Statistical inference with graphical models

Jaesik Choi

Probabilistic Artificial Intelligence Lab (PAIL)

Ulsan National Institute of Science and Technology (UNIST)

http://pail.unist.ac.kr/jaesik

*** Includes some slides from Lise Getoor (U of Maryland)**

Contents

- **Machine Learning Revisited**
- Bayesian Learning
- Graphical Models and Inference Algorithms
- Lifted Graphical Models and Inference
- Relational Kalman Filtering
- Appendix: Kaggle Competition

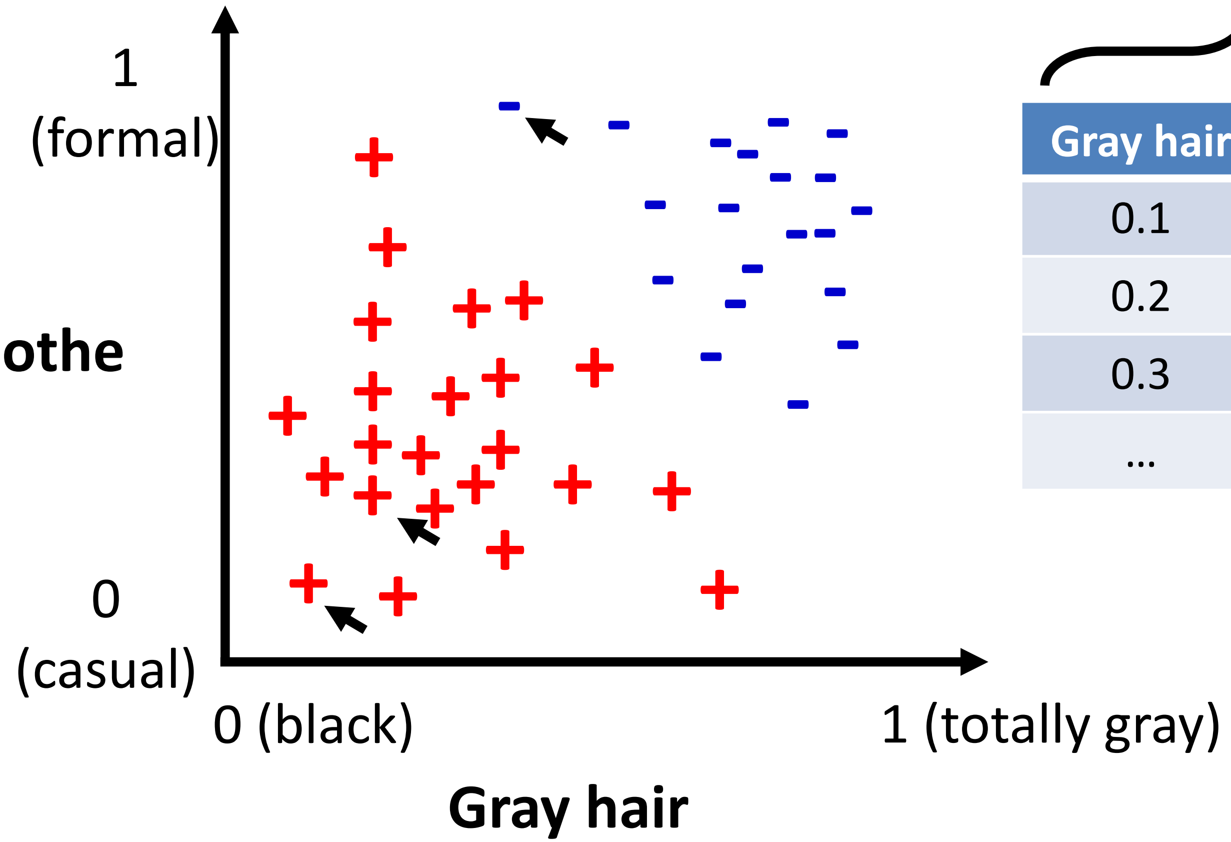
What is the Machine Learning problem?
Guessing game

- We want to correctly classify an event.
- E.g., Guessing game: occupation of the attendee.
  - Input features x:
    - Glasses
    - Height
    - Gray hair
    - Clothe
    - …
  - Output y:
    - Student or not student.

What are relevant features?
Features: "Gray Hair" and "Clothe"

- We want to classify data and label correctly.
- E.g., Guessing game: occupation of the attendee.
  - Input features x:
    - Glasses
    - Height
    - Gray hair (black:0 ~ completely gray:1)
    - Clothe (casual:0 ~ formal:1)
    - …
  - Output y:
    - Student (+) or not student(-).

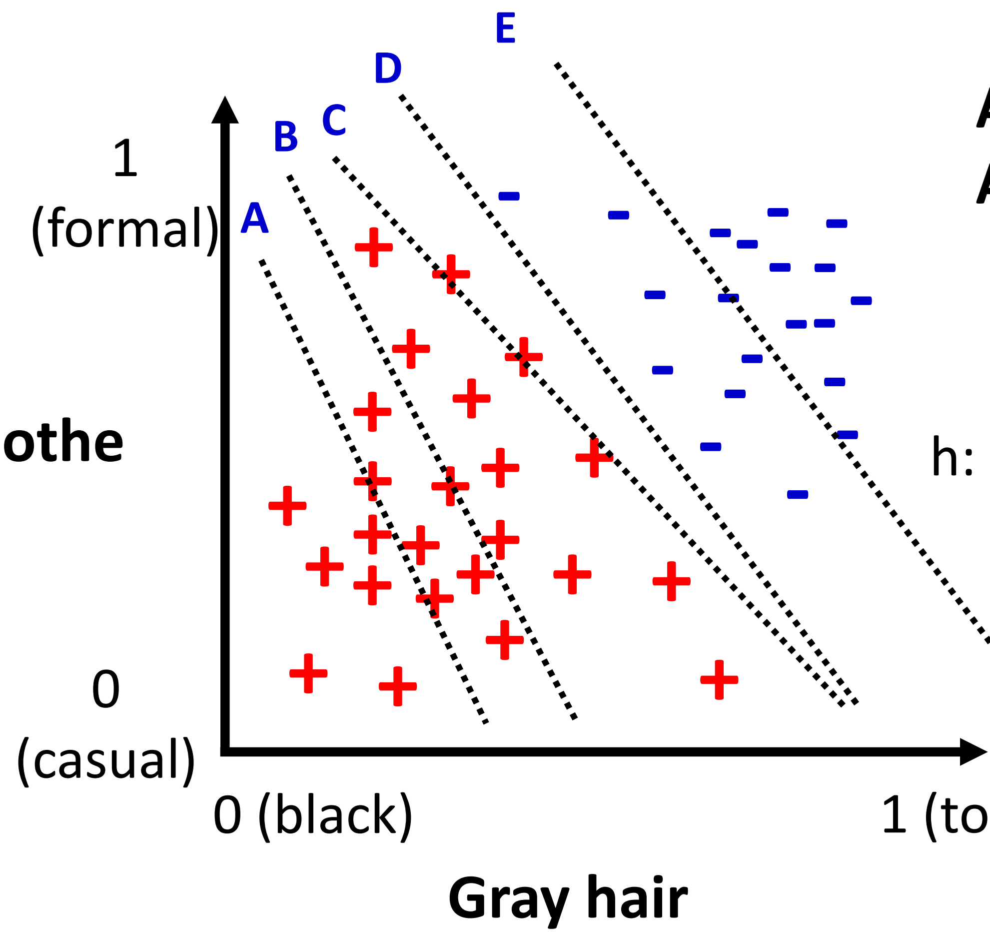
How does the data look like?
Let’s plot the data (x,y)



Features: x — Label: y

Gray hair	Clothe	Student
0.1	0.1	Yes
0.2	0.3	Yes
0.3	1	No
…	…	…

How can we classify the data points?
With a hypothesis space: e.g., lines



A hypothesis space H:
A set of all separating lines

$\mathrm{h} \in \mathrm{H}$

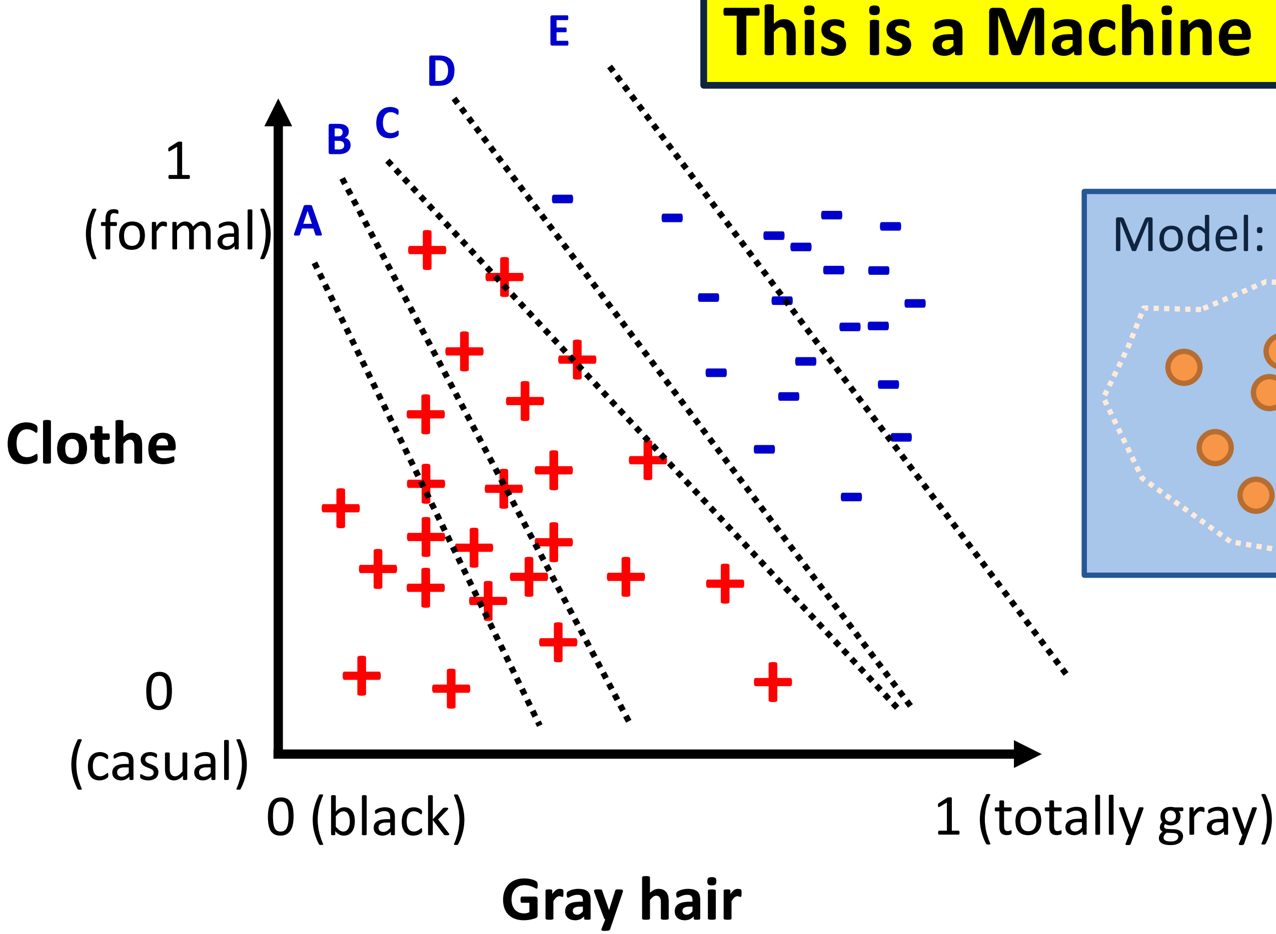
h: w_{gh}*[Gray hair]+w_c*[Clothe] < θ

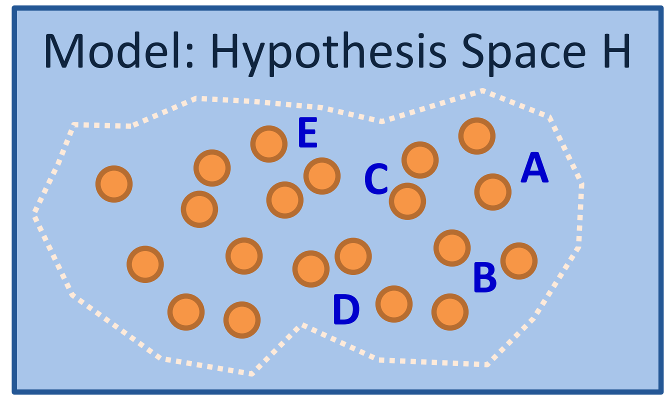
$$f_h(x) = \begin{cases} +, & w \cdot x < \theta \\ -, & \text{othewise} \end{cases}$$

How can we classify the data points?

With a hypothesis space: e.g., lines

This is a Machine Learning Model

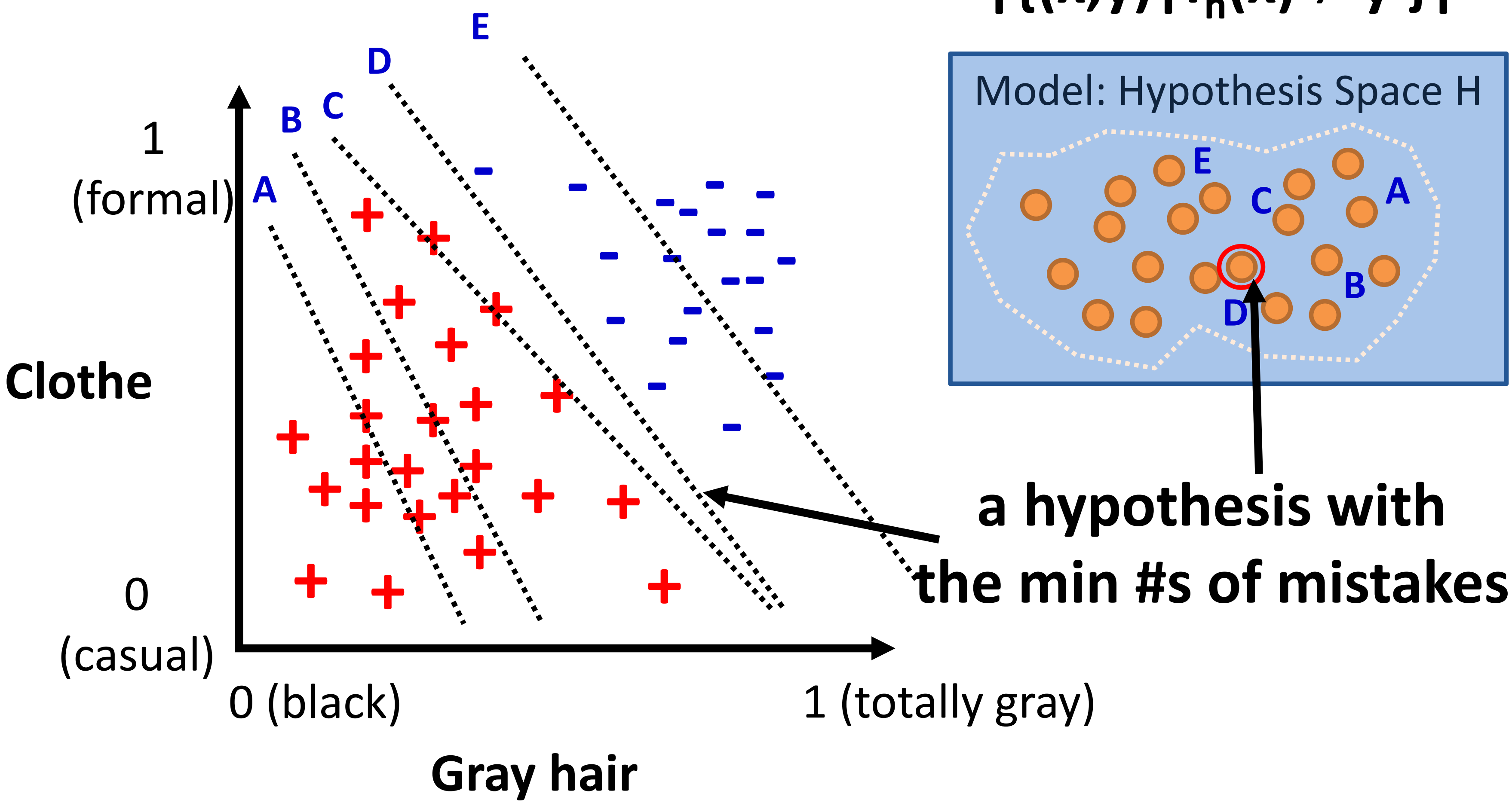




What is the **best** hypothesis?

