongyan Liu[a,b,*], Hui Xiong[c]

[a] Research Center for Contemporary Management, Tsinghua University, Beijing, China
[b] School of Economics and Management, Tsinghua University, Beijing, China
[c] Management Science and Information Systems Department, Rutgers, The State University of New Jersey, United States

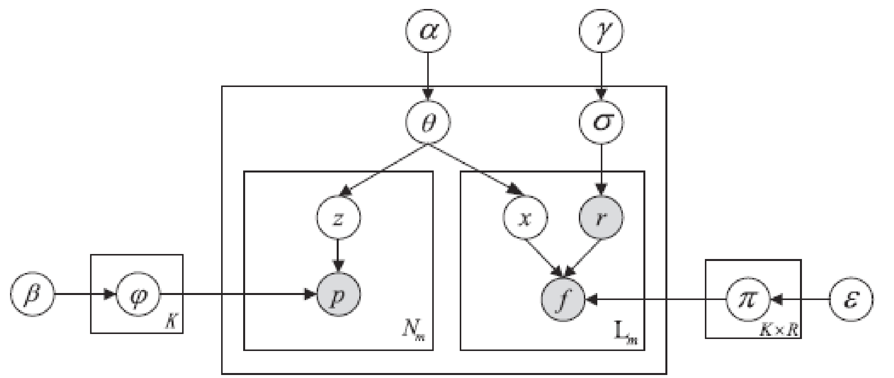


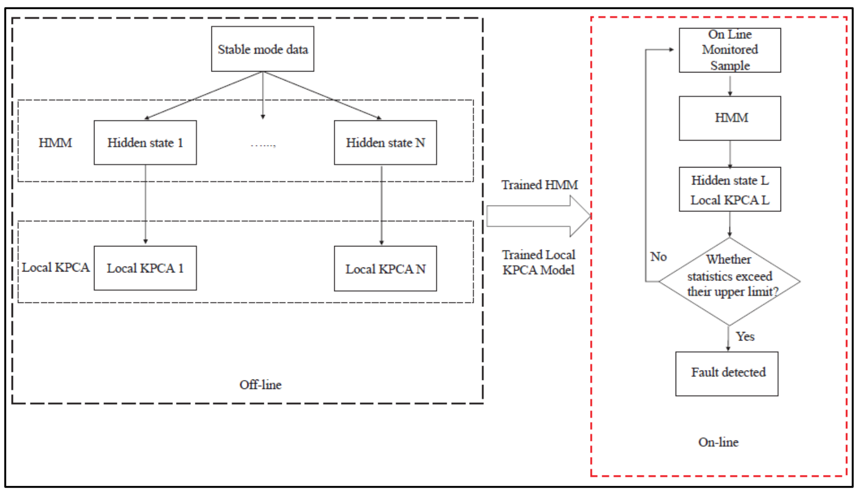
Fig. 4. Graphical representation of *socoLDA*. The Observed variables *p*, *f*, and *r* denote travel packages, co-travelers, and relationship types between users and co-travelers, separately. The hidden variables *x* and *z* denote topics assigned for travel packages and topics assigned for co-travelers, separately.