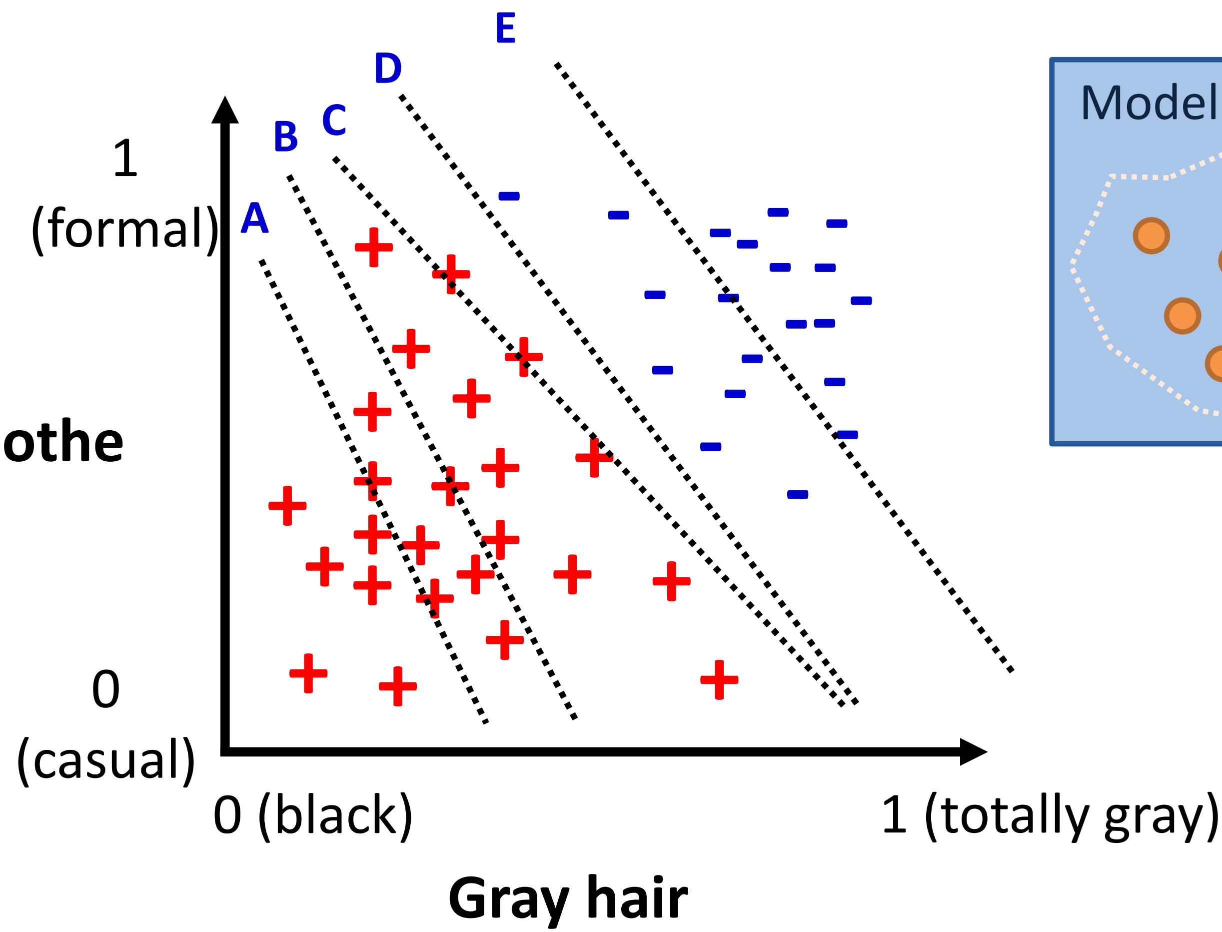
Based on a measure: e.g., # of mistakes

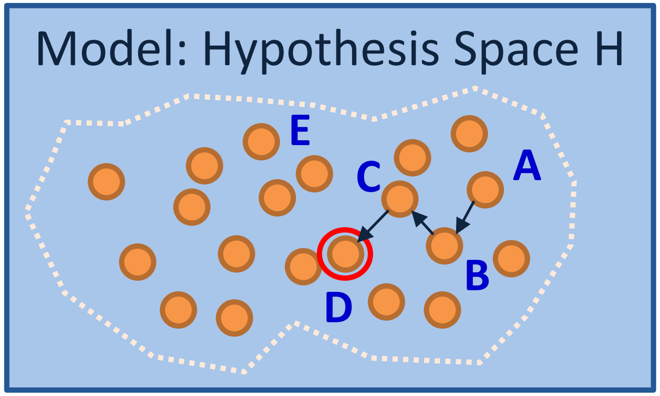
$= |\{(x,y) \mid f_h(x) \neq y \}|$



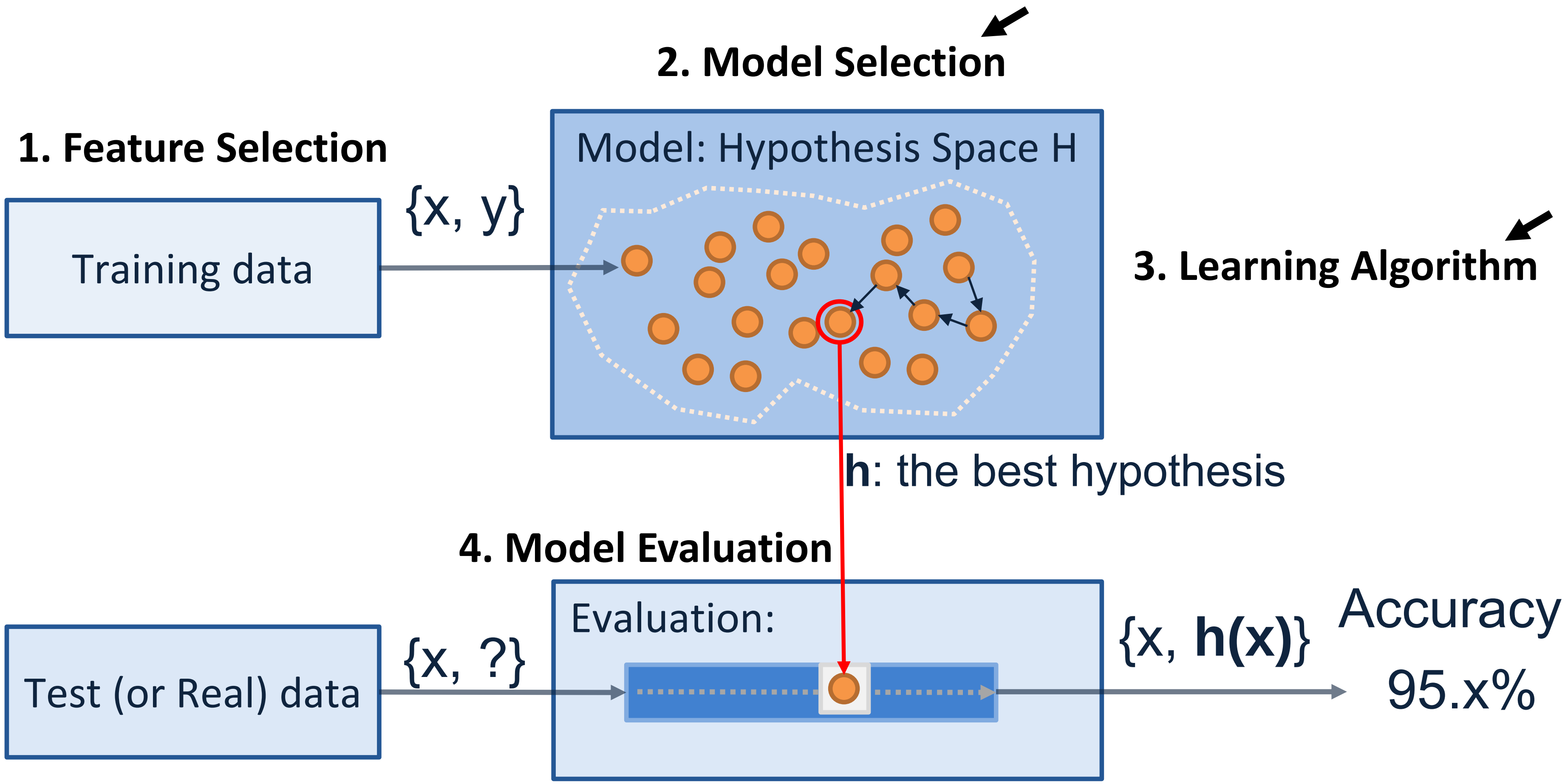
How can we learn the best Hypothesis?

With learning (searching) algo.: e.g., Gradient Decent

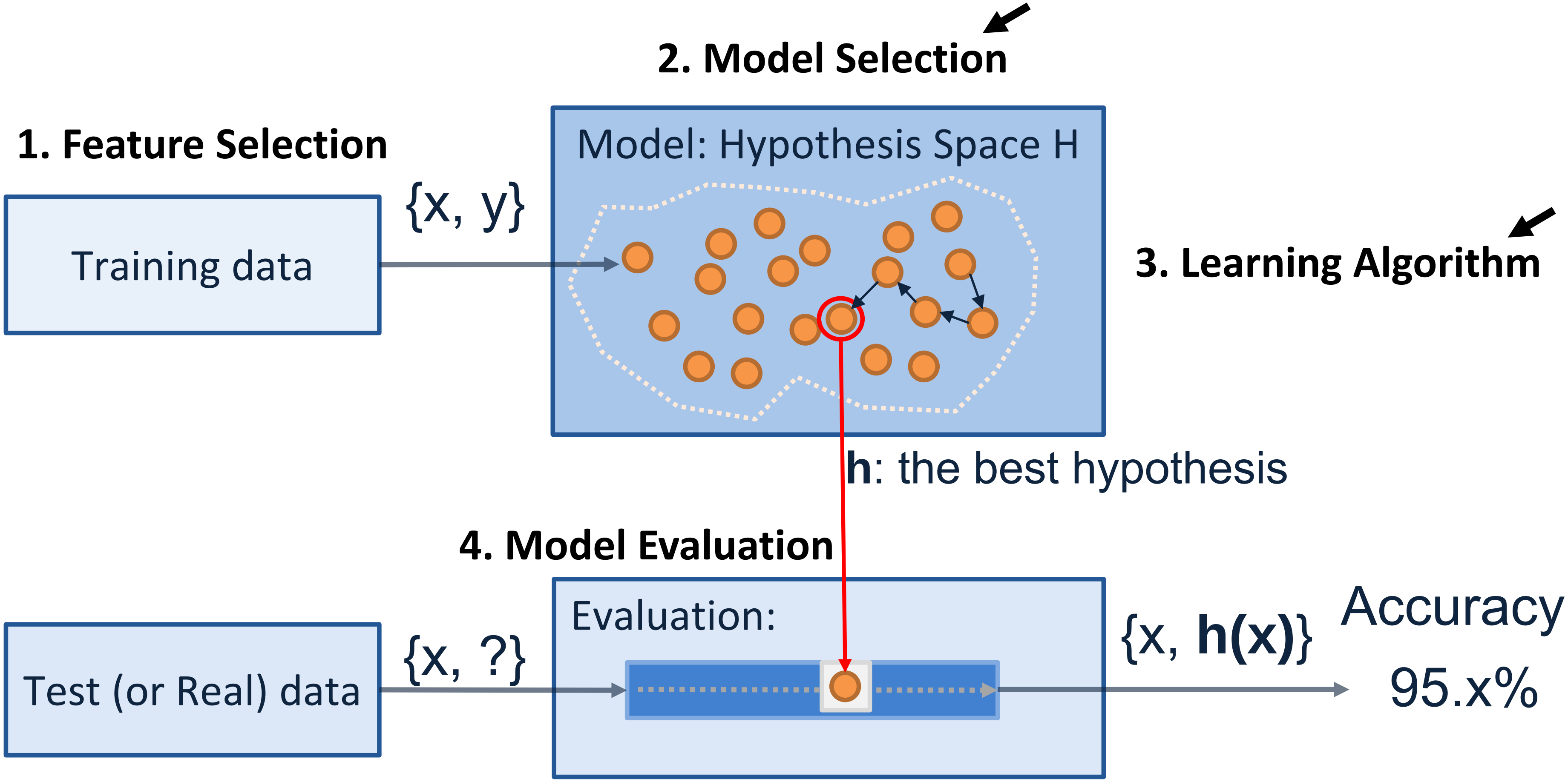




Machine Learning: Big Picture
Discriminative Model



Machine Learning: Big Picture
Discriminative Model



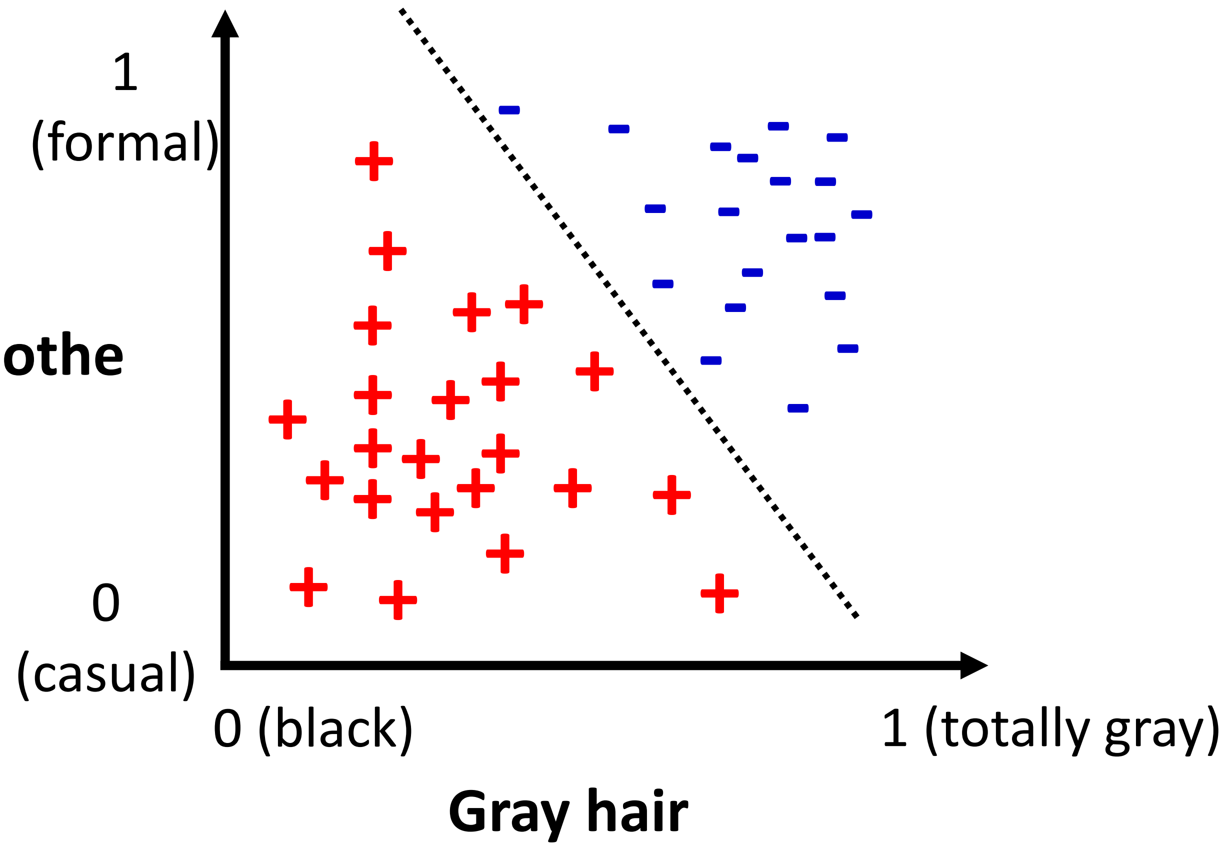
Well, are we done now?
What if, we want to model the data distribution?

Contents

- Machine Learning Revisited
- **Bayesian Learning**
- Graphical Models and Inference Algorithms
- Lifted Graphical Models and Inference
- Relational Kalman Filtering
- Appendix: Kaggle Competition

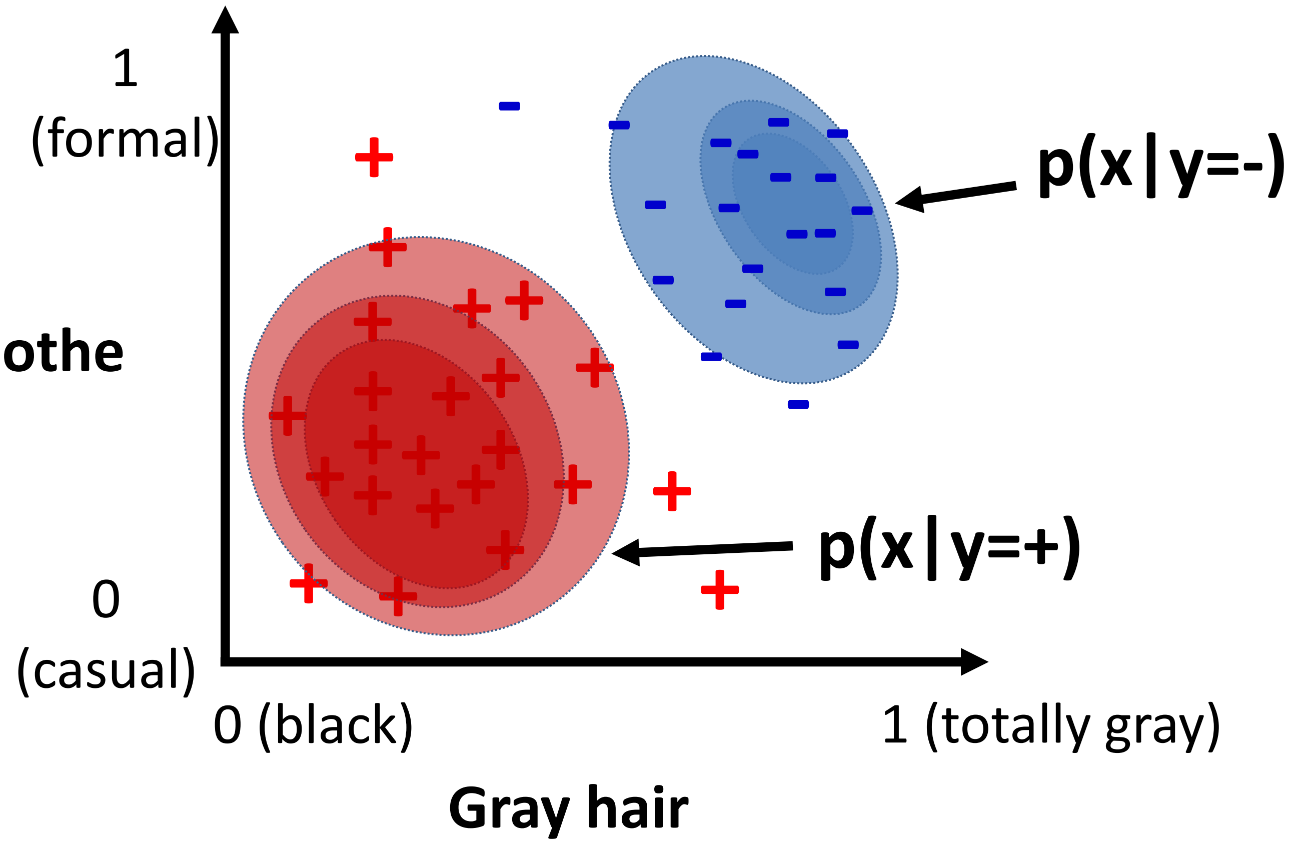
How can we model the data distribution?

Model the distribution p(x|y)



How can we model the data distribution?

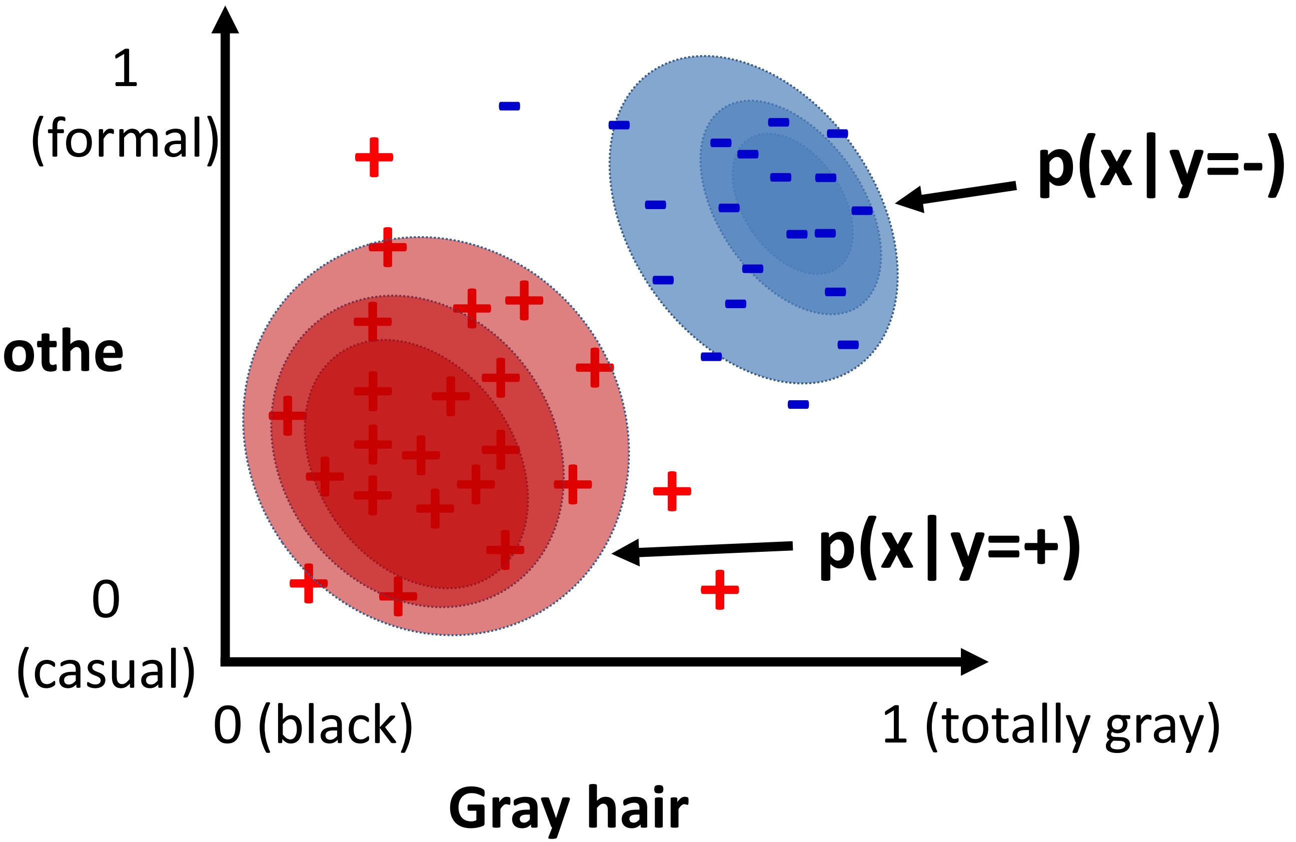
Model the distribution p(x|y)



How can we model the data distribution?

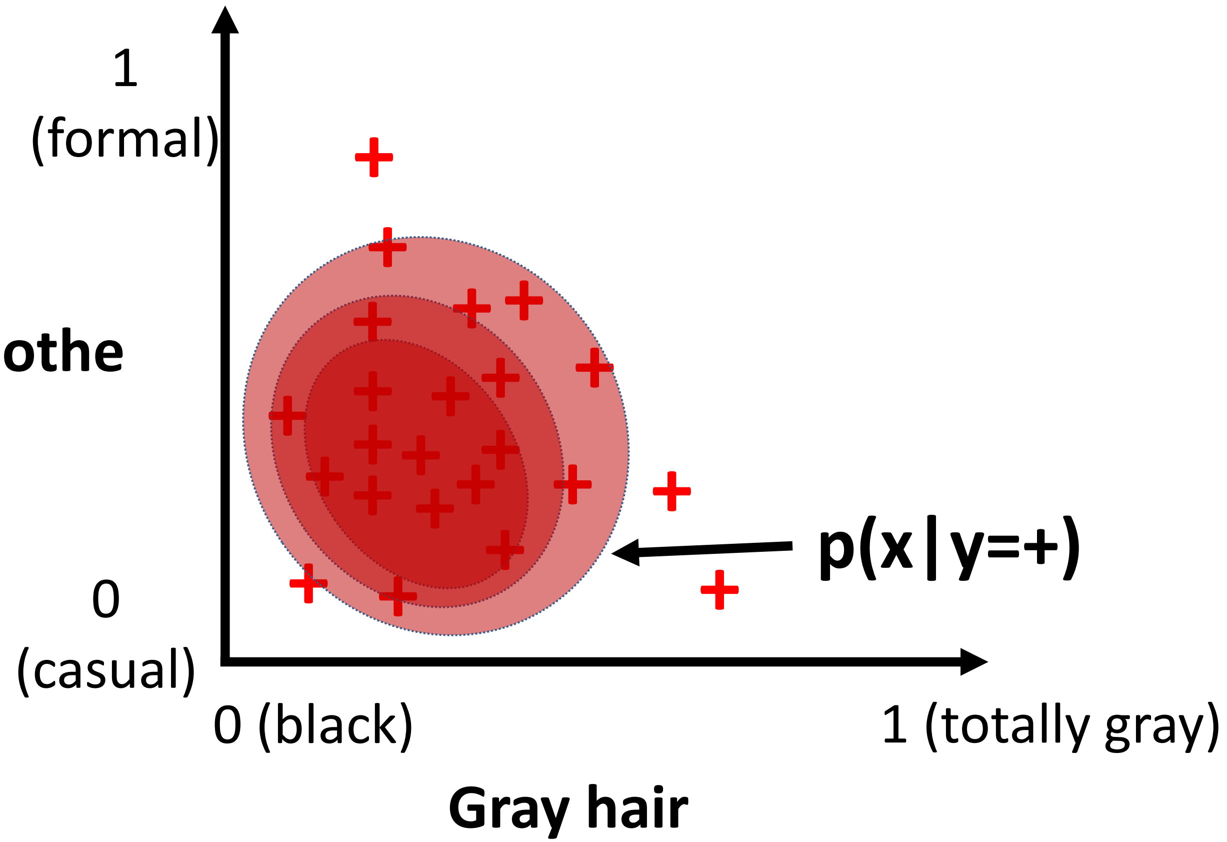
Model the distribution p(x|y)

If we can model p(x|y=-) and p(x|y=+), we can answer p(y|x).



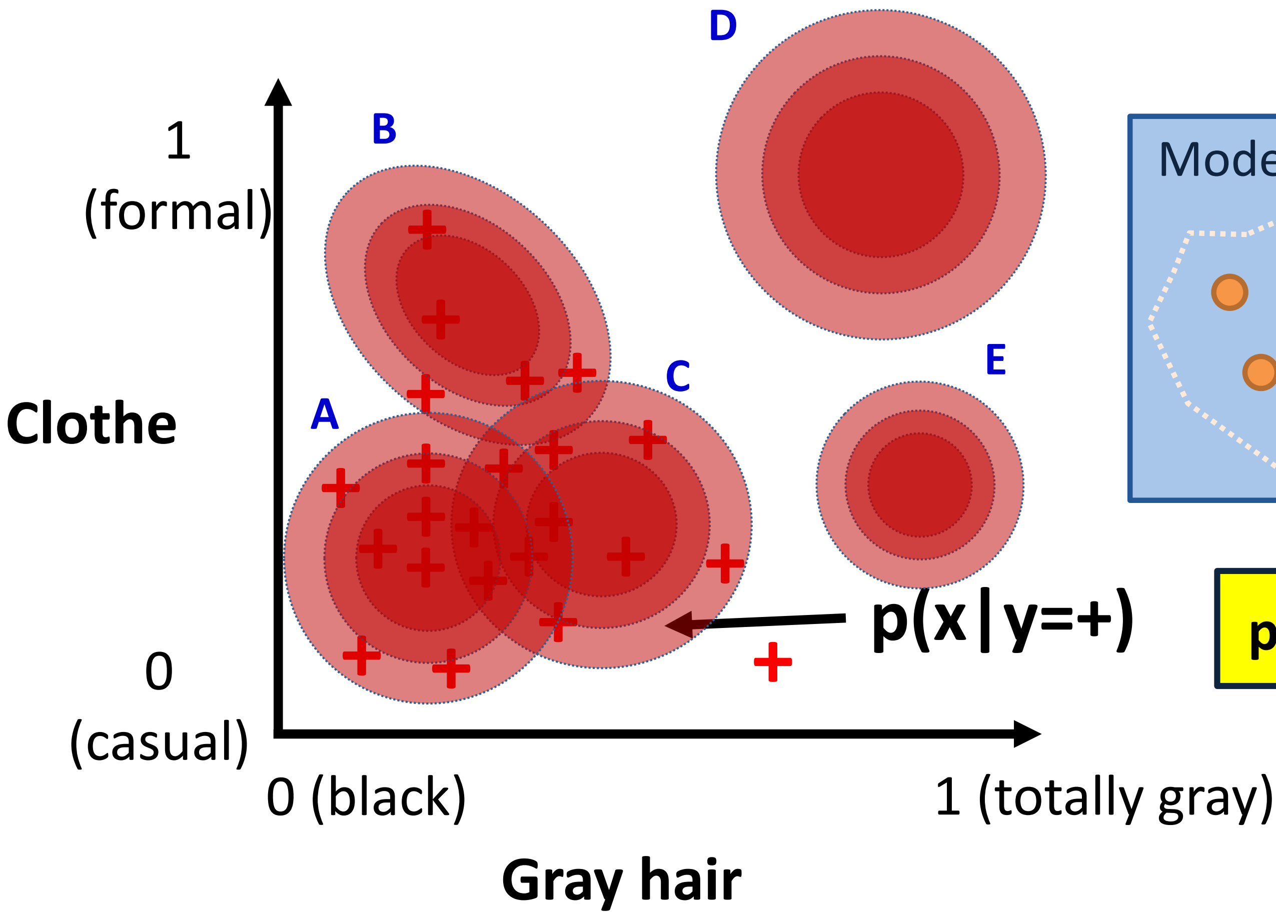
How can we model the data distribution?

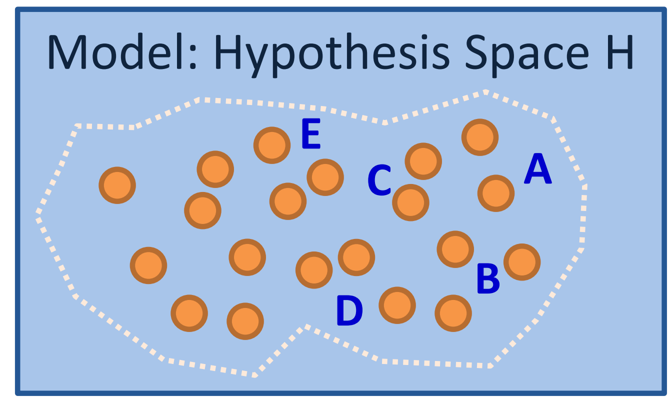
Model the distribution p(x|y=+)



How can we model the data distribution?

Model p(x|y=+): e.g., Mixture of Gaussian



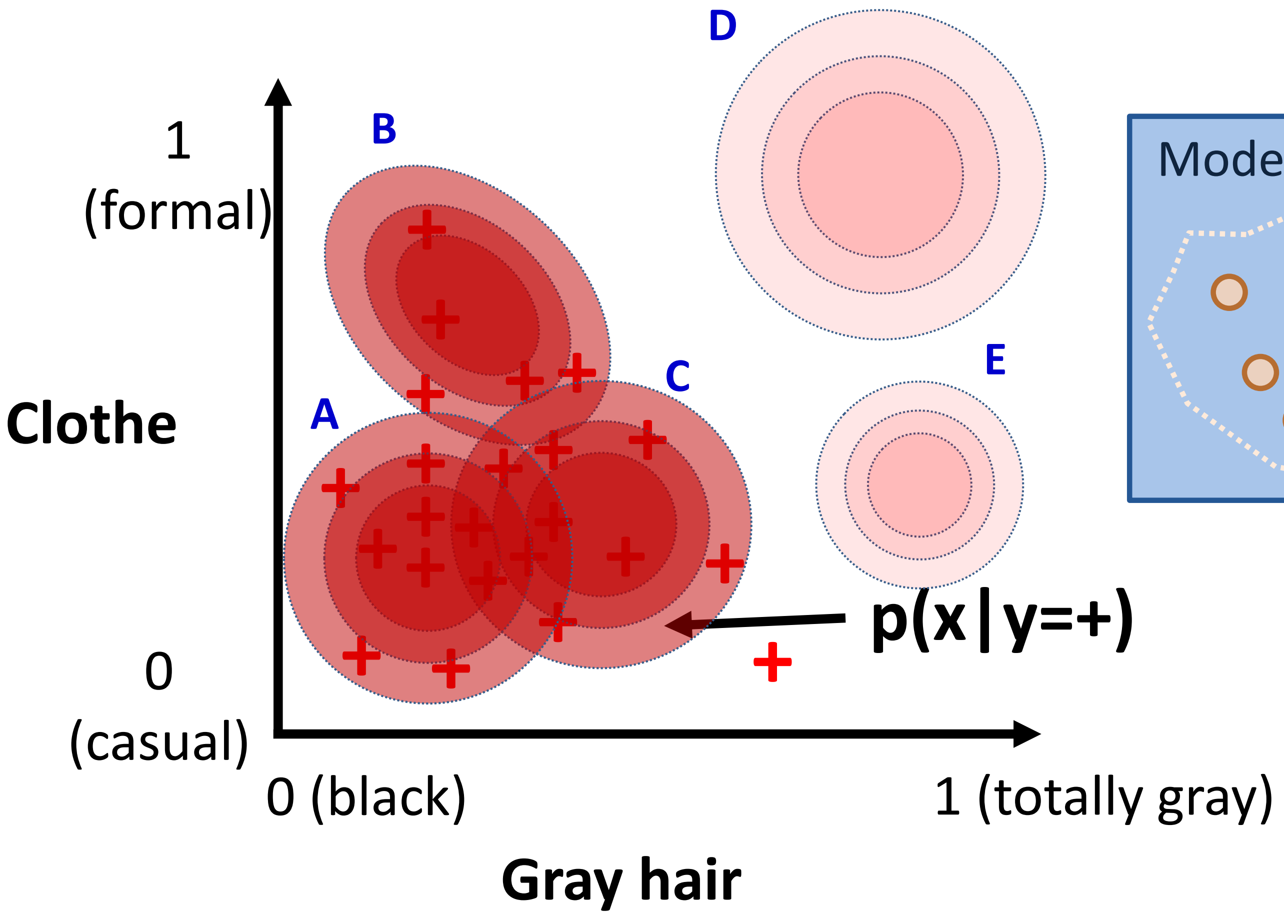


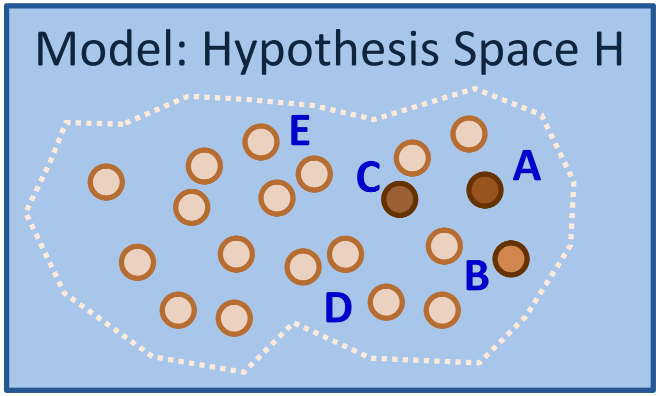
$p(x|h) = f_N(x|\mu, \sigma^2)$

Hope that p(x|y=+) ≈ p(x|h)p(h|y=+)

How can we model the data distribution?

Model p(x|y=+): e.g., Mixture of Gaussian

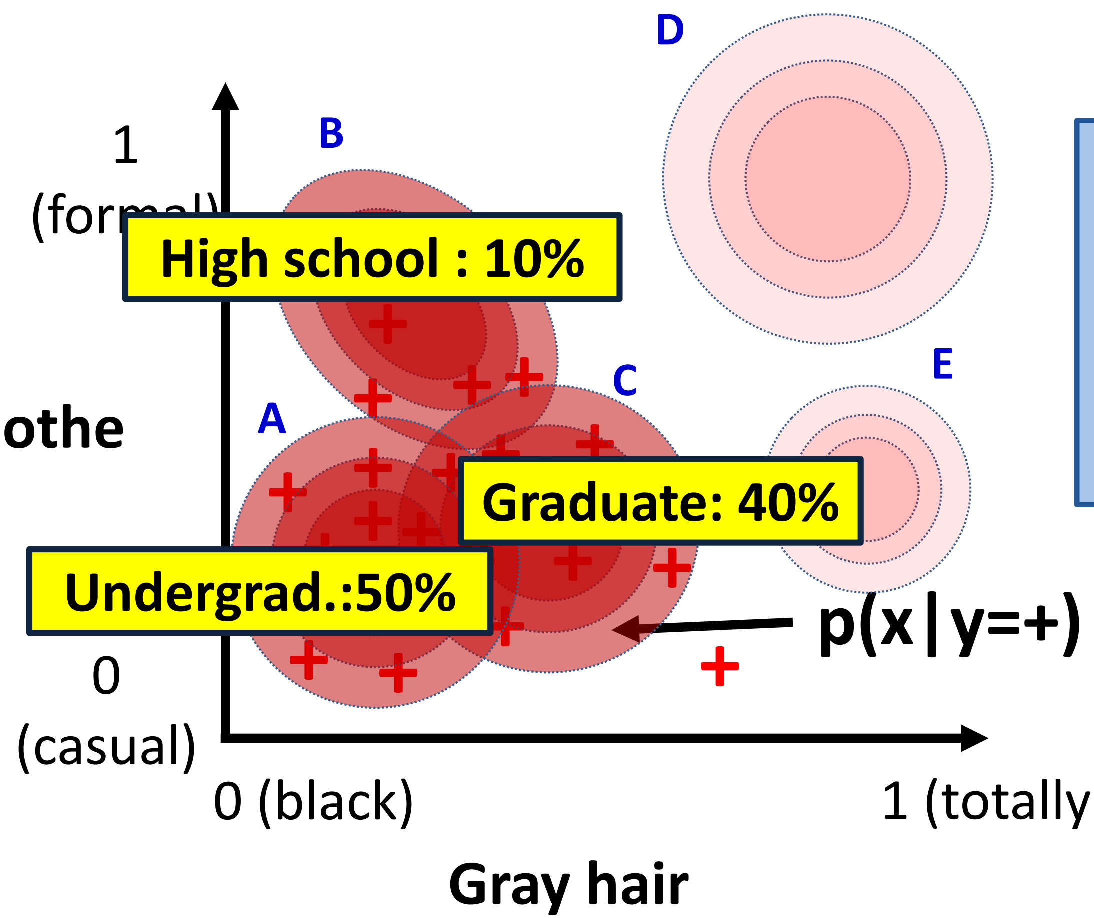


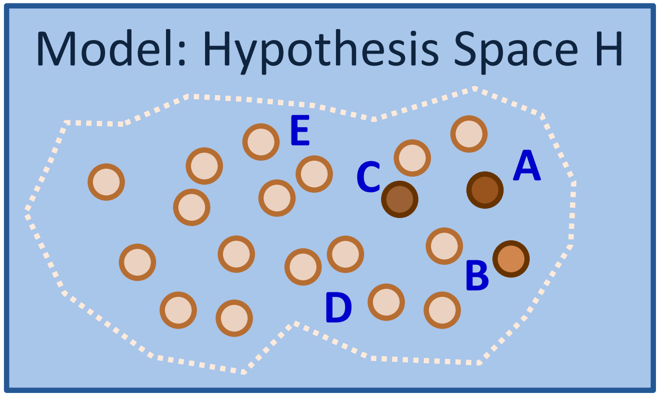


Hope that $p(x|y=+) \approx p(x|h)p(h|y=+)$

How can we model the data distribution?

Model p(x|y=+): e.g., Mixture of Gaussian

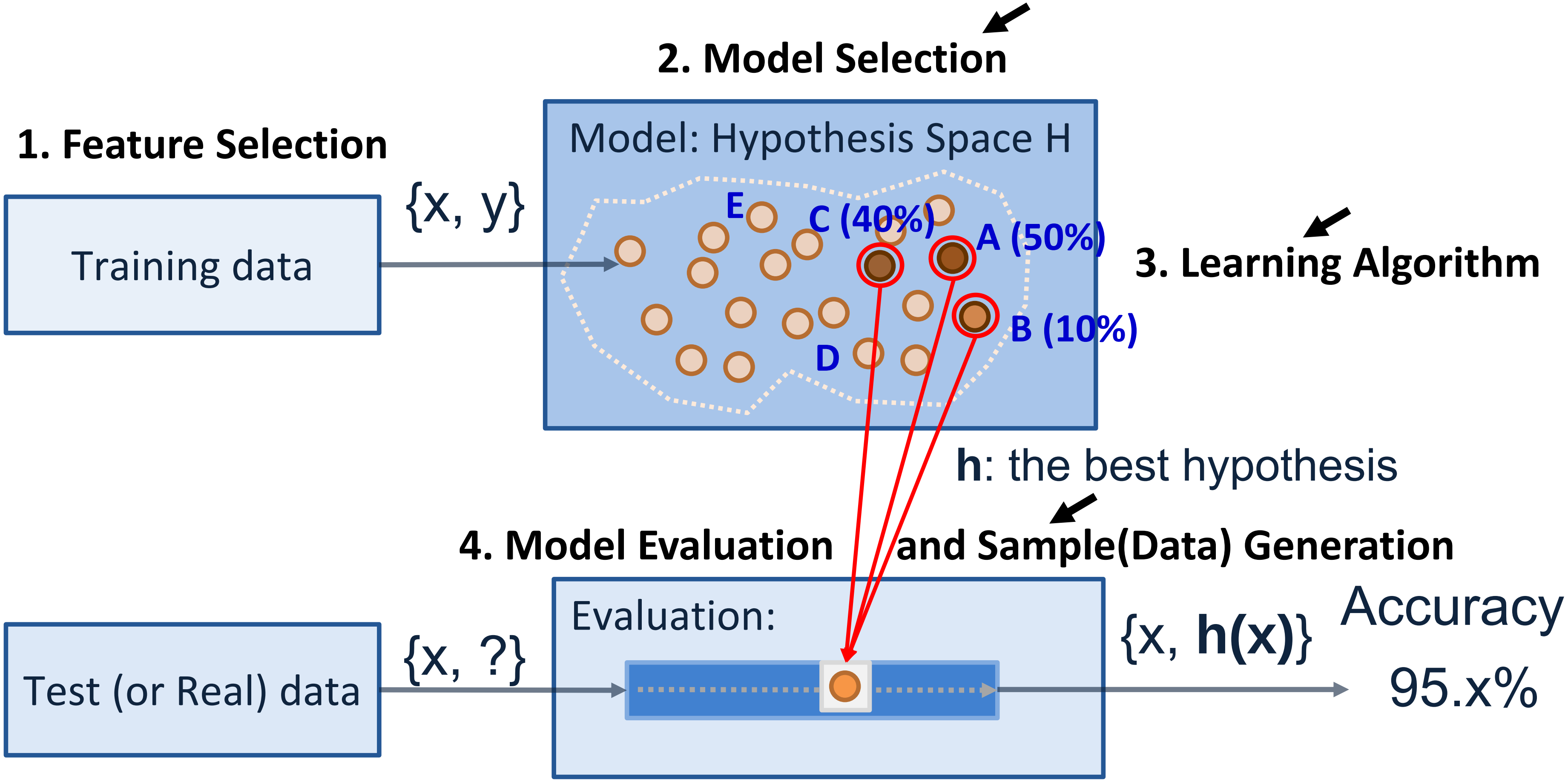




Hope that $p(x|y=+) \approx p(x|h)p(h|y=+)$

Machine Learning: Big Picture

Generative Model



Well, are we done now?
What if, x includes many features beyond gray hair & clothe?

Contents

- Machine Learning Revisited
- Bayesian Learning
- **Graphical Models and Inference Algorithms**
- Lifted Graphical Models and Inference
- Relational Kalman Filtering
- Appendix: Kaggle Competition

How can we represent p(X), if X includes many features?

- A naïve way is to represent the full joint probability.

$$p(x_1, x_2, \ldots, x_n)$$

- For discrete variables (e.g., n binary variables)
  - We need a probability distribution table with 2^n entries.

- For continuous variables (e.g., n-variate Gaussian)
  - We need a covariance matrix with n^2 entries

How can we represent p(X), if X includes many features?

- A naïve way is to represent the full joint probability.

$$p(x_1, x_2, \ldots, x_n)$$

- For discrete variables (e.g., n binary variables)
  - We need a probability distribution table with 2^n entries.
- For continuous variables (e.g., n-variate Gaussian)
  - We need a covariance matrix with n^2 entries

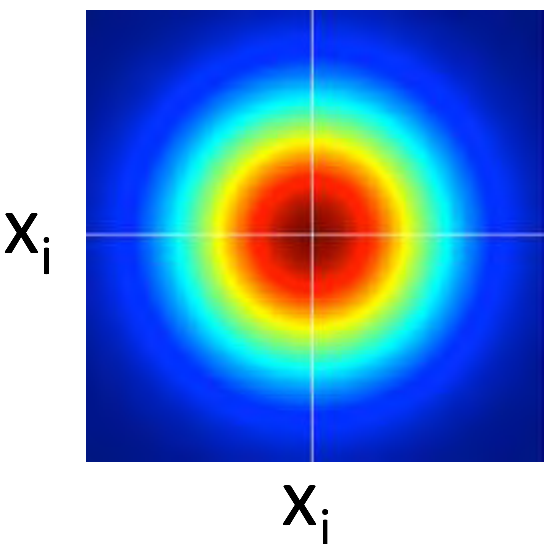
Problem: (1) It's hard to learn the entries; (2) they may be independent each other.

Independent Random Variables

- Independence of variables

$$p(x_i,x_j)=p(x_i)p(x_j)$$

- E.g., Bivariate Gaussian, $p(x_i,x_j)=f_N(x_i,x_{j;}\mu,\Sigma)$
- Which two variables are independent?