感谢清华大学自动化系的古槿老师在授课期间和科研建模中给我的耐心指导；古老师讲授的《概率图模型》一课深入浅出，奠定了我博士研究的方法基础。感谢清华经管学院

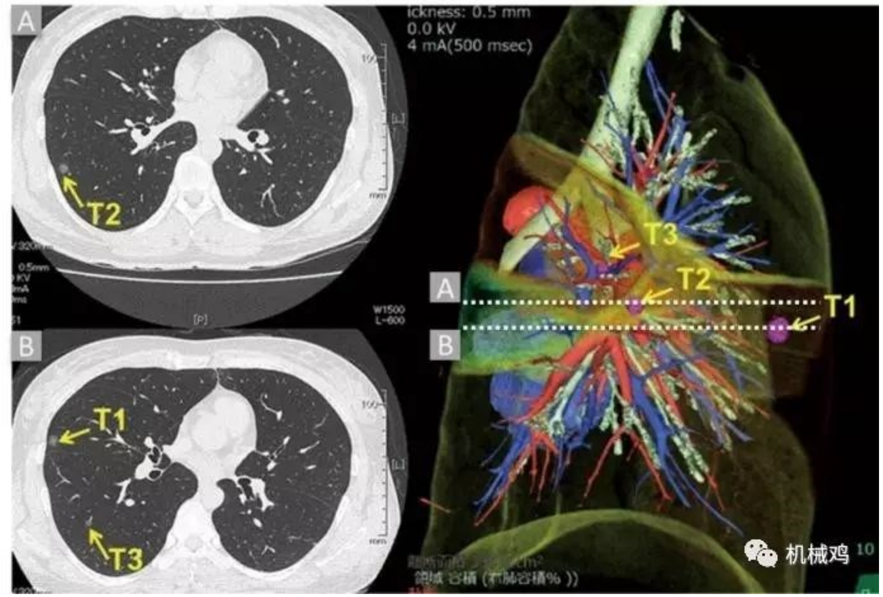
Previous Projects

Hidden Markov Model Combined with Kernel Principal Component Analysis for Nonlinear Multimode Process Fault Detection

Peng Peng, Jiaxin Zhao, Yi Zhang, Heming Zhang



已接收会议论文：
结合核PCA的HMM模型用于故障诊断



Kaggle challenge: NCI为数据科学竞赛提供了2000个匿名的高分辨率CT扫描，每个图像包含千兆字节的数据。

Contact Information

- Jin Gu 古槿
  - Office: FIT 1-111/112C
  - jgu@tsinghua.edu.cn
  - Phone: 62794294-866
  - **08:30~09:30, Friday**

Beijing National Research Center for Information Science and Technology
北京信息科学与技术国家研究中心
Department of Automation
清华大学自动化系

- 助教队伍
  - **Wenbo Guo 郭文博**
    - FIT Building, 1-111/112
    - gwb17@mails.tsinghua.edu.cn
  - **Xinqi Li 李欣琪**
    - FIT Building, 1-111/112
    - lxq19@mails.tsinghua.edu.cn

Textbooks & References

- Textbooks
  - Koller D, Friedman N. **Probabilistic Graphical Models: Principles and Techniques**. *The MIT Press* 2009.
  - Murphy K. **Machine Learning: A Probabilistic Perspective**. *The MIT Press* 2012.
- Other references
  - Goodfellow I, Bengio Y, Courville A. **Deep learning**. *The MIT Press* 2016.
  - Wainwright MJ, Jordan MI. **Graphical Models, Exponential Families, and Variational Inference**. *Foundations and Trends in Machine Learning* 2008, 1(1-2):1-305.
  - Hastie T, Tibshirani R, Friedman J. **The Elements of Statistical Learning**. *The Springer Press* 2009.

Some Practical E-Books & Resources

- Ankur Ankan & Abinash Panda. **Mastering probabilistic graphical models using python**. *Packt Publishing* 2015.
- David Bellot. **Learning probabilistic graphical models in R**. *Packt Publishing* 2016.

- Software listed by Kevin Murphy
  - http://people.cs.ubc.ca/~murphyk/Software/bnsoft.html
- TensorFlow, PyTorch, …
- Many MOOC courses

作业基本要求

- 请尽快在网络学堂维护个人邮箱与手机号
- 书面作业可以提交纸版或者电子版（**尽量提交纸版**，便于批改，可打印）
- 纸版请在**下一次上课前交给助教**，电子版请在**周日晚12点前从网络学堂**提交
- 截止时间以后（第14周课程前）补交作业，按照0.6系数折算
- 如发现抄袭，**抄袭与被抄袭作用均判为0分，相关情况上报院系处理**

重要：教学日历调整

- **9月18日周六上课！**
- **10月4日周一上课！**