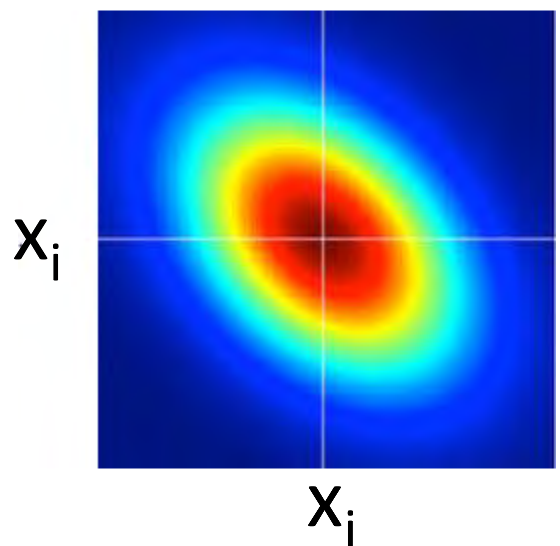


Image sources: stats.stackexchange.com

Factor Graphs

- The probability of a joint assignment of values x to the set of variables X is computed as:

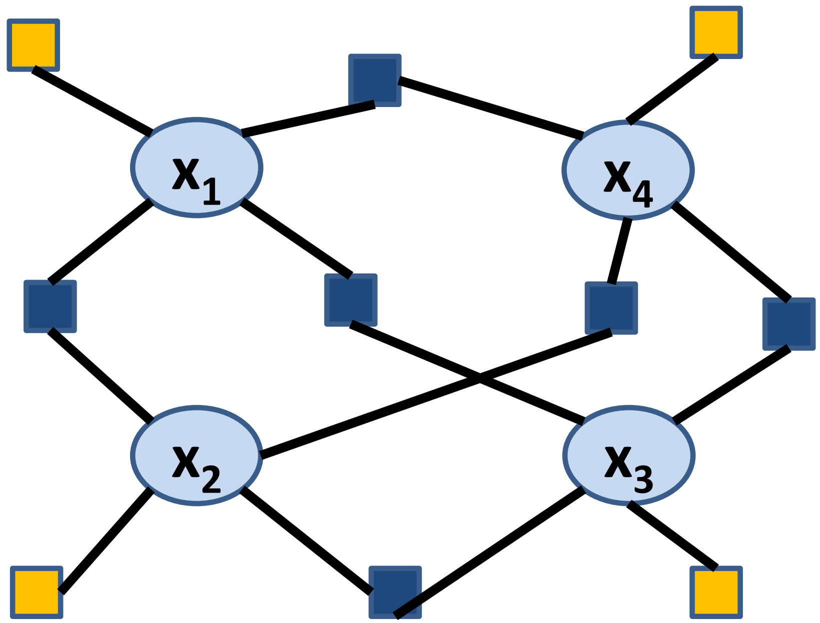
$$p(X = x) = \frac{\prod_{f \in Factors} f(x\{f\})}{Z}$$

Normalizing constant
(or Partition function)

Subset of variables that
participate in the
computation of f

[slide courtesy of Lise Getoor]

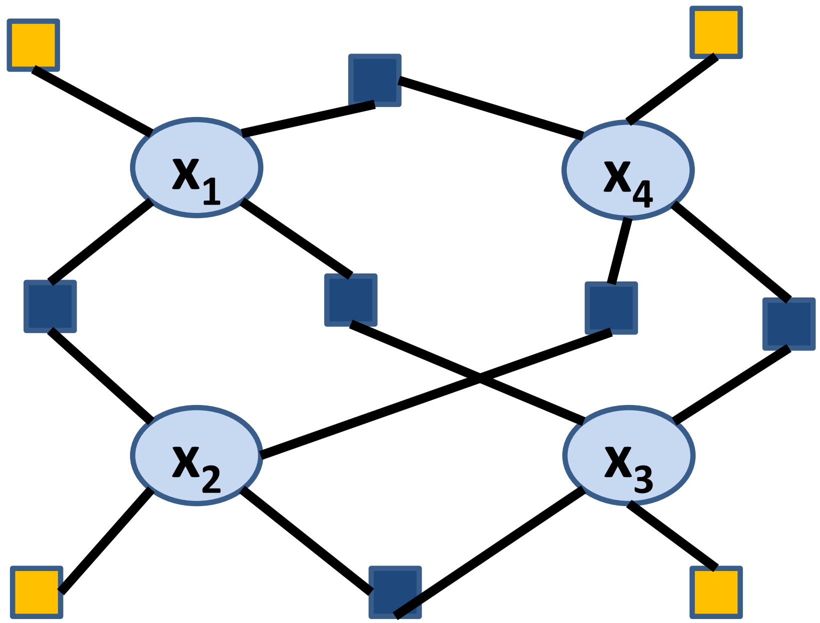
Factor Graphs: Log-Linear Rep



- Each  represents
  $\exp(\theta_L, f_i(x_i))$
- Each  represents
  $\exp(\theta_G, f_{i,j}(x_i, x_j))$

[slide courtesy of Lise Getoor]

Factor Graphs: Log-Linear Rep



For example, in the Ising Model the possible assignments are {-1,+1} and one has

$$\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$$

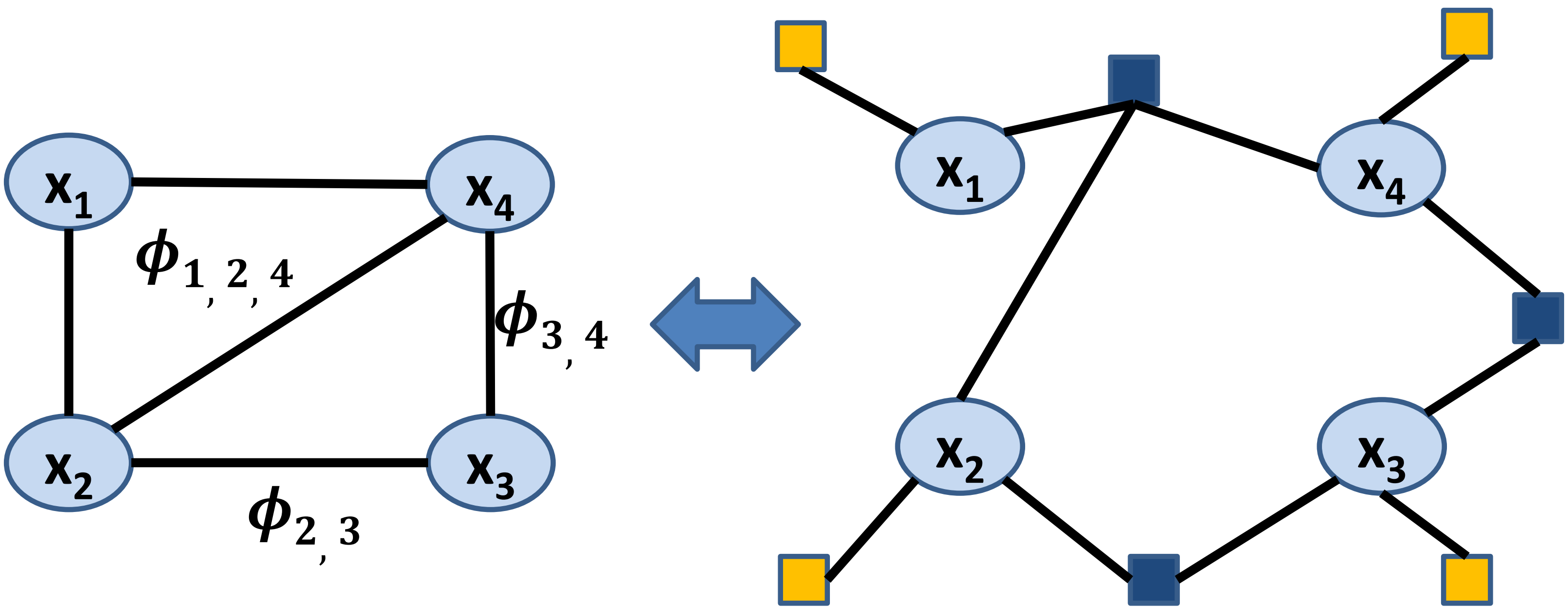
Positive $\theta_{i,j}$ encourages neighboring nodes to have the same assignment

Negative $\theta_{i,j}$ encourages contrasting assignment

[slide courtesy of Lise Getoor]

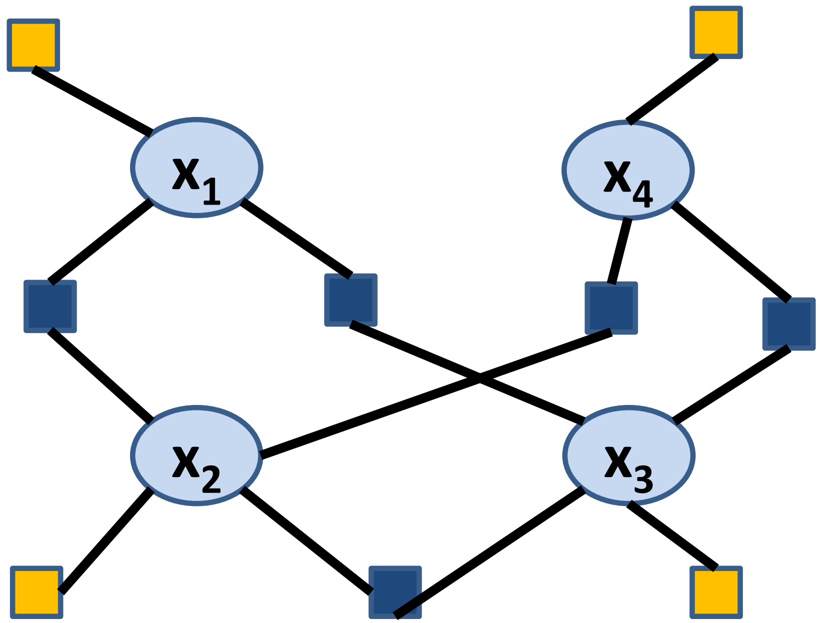
Markov Networks

- Markov networks (aka Markov random fields) can be viewed as a special cases of factor graphs.



[slide courtesy of Lise Getoor]

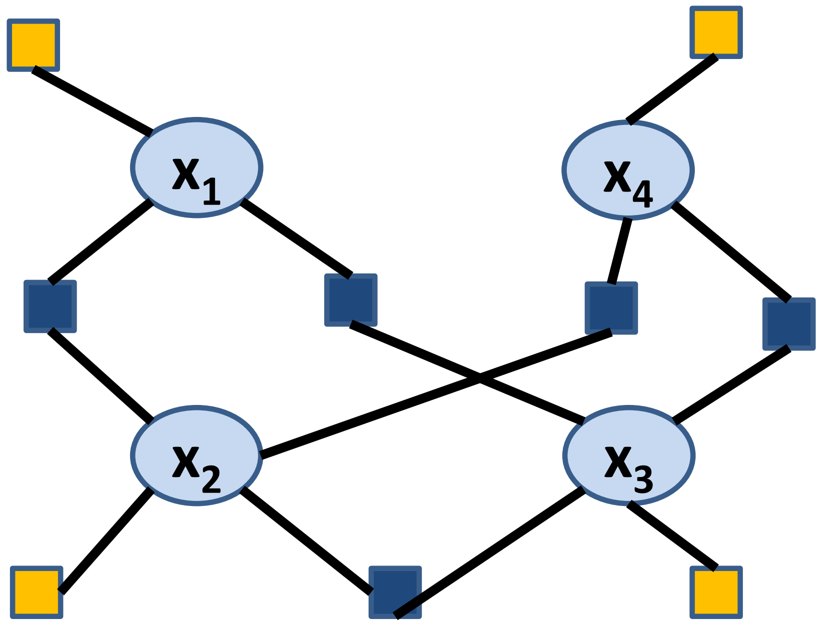
How can we Inference with Factor Graphs?



e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_1, x_2, x_3, x_4) = \phi_1\ \phi_2\ \phi_3\ \phi_4\ \phi_{1,2}\ \phi_{1,3}\ \phi_{2,3}\ \phi_{2,4}\ \phi_{3,4}$$

How can we Inference with Factor Graphs?

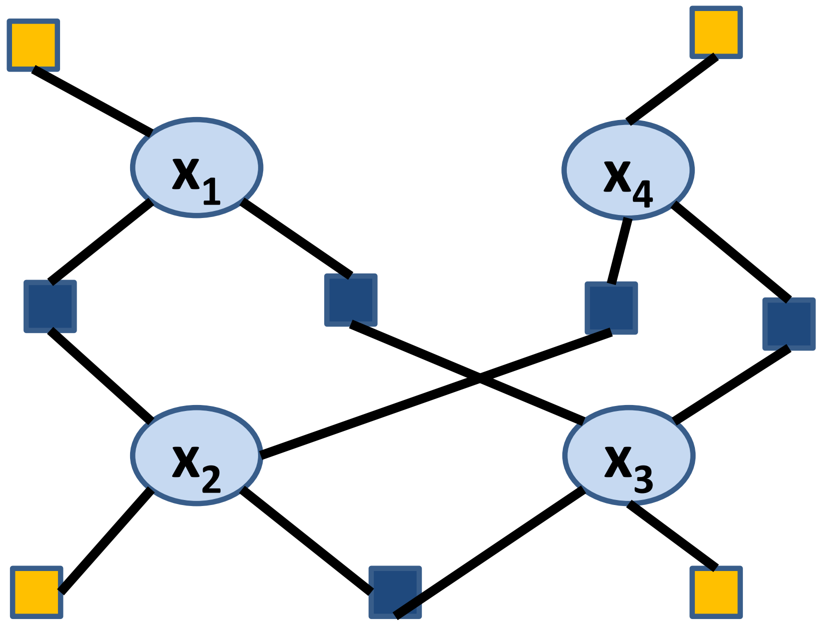


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_1, x_2, x_3, x_4) = \phi_1\ \phi_2\ \phi_3\ \phi_4\ \phi_{1,2}\ \phi_{1,3}\ \phi_{2,3}\ \phi_{2,4}\ \phi_{3,4}$$

$$p(x_4) = \sum_{x_1, x_2, x_3} p(x_1, x_2, x_3, x_4) = \sum_{x_3} \sum_{x_2} \sum_{x_1} p(x_1, x_2, x_3, x_4)$$

How can we Inference with Factor Graphs?

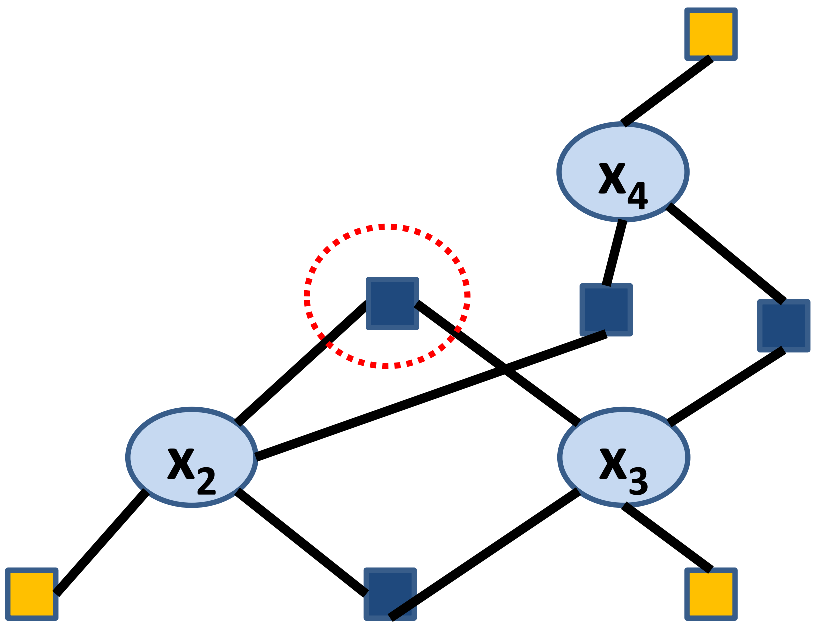


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3}\sum_{x_2}\sum_{x_1} p(x_1, x_2, x_3, x_4)$$

$$= \sum_{x_3}\sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4} \sum_{x_1} \phi_1\, \phi_{1,2}\, \phi_{1,3}$$

How can we Inference with Factor Graphs?

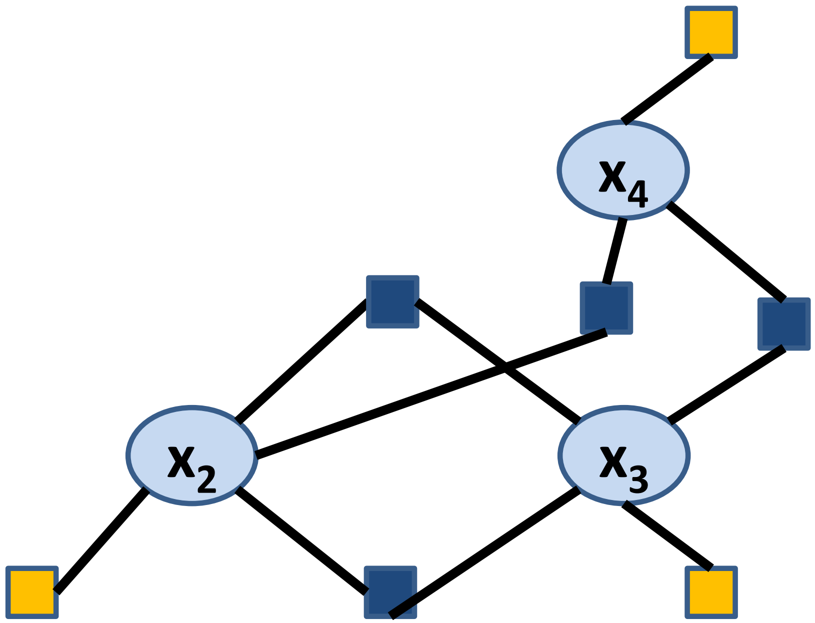


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3}\sum_{x_2}\sum_{x_1} p(x_1, x_2, x_3, x_4)$$

$$= \sum_{x_3}\sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4}\, \phi'_{2,3}$$

How can we Inference with Factor Graphs?

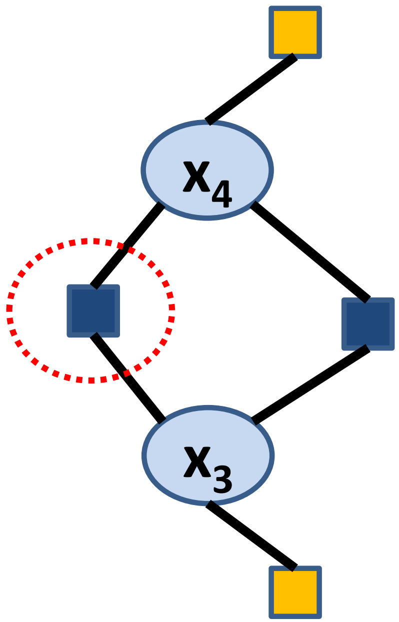


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3} \sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4}\, \phi'_{2,3}$$

$$= \sum_{x_3} \phi_3 \phi_4 \phi_{3,4} \sum_{x_2} \phi_2\, \phi_{2,3}\, \phi_{2,4}\, \phi'_{2,3}$$

How can we Inference with Factor Graphs?

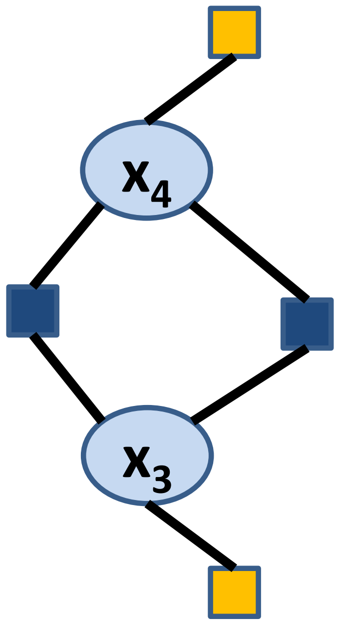


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3} \sum_{x_2} \phi_2 \, \phi_3 \, \phi_4 \phi_{2,3} \, \phi_{2,4} \, \phi_{3,4} \, \phi'_{2,3}$$

$$= \sum_{x_3} \phi_3 \phi_4 \phi_{3,4} \, \phi'_{3,4}$$

How can we Inference with Factor Graphs?

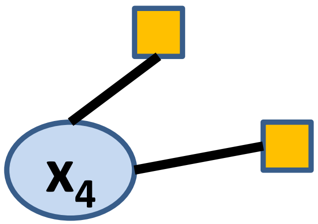


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3} \sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4}\, \phi'_{2,3}$$

$$= \sum_{x_3} \phi_3 \phi_4 \phi_{3,4}\ \phi'_{3,4} = \phi_4 \sum_{x_3} \phi_3 \phi_{3,4}\ \phi'_{3,4}$$

How can we Inference with Factor Graphs?

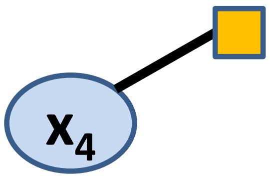


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3} \sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4}\, \phi'_{2,3}$$

$$= \sum_{x_3} \phi_3 \phi_4 \phi_{3,4}\ \phi'_{3,4} = \phi_4 \phi'_4$$

How can we Inference with Factor Graphs?

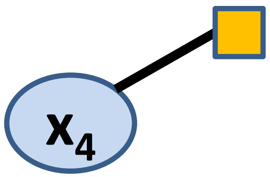


e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

$$p(x_4) = \sum_{x_3} \sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4}\, \phi'_{2,3}$$

$$= \sum_{x_3} \phi_3 \phi_4 \phi_{3,4}\ \phi'_{3,4} = \phi_4 \phi'_4 = \phi''_4$$

How can we Inference with Factor Graphs?



e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

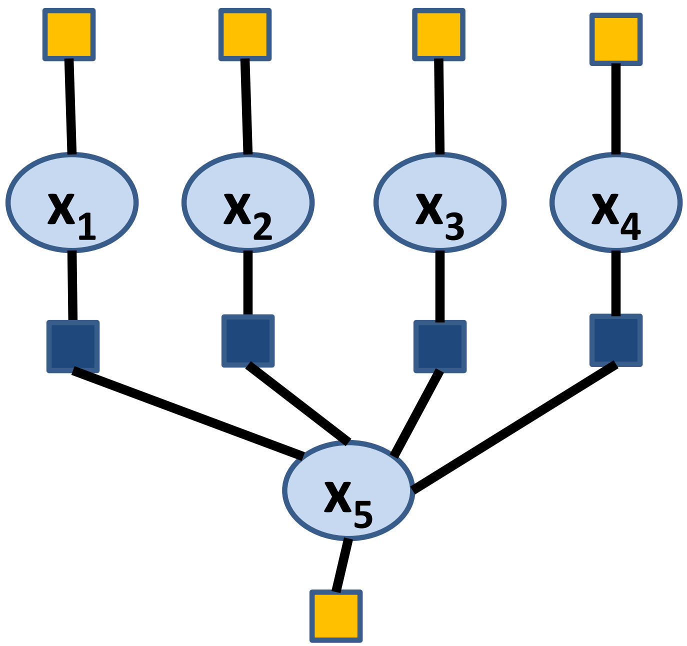
We summed (eliminated) out variables.
It is called "Variable Elimination"

$$p(x_4) = \sum_{x_3} \sum_{x_2} \phi_2\, \phi_3\, \phi_4 \phi_{2,3}\, \phi_{2,4}\, \phi_{3,4}\, \phi'_{2,3}$$

$$= \sum_{x_3} \phi_3 \phi_4 \phi_{3,4}\ \phi'_{3,4} = \phi_4 \phi'_4 = \phi''_4$$

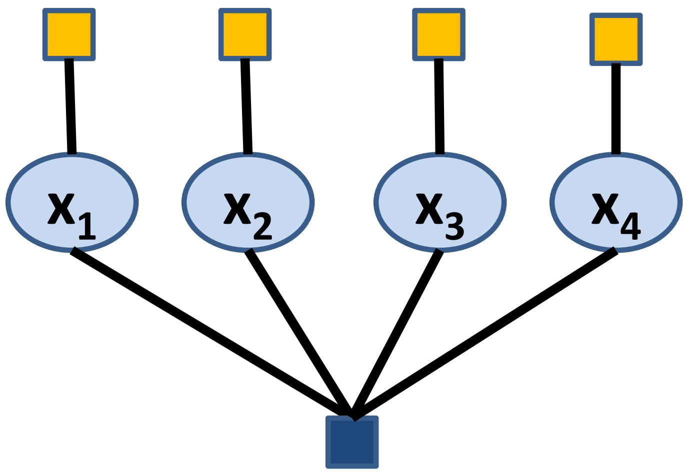
How can we Inference with Factor Graphs? What can happen?

- The order of elimination does matter!
- What if we have a bad elimination order?



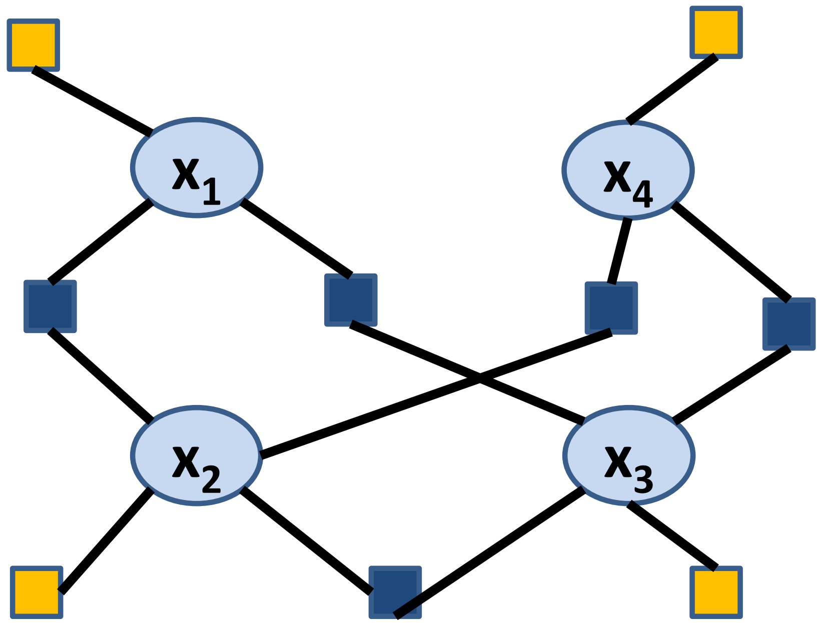
How can we Inference with Factor Graphs? What can happen?

- The order of elimination does matter!
- What if we have a bad elimination order? If we eliminate x5…



The computational complexity is exponential to the treewidth (the size of largest clique in a junction tree)

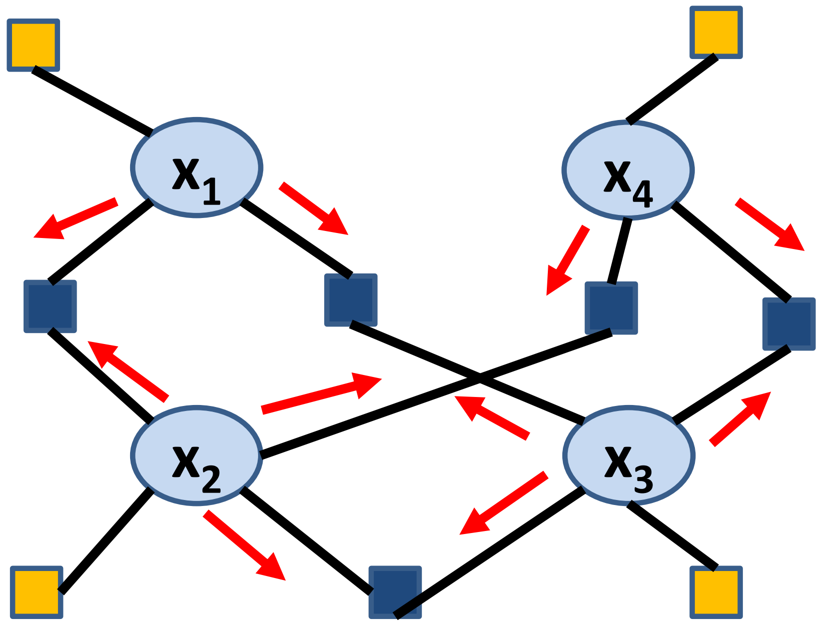
How can we Inference with Factor Graphs? Any alternative? Belief propagation



e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

Intuition: Let's exchange our opinions!

How can we Inference with Factor Graphs? Any alternative? Belief propagation



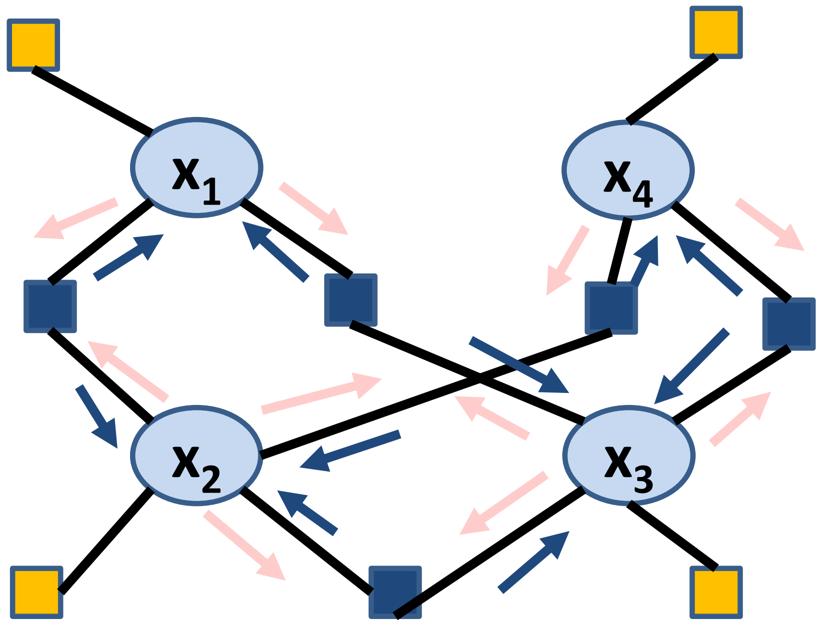
e.g., $\phi_{i,j} = \exp(\theta_{i,j} x_i x_j)$

Intuition: Let's exchange our opinions!

Time 0: every one has own belief about initial status.

Time 1: Send my belief to my neighbors

How can we Inference with Factor Graphs? Any alternative? Belief propagation



e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

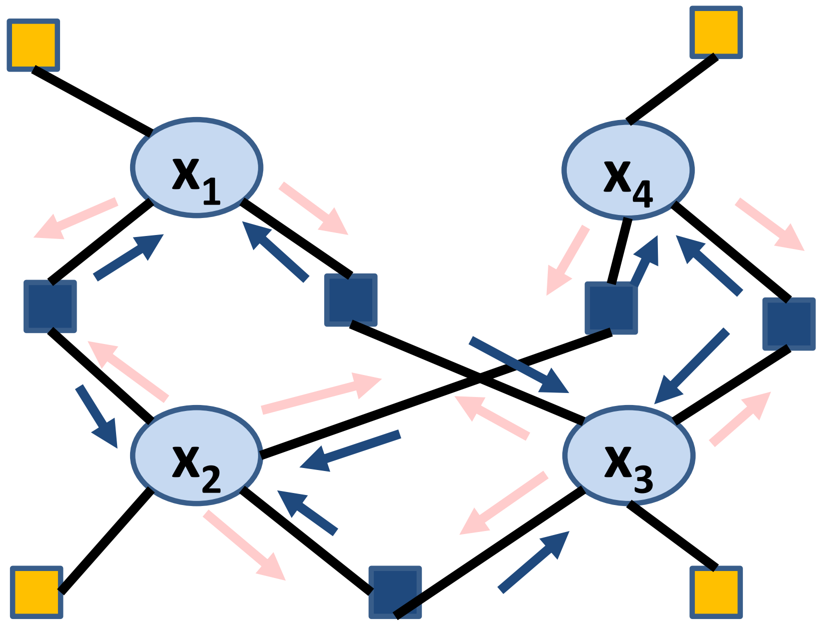
Intuition: Let's exchange our opinions!

Time 0: every one has own belief about initial status.

Time 1: Send my belief to my neighbors.

Time 1.5: Update my belief based on my neighbors.

How can we Inference with Factor Graphs? Any alternative? Belief propagation



e.g., $\phi_{i,j} = \exp(\theta_{i,j} x_i x_j)$

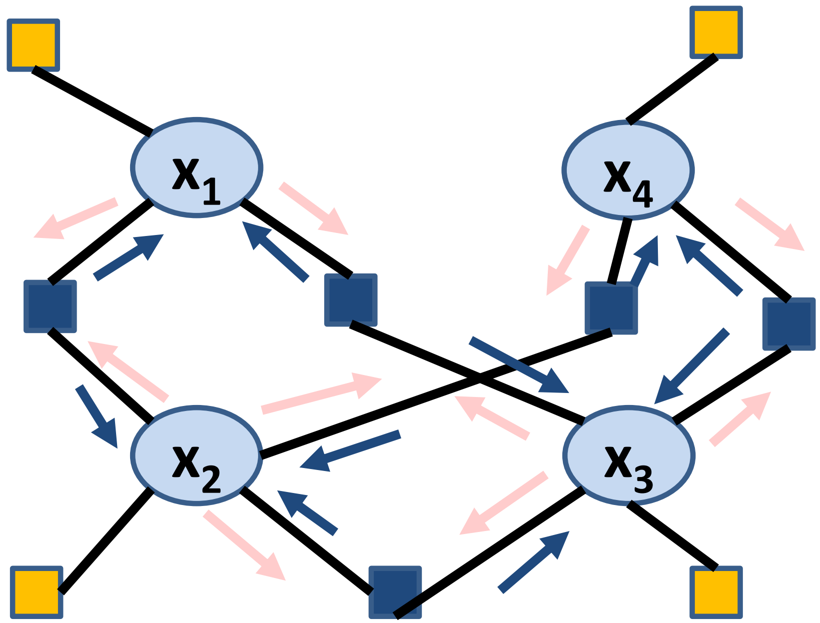
Intuition: Let's exchange our opinions!

Time 0: every one has own belief about initial status.

Time t: Send my belief to my neighbors.

Time t.5: Update my belief based on my neighbors.

How can we Inference with Factor Graphs? Any alternative? Belief propagation



e.g., $\phi_{i,j} = \exp\left(\theta_{i,j} x_i x_j\right)$

We hope that it converges to a close-to-optimal value.

Time t: Send my belief to my neighbors.

Time t.5: Update my belief based on my neighbors.

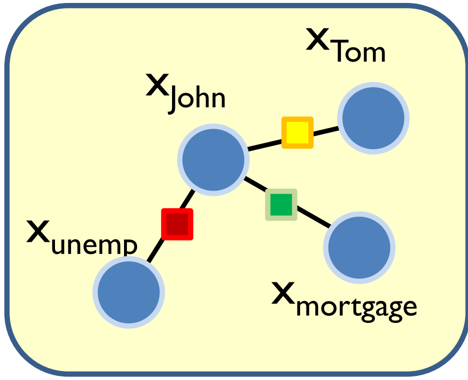
Contents

- Machine Learning Revisited
- Bayesian Learning
- Graphical Models and Inference Algorithms
- **Lifted Graphical Models and Inference**
- Relational Kalman Filtering
- Appendix: Kaggle Competition

How to Estimate Future Events with Graphical Models?

- Choose a graphical model: e.g.,
  - Bayesian Networks,
  - Markov Random Fields,
  - Kalman Filter

- Collect observations:
  - Tom sold his home at $0.5 million.
  - The mortgage rates increased(↑) to average 5.5%.
  - The unemployment rate downed(↓) to 7%.



- Compute conditional probabilities by relationships:

P(value of John's home | observations)

Estimating Future Events with Large-Scale Graphical Models

- Estimating future events is essential in
  - Financial markets (housing)
  - Environment (extreme weather, groundwater)
  - Energy (smart grid)

housing

weather

energy

Challenges: Large-Scale Models

- Hard to handle large numbers of elements
  - US housing market: 75.56 million house units
  - Hurricane Sandy: spanning 1,100 miles (1,800 km)

- Computational Complexities
  - Kalman filter:

$$O(n^3) = O(75.56^3 \cdot 10^6 \text{ trillion})$$

  - Dynamic Bayesian Networks and Markov Random Field:

$$O(\exp^n) = O\left(\exp^{75.56 \text{ million}}\right)$$

Some Elements Share Relationships

- Elements share Relationships
  - If mortgage ↑ 1% → price of Tom's home ↓ 3%
  - If mortgage ↑ 1% → price of John's home ↓ 3%
  - If mortgage ↑ 1% → price of any home in the town ↓ 3%

- Relations over clusters
  - Town = {Tom, John, … }
  - Δ(price of *name*'s home)

    $$= -3\Delta\text{Mortgage} + \varepsilon$$
  - *name* ∈ Town
  - $\varepsilon \sim N(0, \Sigma)$, Gaussian Noise

Relational Graphical Models
or Statistical Relational Learning (SRL)

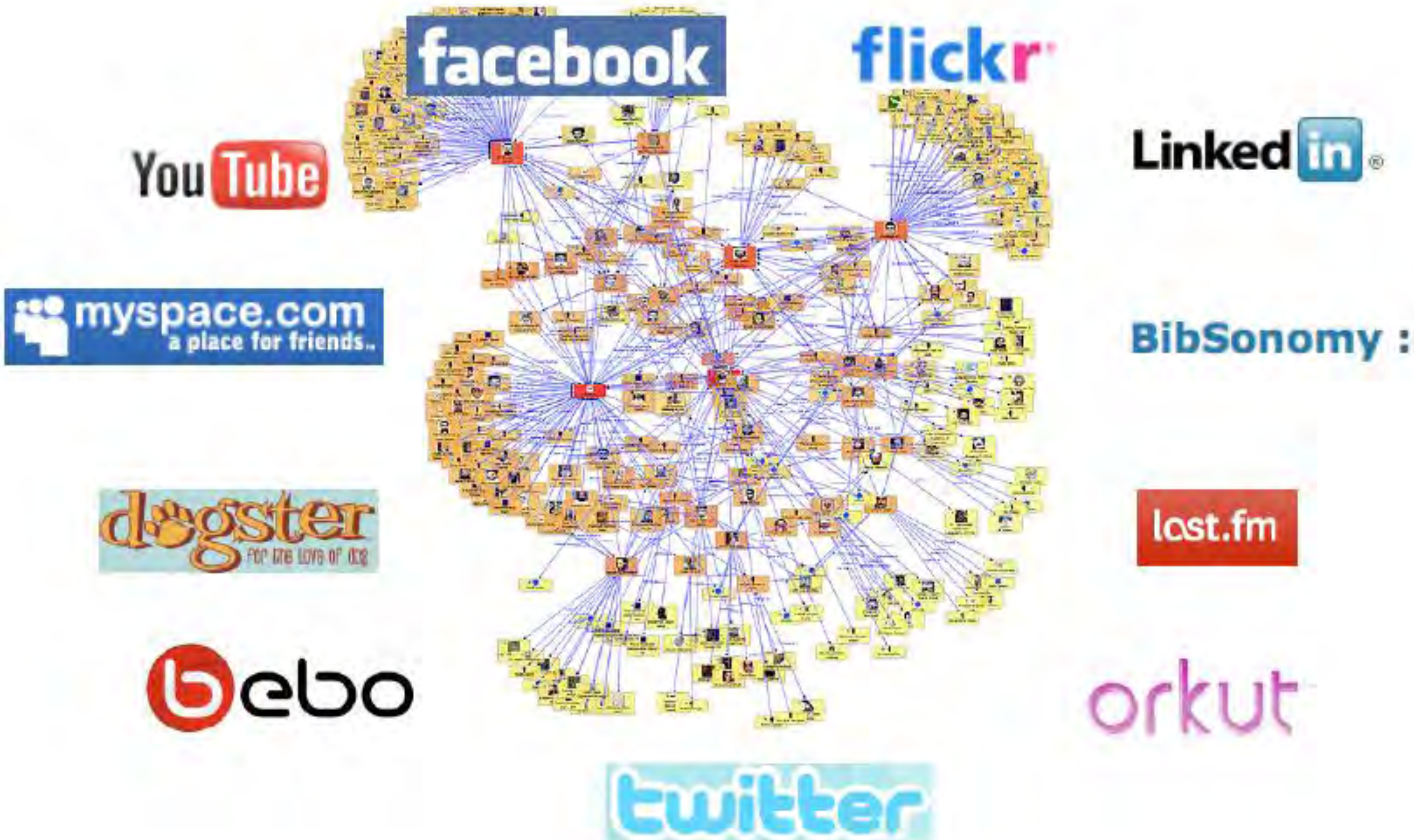
- Traditional statistical machine learning approaches assume
  - A random sample of homogeneous objects from single relation
  - Independent, identically distributed (IID)

- Traditional relational machine learning approaches assume:
  - Logical language for describing structure in sample
  - No noise and no uncertainty

- Real world data sets:
  - Multi-relational and heterogeneous
  - Noisy and uncertainty

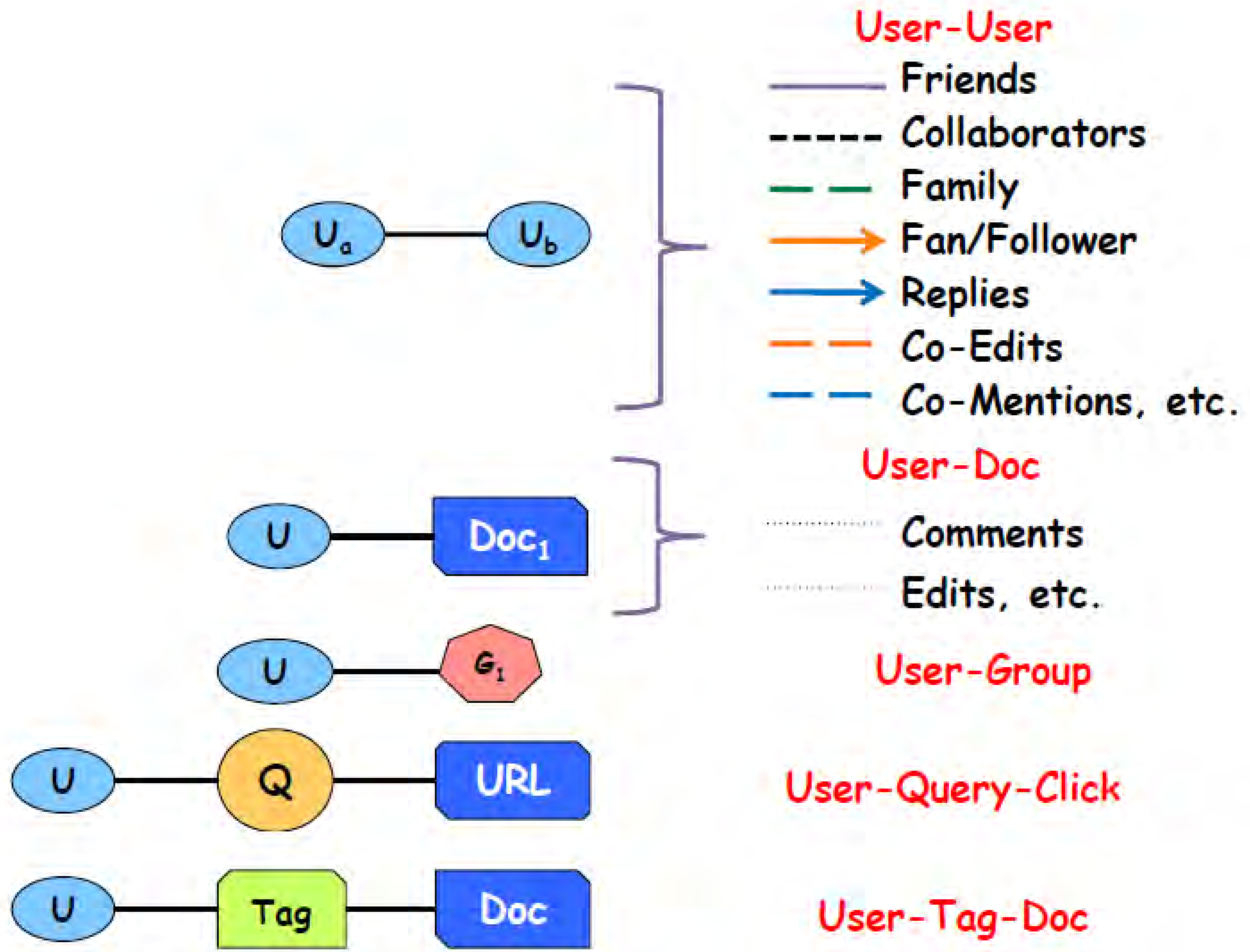
[slide courtesy of Lise Getoor]

Example: Social Media



[slide courtesy of Lise Geetor]

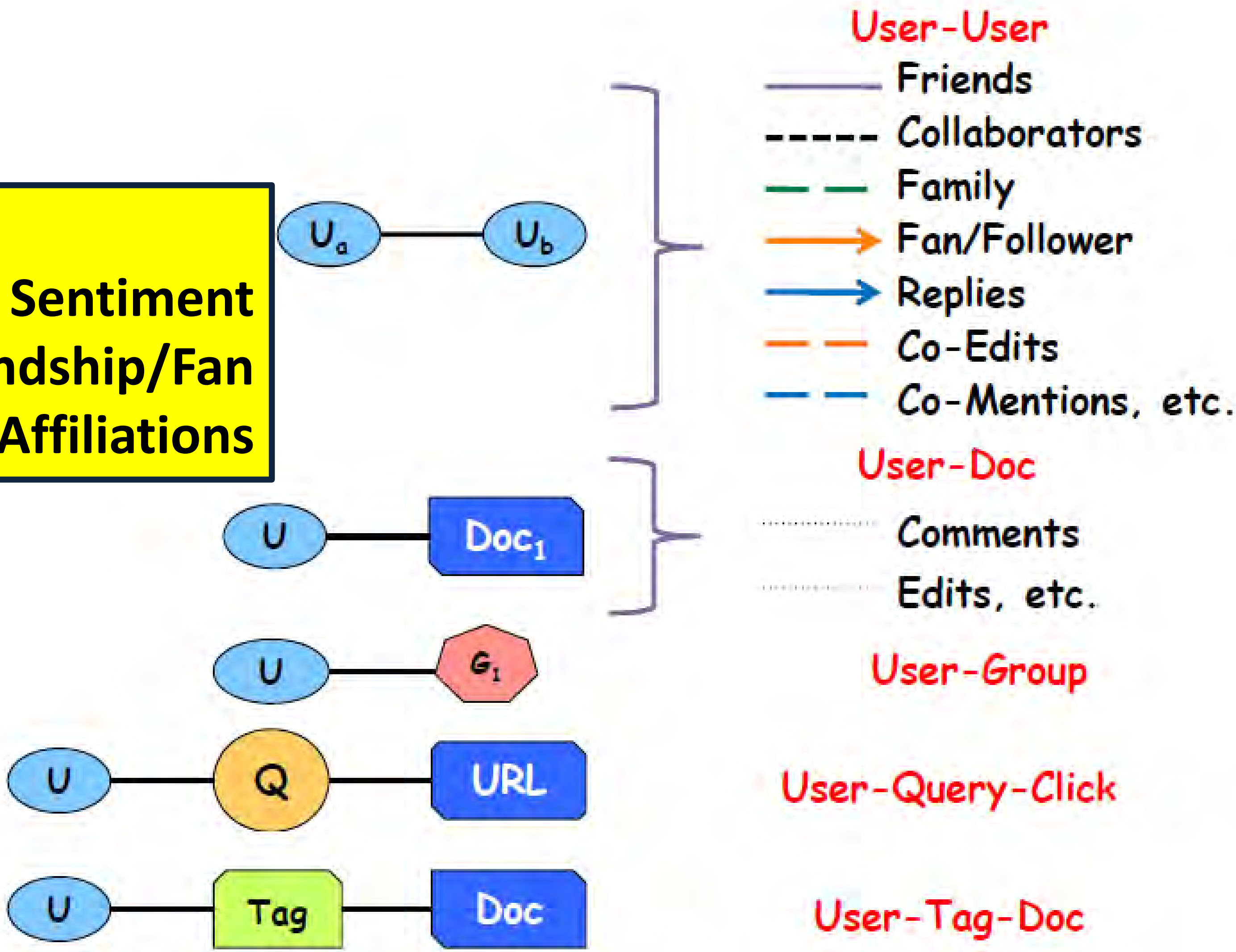
Social Media Relationships



[slide courtesy of Lise Getoor]

Social Media Relationships

Predict:
Sentiment
Friendship/Fan
Affiliations

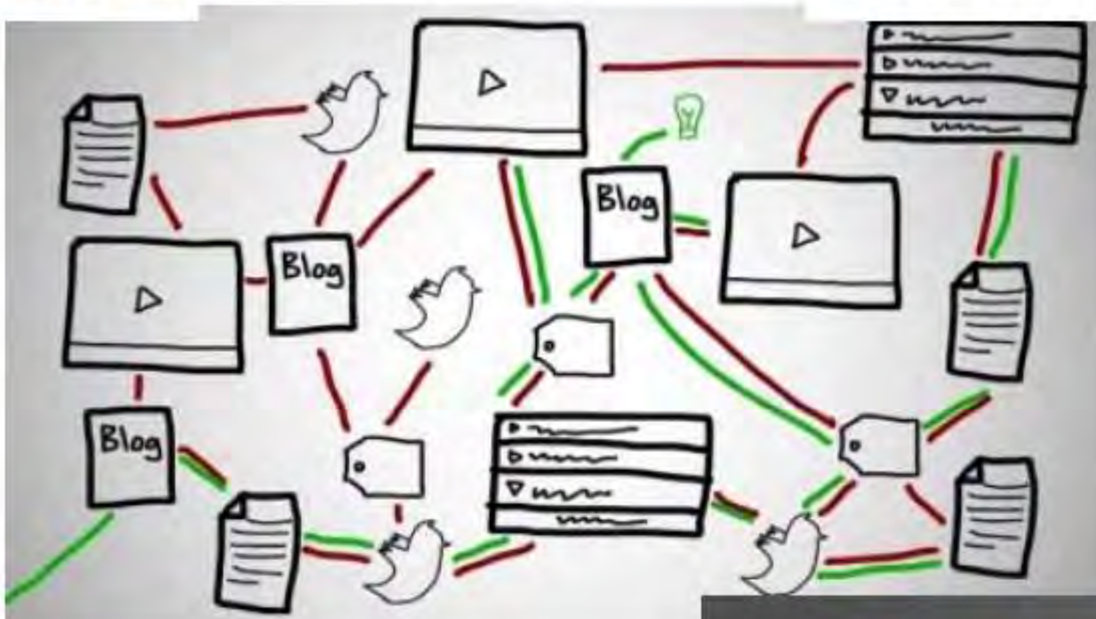


[slide courtesy of Lise Getoor]

Massively Open Online Courses (MOOCs)

coursera

Codecademy

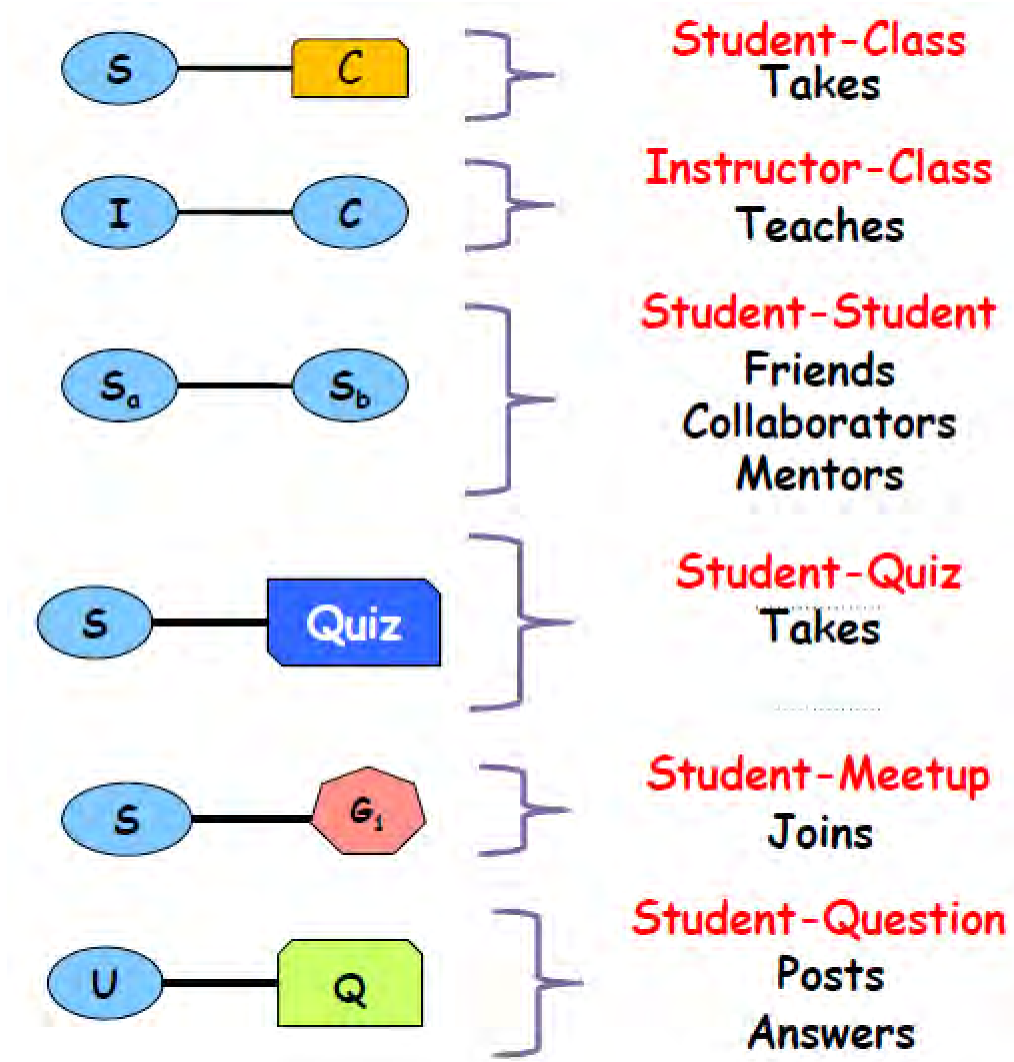


edX

UDACITY

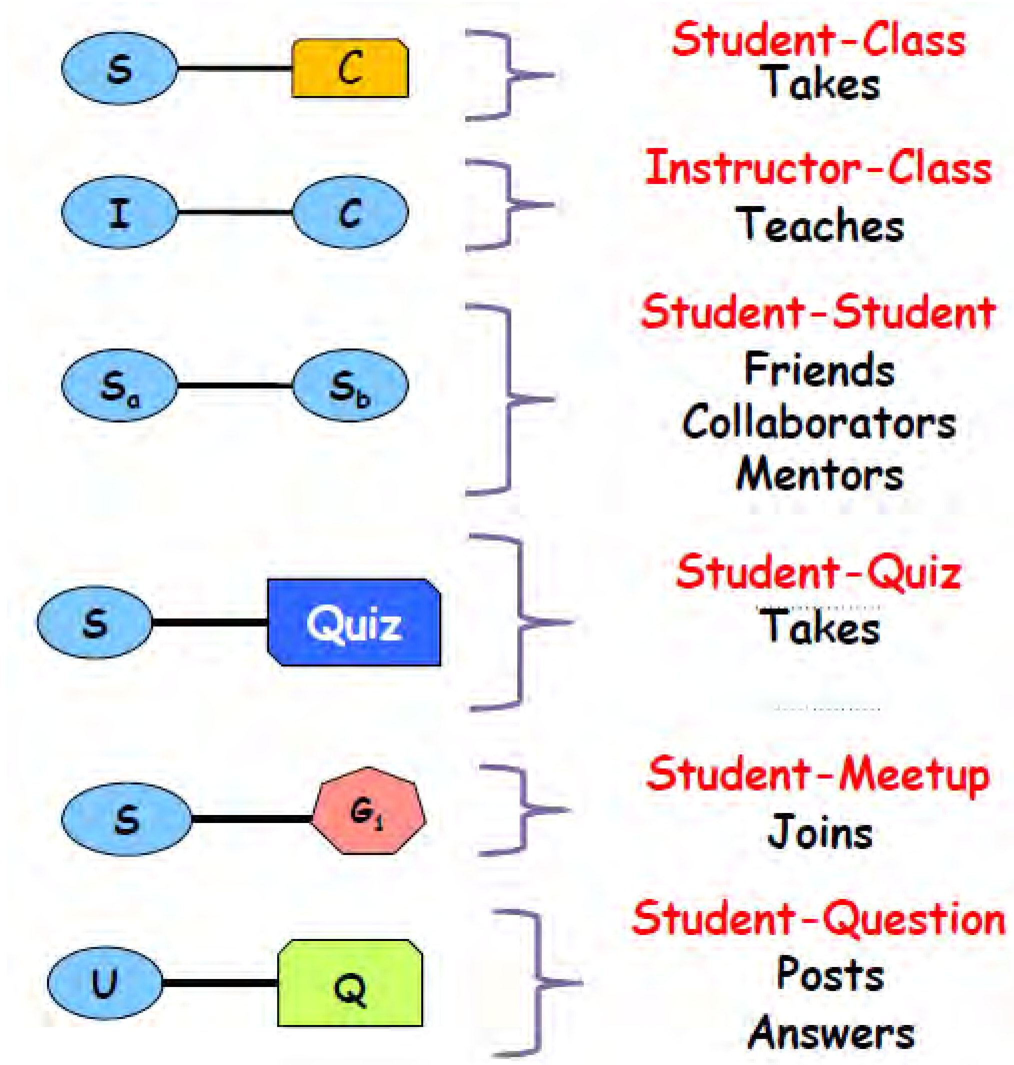
[slide courtesy of Lise Getoor]

MOOC Relationships



[slide courtesy of Lise Getoor]

MOOC Relationships



Predict:

Performance

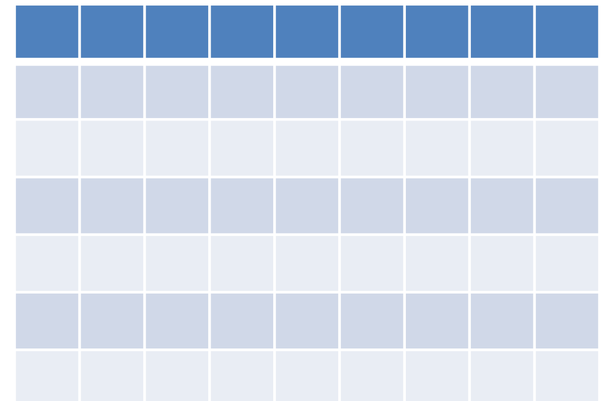
Role

Participation

[slide courtesy of Lise Getoor]

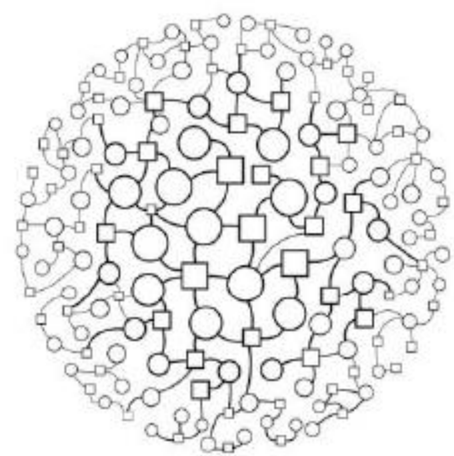
What is the difference?

Most of the data that is available in the newly emerging ear of big data does not look like this



Or even like this

It looks more like this



[slide courtesy of Lise Getoor]

What is Statistical Relational Learning?

- Collection of techniques which combine rich relational knowledge AI/DB representations with statistical models
  - First-order logic, SQL, graphs,
  - Graphical models, directed, undirected, mixed; relational decision trees, etc.
- Example:
  - Markov Logic Networks (Washington and Texas), Bayesian Logic Programs (Berkeley & MIT), Probabilistic Relational Models (Stanford), Factorie (UMass), **Relational Kalman Filtering (U of Illinois & UNIST)**, and …
- Key ideas
  - Relational feature construction
  - Collective reasoning
  - 'Lifted' representation, inference and learning

[modified from a slide courtesy of Lise Getoor]

Lifted Inference
(or First-Order Probabilistic Inference)

Input:

Compact: **Relational Models**

Ground

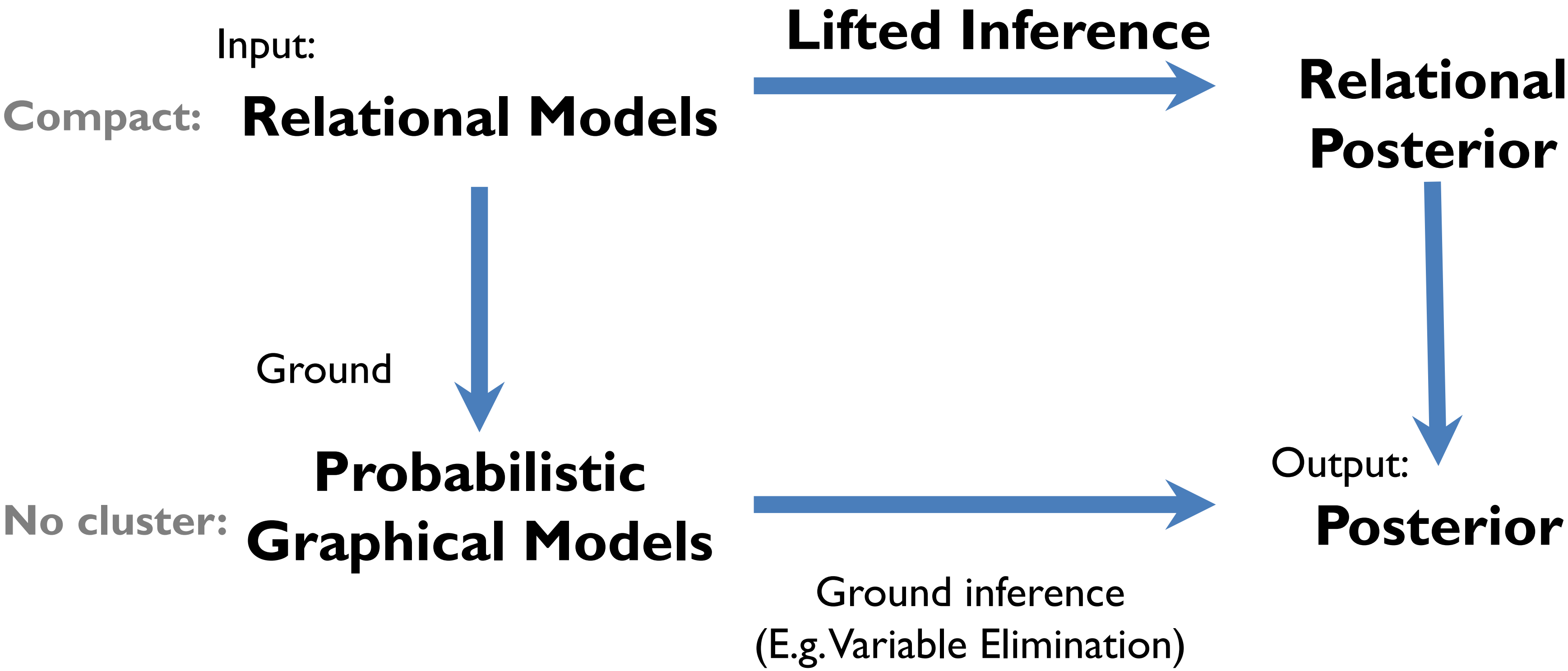
No cluster: **Probabilistic Graphical Models**

Ground inference
(E.g. Variable Elimination)

Output:
Posterior

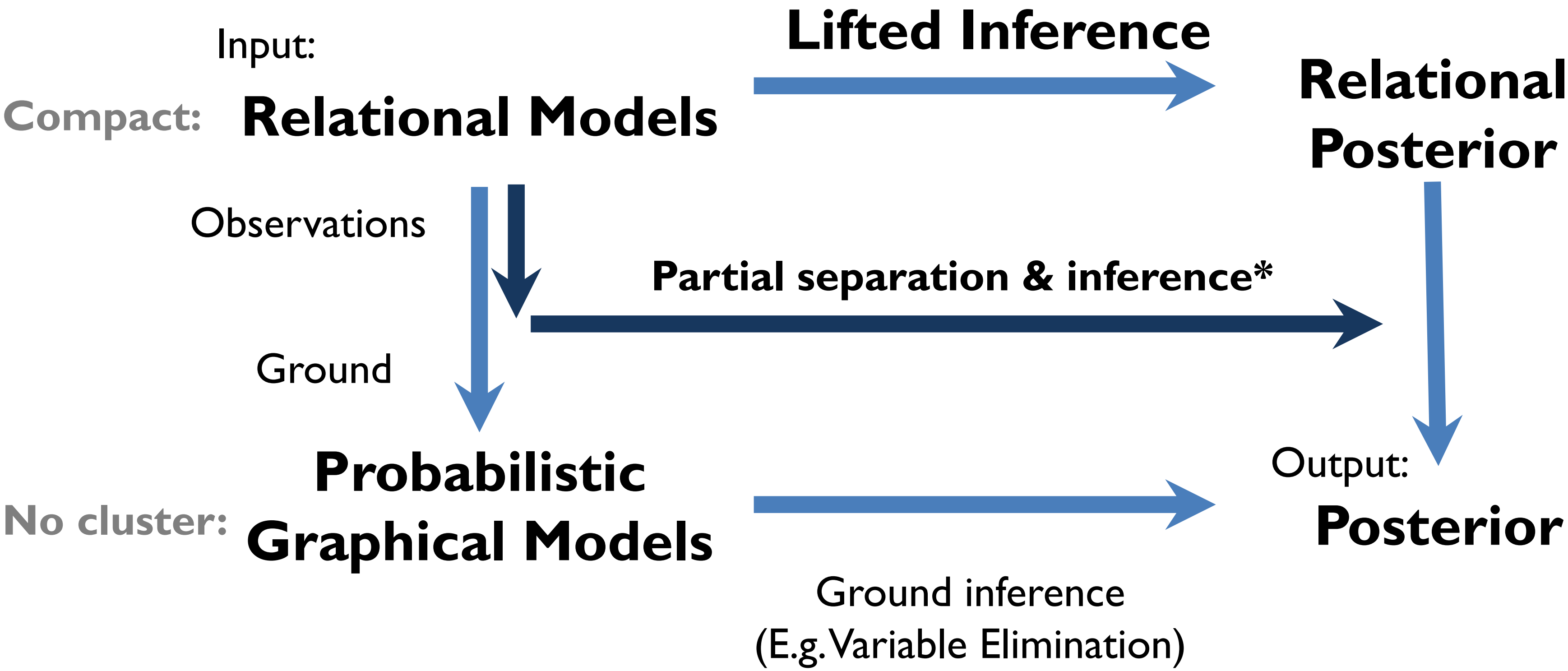
[slide courtesy of David Poole]

Lifted Inference
(or First-Order Probabilistic Inference)



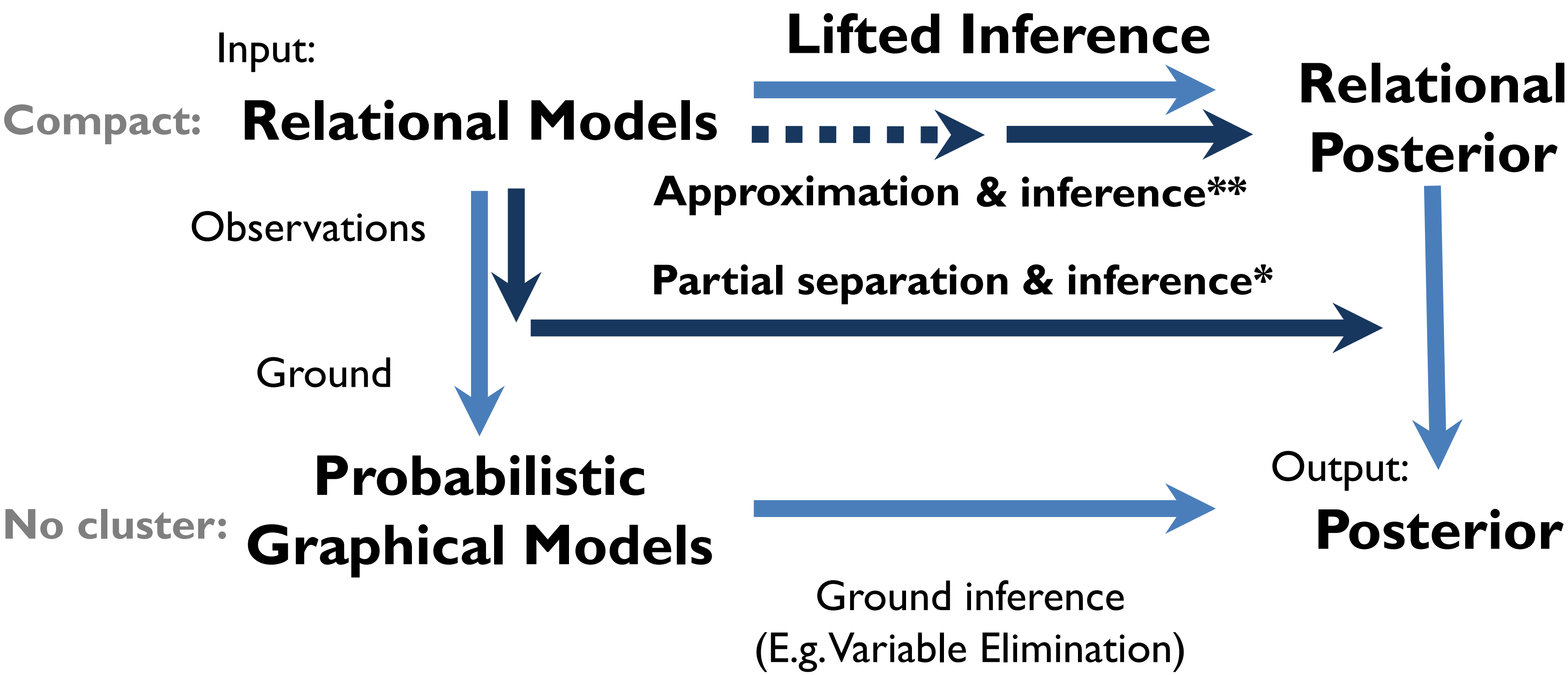
[slide courtesy of David Poole]

Lifted Inference
(or First-Order Probabilistic Inference)



* **Before** $O(n^3) \rightarrow$ **Now** $O(nm^2)$

Lifted Inference
(or First-Order Probabilistic Inference)



Main Ideas and Contributions

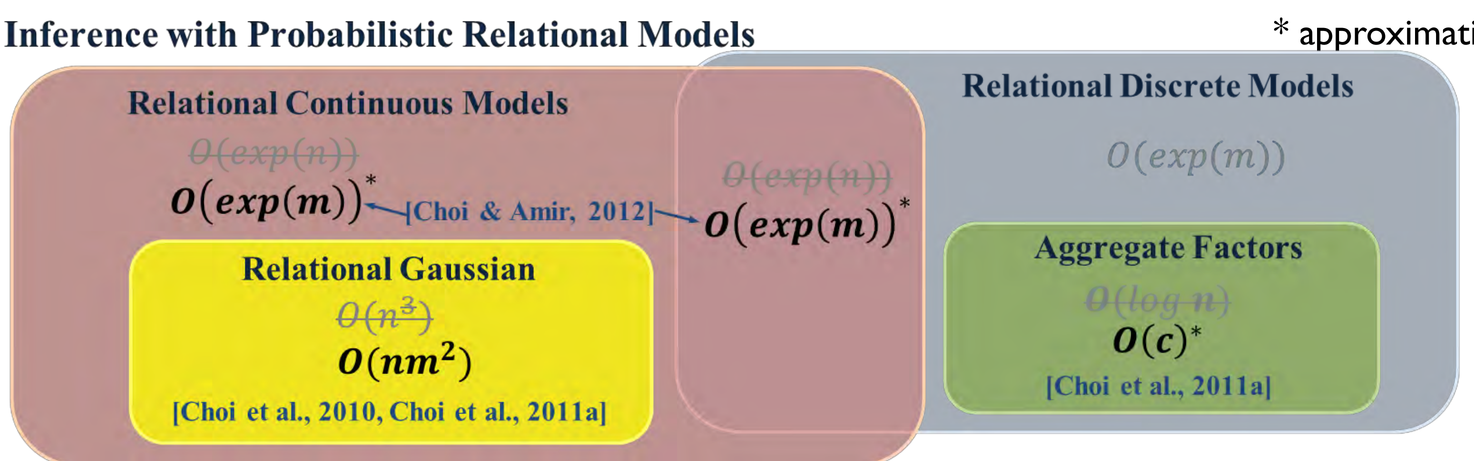
- **Representing and maintaining compact structures over clusters** is feasible, and thus can lead to accurate and efficient estimations of future events.
  - Scalable Kalman filter

$$\text{Before: } \boldsymbol{O(n^3)} \rightarrow \text{Now: } \boldsymbol{O(nm^2)}$$

  - Unified, efficient estimations of discrete-continuous models

$$\text{Before: } \boldsymbol{O(exp(n))} \rightarrow \text{Now: } \boldsymbol{O(exp(m))}$$

* approximation



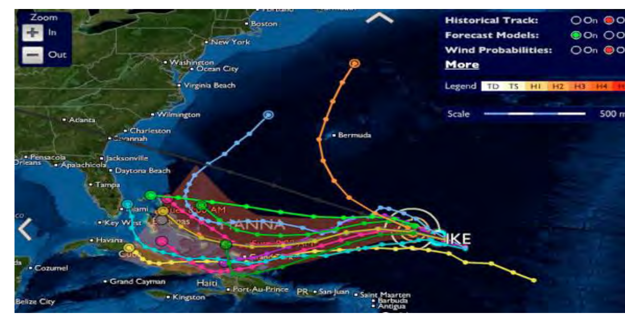
n: # of random variables (RVs), m: # of clusters of RVs & c: constant

Contents

Kalman Filter

- Kalman Filter is an algorithm which produces estimates of unknown variables given a series of measurements (w/ noise) over time.

- Numerous applications in
  - Robot localization
  - Autopilot
  - Econometrics (time series)
  - Military: rocket and missile guidance
  - Weather forecasting
  - Speech enhancement
  - …

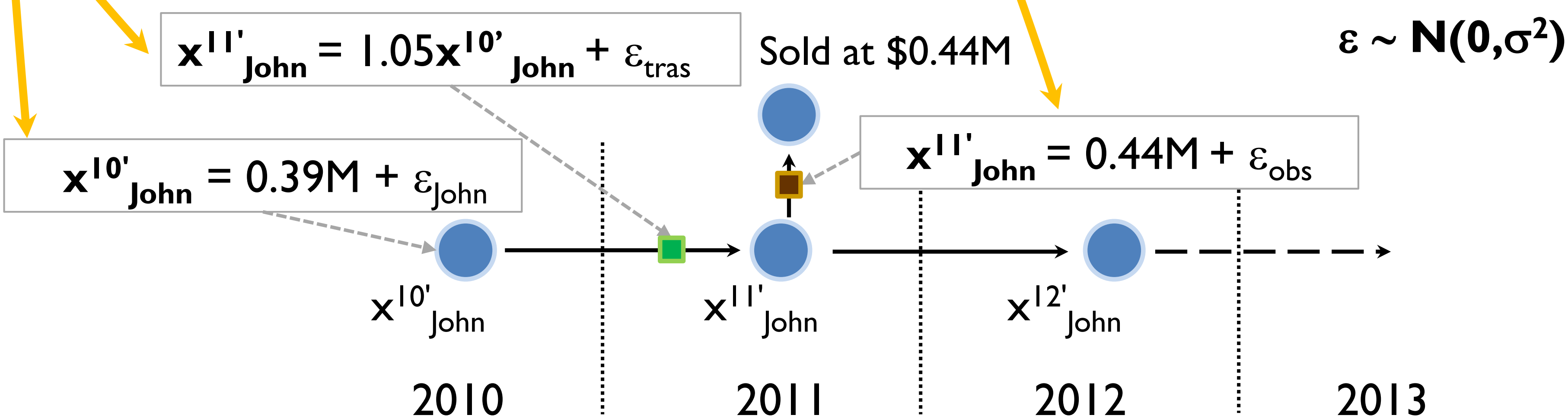


Example – Kalman Filter for John's Home

- Input statements
  - **John**'s house price was **$0.39M at 2010**.
  - Each year, **John**'s house price **increases 5%**.
  - **John**'s house price is around the sold price.
  - **John**'s house is sold sporadically.
- Question: what is the price of John's house each year?

Example – Kalman Filter for John's Home

- Input statements
  - **John**'s house price was **$0.39M at 2010**.
  - Each year, **John**'s house price **increases 5%**.
  - **John**'s house price is around the sold price.
  - **John**'s house is sold sporadically.
- Question: what is the price of John's house each year?



Why Kalman Filter Takes $O(n^3)$ operations?

- Kalman Filtering steps
  1. **Input: prior belief,** $X^t \sim N(\mu^t, \Sigma^t)$

     n variables, $X^t = \{x^t_{John}, x^t_{Tom}, x^t_{Ann}, \ldots\}$.
  2. **Take the transition model:**

     $X^{t+1} = A_T X^t + \varepsilon_{trans}$ when $\varepsilon_{trans} = N(0, \Sigma_T)$.
  3. **Updated covariance matrix:** $\Sigma^{t'} = A_T \Sigma^t A_T^T + \Sigma_T$.
  4. **Take the observation model:**

     $X^{t+1} = A_O Obs^{t+1} + \varepsilon_{obs}$ when $\varepsilon_{obs} = N(0, \Sigma_O)$.
  5. **Kalman gain:** $K = \Sigma^{t'} A_O^T (A_O \Sigma^{t'} A_O^T + \Sigma_O)^{-1}$.
  6. **Output: update belief,** $X^{t+1} \sim N(\mu^{t+1}, \Sigma^{t+1})$

     New mean: $\mu^{t+1} = \mu^t + K(Obs^{t+1} - \mu^t)$

     New covariance: $\Sigma^{t+1} = (I - KA_O) \Sigma^{t'}$

Why Kalman Filter Takes $O(n^3)$ operations?

- Kalman Filtering steps
  1. **Input: prior belief,** $X^t \sim N(\mu^t, \Sigma^t)$

     n variables, $X^t = \{x^t_{John}, x^t_{Tom}, x^t_{Ann}, \ldots\}$.
  2. **Take the transition model:**

     $X^{t+1} = A_T X^t + \varepsilon_{trans}$ when $\varepsilon_{trans} = N(0, \Sigma_T)$.
  3. **Updated covariance matrix:** $\Sigma^{t'} = A_T \Sigma^t A_T^T + \Sigma_T$.
  4. **Take the observation model:**

     $X^{t+1} = A_O Obs^{t+1} + \varepsilon_{obs}$ when $\varepsilon_{obs} = N(0, \Sigma_O)$.
  5. **Kalman gain:** $K = \Sigma^{t'} A_O^T (A_O \Sigma^{t'} A_O^T + \Sigma_O)^{-1}$
  6. **Output: update belief,** $X^{t+1} \sim N(\mu^{t+1}, \Sigma^{t+1})$

     New mean: $\mu^{t+1} = \mu^t + K(Obs^{t+1} - \mu^t)$

     New covariance: $\Sigma^{t+1} = (I - KA_O)\,\Sigma^{t'}$

Inversions and multiplications of the n by n matrix need $O(n^3)$ operations.

$$\Sigma^t = \begin{bmatrix} \sigma^2_{1,1} & \cdots & \sigma^2_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma^2_{n,1} & \cdots & \sigma^2_{n,n} \end{bmatrix}$$

Relational Kalman Filter

A set of element shares relationship!

- Input statements
  - **Town is a set of houses.**
  - **Town's** houses have initial prices **at 2010**.
  - Each year, **Town**'s house prices **increase 5%**.
  - **Town**'s house prices are around sold prices.
  - **Town**'s houses are sold sporadically.
- Question: what is the prices of Town's houses each year?

Relational Kalman Filter (IJCAI-11):
New Transition Models & Observation Models

- Input statements
  - **Town is a set of houses.**
  - **Town's** houses have initial prices **at 2010**.
  - Each year, **Town**'s house prices **increase 5%**.
  - **Town**'s house prices are around sold prices.
  - **Town**'s houses are sold sporadically.

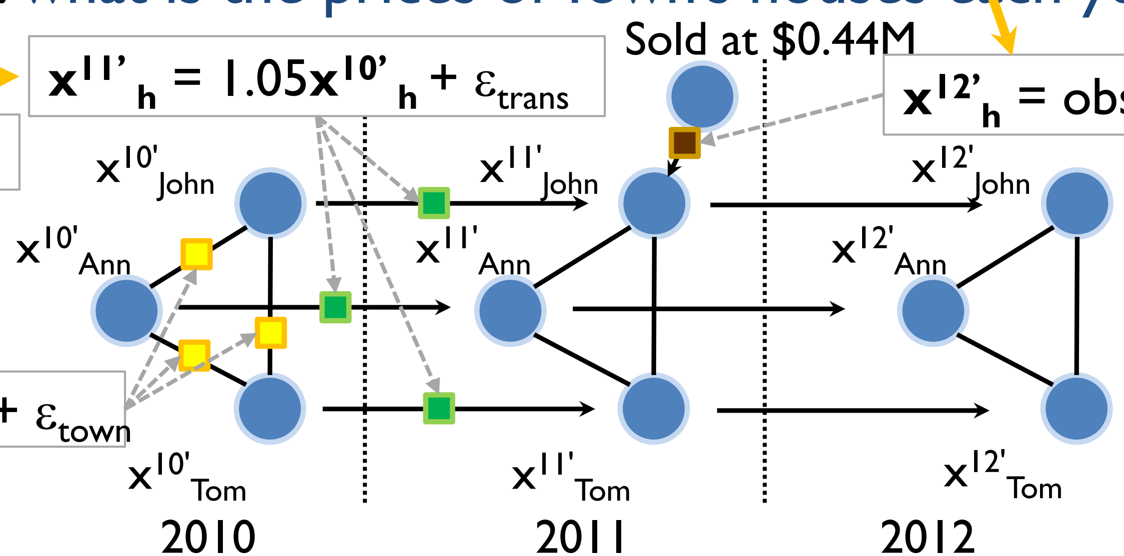
- Question: what is the prices of Town's houses each year?

$$x^{11'}_{h} = 1.05x^{10'}_{h} + \varepsilon_{trans}$$

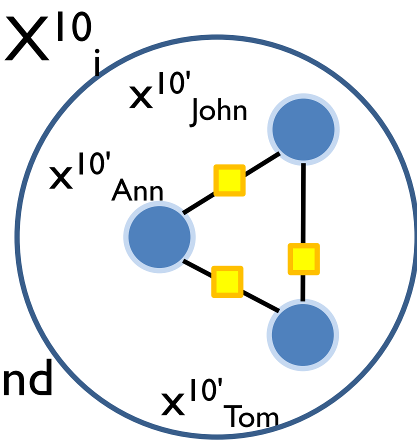
$$x^{12'}_{h} = obs^{12'}_{h} + \varepsilon_{obs}$$

$$h, h' \in \mathbf{Town}$$

$$X^{10'}_{h} = x^{10'}_{h'} + \varepsilon_{town}$$



Relational Gaussian Models (UAI-10)



- Definitions
  - X^t is a disjoint union of m clusters of state variables X^t_i.

    $X^t = \cup_i X^t_i$, e.g., $X^{10}_i = \{x^{10'}_{Ann}, x^{10'}_{John}, x^{10'}_{Tom}\}$.

  - Any two state variables in a cluster have the same variance and covariances.

    For $x, x' \in X^t_i$, $\sigma^2_{x,x} = \sigma^2_{x',x'}$ and for any y, $\sigma^2_{x,y} = \sigma^2_{x',y}$

  - Property: any multivariate Gaussian of X_t can be represented as a product of <u>pairwise linear Gaussian</u>, i.e., quadratic exponentials. (UAI-10)

$$P(X_t) \propto \prod_{i,j} \prod_{\substack{x \in X_{t,i} \\ y \in X_{t,j}}} \exp\left(\frac{(x - \beta_{RPM_{i,j}} y - \mu_{RPM_{i,j}})^2}{2 \cdot \sigma^2_{RPM_{i,j}}} \right) \prod_{x \in X_t} \exp\left(\frac{(x - \mu_x)^2}{2 \cdot \sigma^2_x} \right)$$

Relational Gaussian Models (UAI-10)

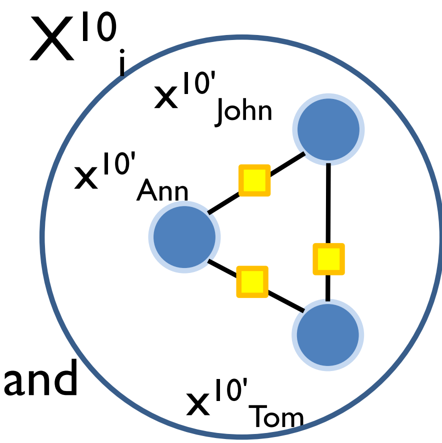
- Definitions
  - X^t is a disjoint union of m clusters of state variables X^t_i.

    $X^t = \cup_i X^t_i$, e.g., $X^{10}_i = \{x^{10'}_{Ann}, x^{10'}_{John}, x^{10'}_{Tom}\}$.

  - Any two state variables in a cluster have the same variance and covariances.

    For $x, x' \in X^t_i$, $\sigma^2_{x,x} = \sigma^2_{x',x'}$ and for any y, $\sigma^2_{x,y} = \sigma^2_{x',y}$

  - Property: any multivariate Gaussian of X_t can be represented as a product of <u>pairwise linear Gaussian</u>, i.e., quadratic exponentials. (UAI-10)

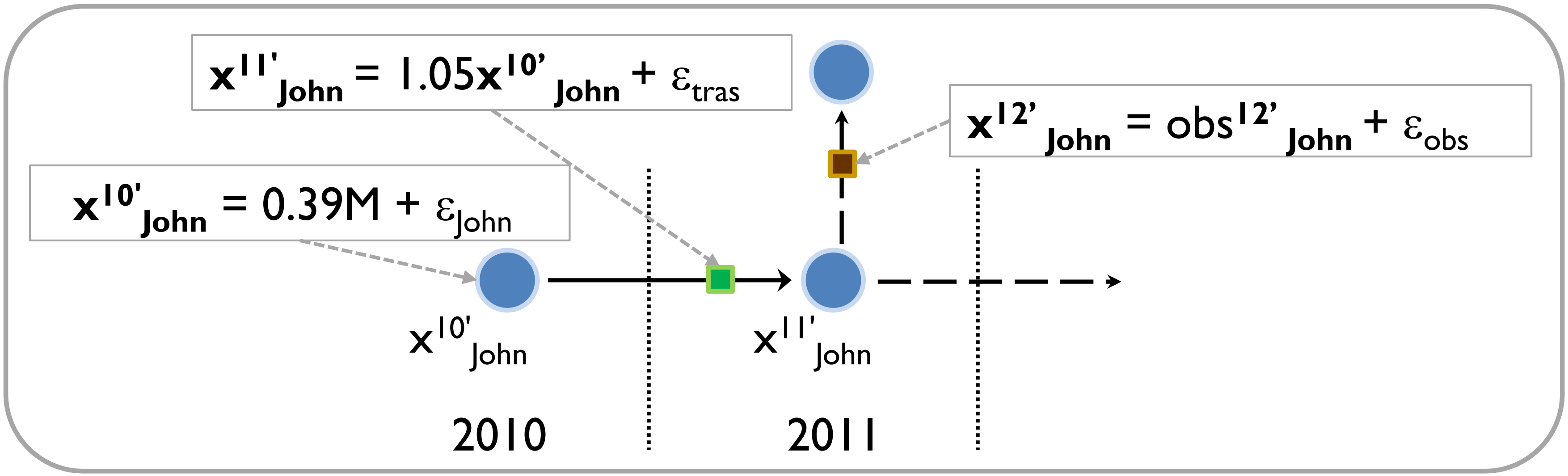


$$P(X_t) \propto \underbrace{\prod_{\boxed{i,j}} \prod_{\substack{x \in X_{t,i} \\ y \in X_{t,j}}} \exp\left(\frac{(x - \beta_{RPM_{i,j}} y - \mu_{RPM_{i,j}})^2}{2 \cdot \sigma^2_{RPM_{i,j}}} \right)}_{\textbf{Group parameters}} \underbrace{\prod_{x \in X_t} \exp\left(\frac{(x - \mu_x)^2}{2 \cdot \sigma^2_x} \right)}_{\textbf{Individual parameters}}$$

Groups i & j

Solve Kalman Filtering by Inference

Product of Gaussian pdfs



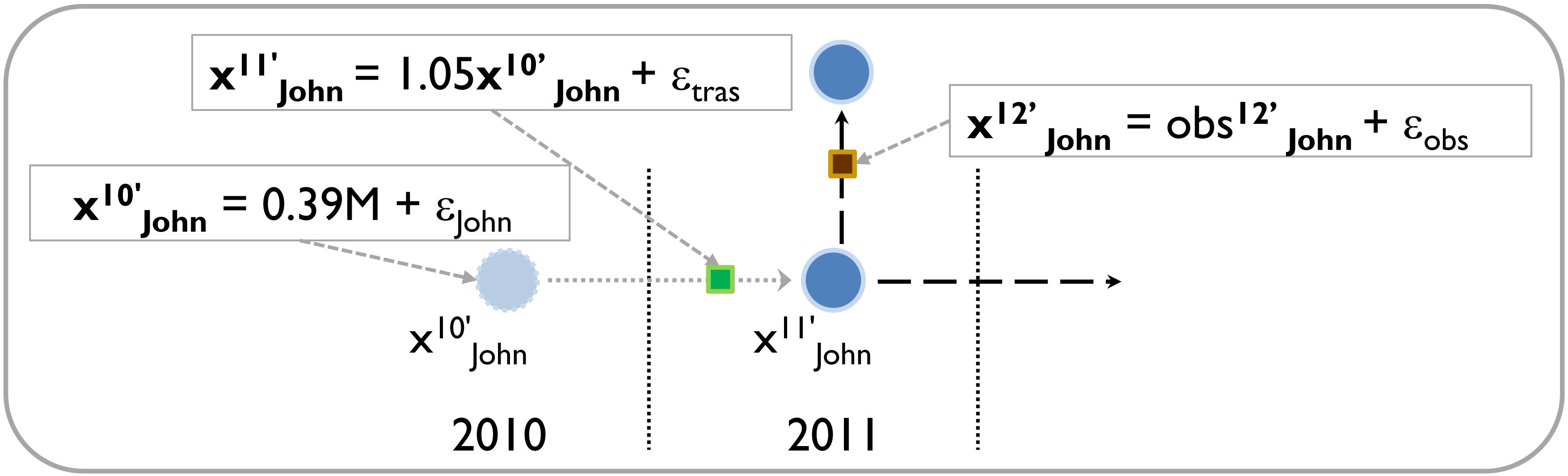
$$f(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$P(x^{10'}_{John}, x^{11'}_{John}) \propto \boxed{f\left(x^{10'}_{John}; 0.39M, \sigma^2_{John}\right)} \cdot \boxed{f\left(x^{11'}_{John} - 1.05x^{10'}_{John}; 0, \sigma^2_{trans}\right)}$$

Prior belief at 10’ Transition model

Solve Kalman Filtering by Inference

Marginalize variables (the previous time step)



$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Marginalize $x^{10'}_{John}$

Posterior $x^{11'}_{John}$

$$P(x^{10'}_{John}, x^{11'}_{John}) \propto f\left(x^{10'}_{John}; 0.39M, \sigma^2_{John}\right) \cdot f\left(x^{11'}_{John} - 1.05x^{10'}_{John}; 0, \sigma^2_{trans}\right)$$

$$P(x^{11'}_{John}) = \boxed{\int_{-\infty}^{\infty} P(x^{10'}_{John}, x^{11'}_{John})\, d\,x^{10'}_{John}} = \boxed{f\left(x^{11'}_{John}; 0.41M, \sigma^2_{John11}\right)}$$

Does RKF take less than $O(n^3)$ operations?

Answer: Not yet.

- Kalman Filtering steps

1. Input: prior belief, $X^t \sim N(\mu^t, \Sigma^t)$

   n variables, $X^t = \{x^t_{John}, x^t_{Tom}, x^t_{Ann}, \ldots\}$.

2. Take the transition model:

   $X^{t+1} = A_T X^t + \varepsilon_{trans}$ when $\varepsilon_{trans} = N(0, \Sigma_T)$.

3. Updated covariance matrix: $\Sigma^{t'} = A_T \Sigma^t A_T^T + \Sigma_T$.

4. Take the observation model:

   $X^{t+1} = A_O Obs^{t+1} + \varepsilon_{obs}$ when $\varepsilon_{obs} = N(0, \Sigma_O)$.

5. Kalman gain: $\mathbf{K = \Sigma^{t'} A_O^T (\boxed{A_O \Sigma^{t'} A_O^T + \Sigma_O})^{-1}}$.

6. Output: update belief, $X^{t+1} \sim N(\mu^{t+1}, \Sigma^{t+1})$

   New mean: $\mu^{t+1} = \mu^t + K(Obs^{t+1} - \mu^t)$

   New covariance: $\Sigma^{t+1} = (I - KA_O)\, \Sigma^{t'}$

$$\Sigma^t = \begin{bmatrix} \sigma_{xx}^2 & \cdots & \sigma_{xy}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{xy}^2 & \cdots & \sigma_{xx}^2 \end{bmatrix}$$

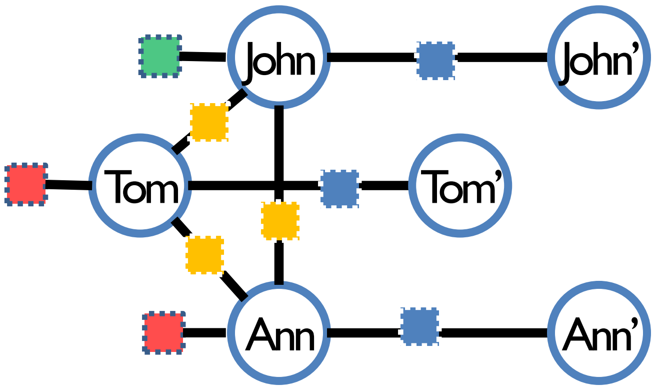
Relational Observation Models

$$\Sigma^{t+1} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1n}^2 \\ \vdots & ? & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}$$

Key Intuition in RKF

Under some conditions, a set of continuous RVs continues to have **the same pairwise relationships** during filtering.

- **Vanilla Kalman filter:**

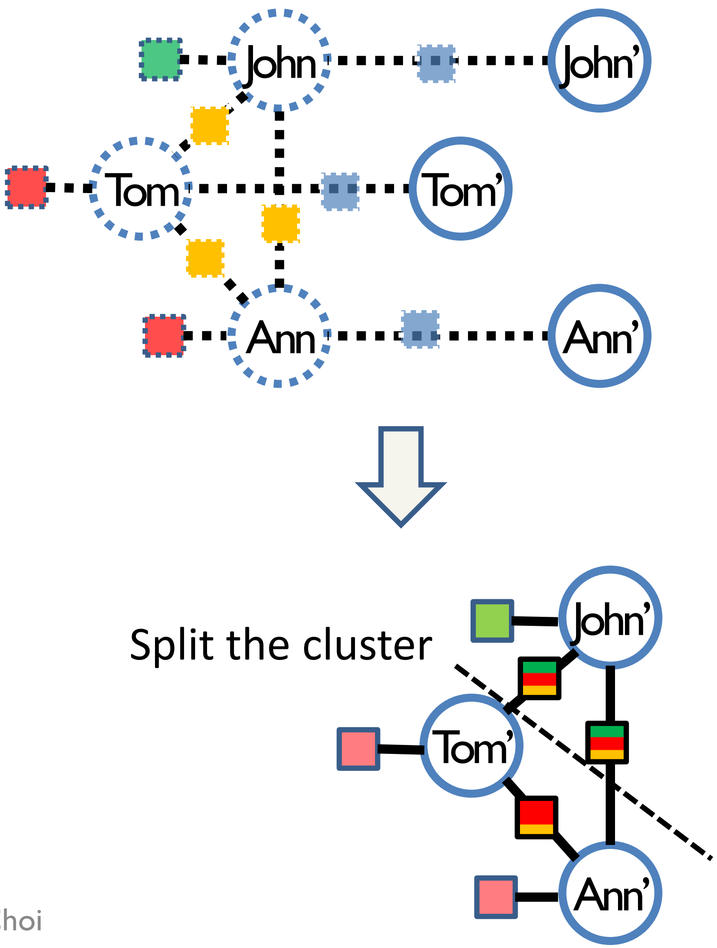


- **New filtering in RKF :**

Key Intuition in RKF

Under some conditions, a set of continuous RVs continues to have **the same pairwise relationships** during filtering.

- **Vanilla Kalman filter:**

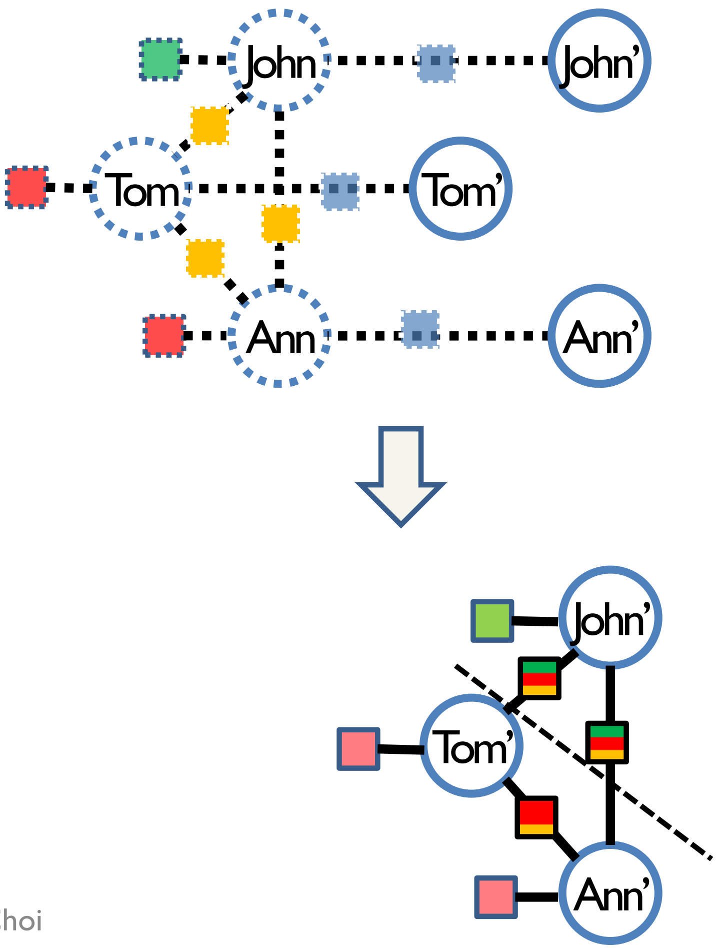


- **New filtering in RKF :**

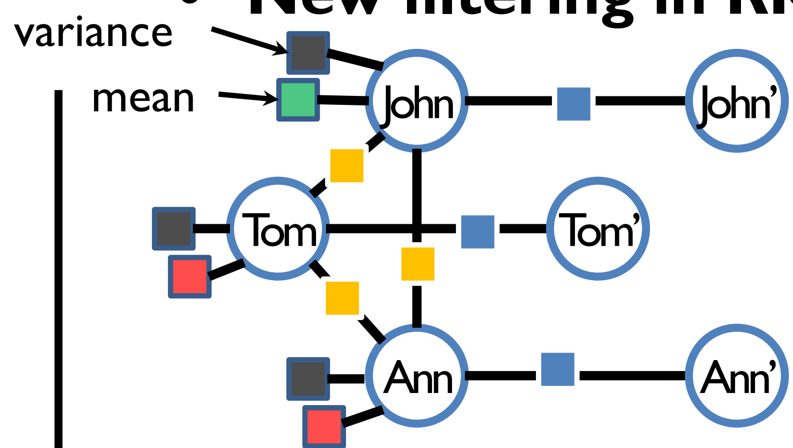
Key Intuition in RKF

Under some conditions, a set of continuous RVs continues to have **the same pairwise relationships** during filtering.

- **Vanilla Kalman filter:**



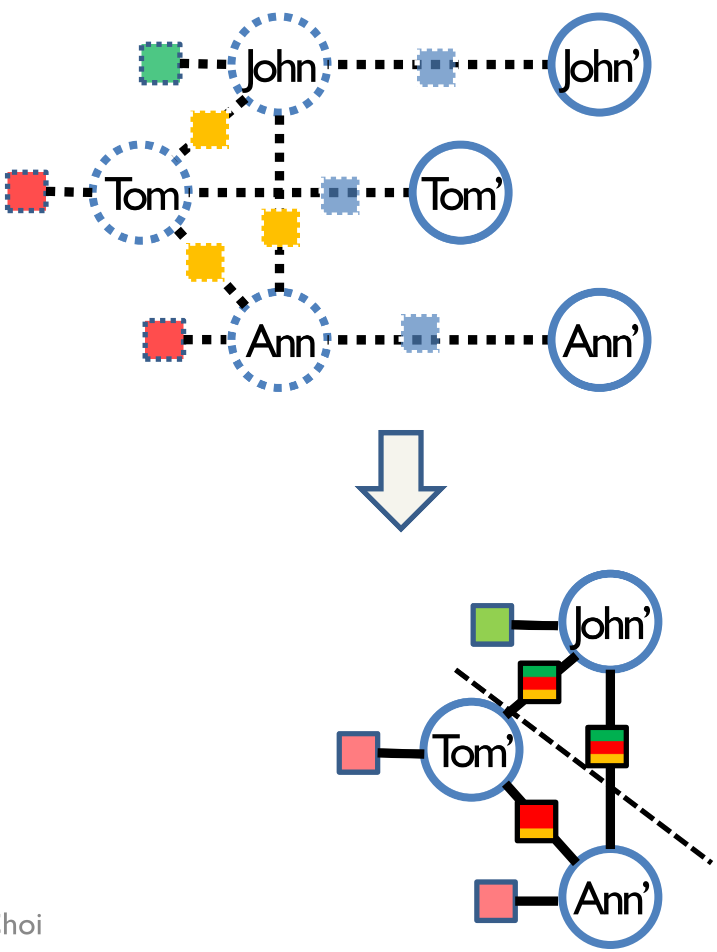
- **New filtering in RKF :**



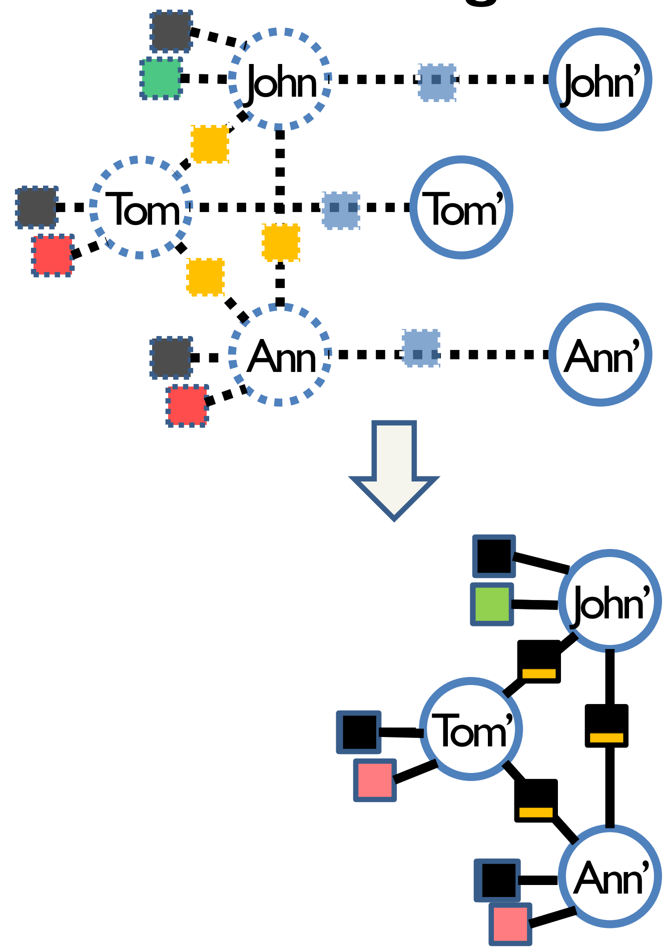
Key Intuition in RKF

Under some conditions, a set of continuous RVs continues to have **the same pairwise relationships** during filtering.

- **Vanilla Kalman filter:**



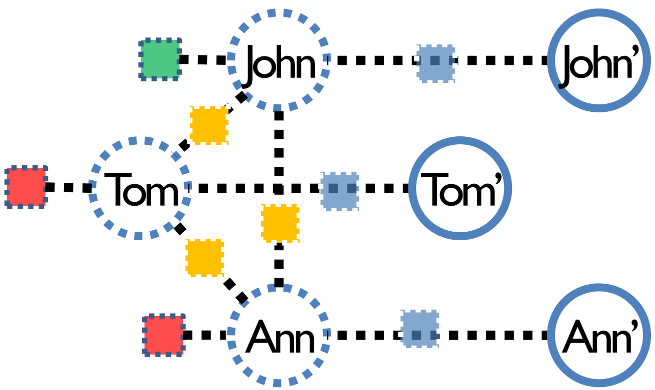
- **New filtering in RKF :**



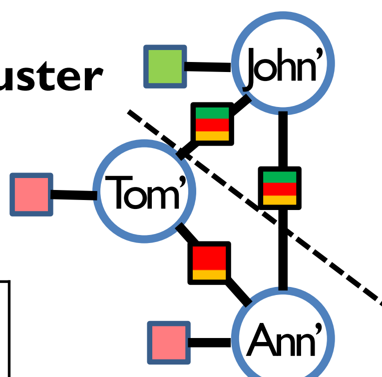
Key Intuition in RKF

Under some conditions, a set of continuous RVs continues to have **the same pairwise relationships** during filtering.

- **Vanilla Kalman filter:**



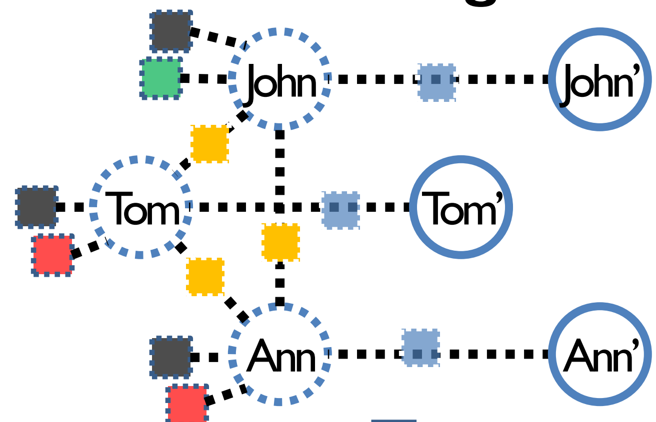
Split the cluster

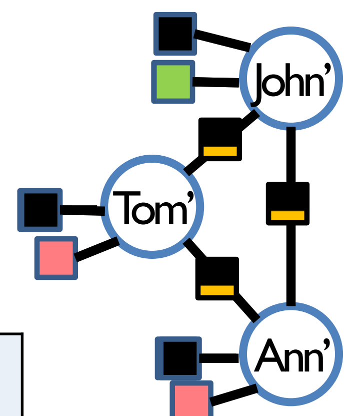


n: # of variables

$\boxed{O(\mathbf{n}^3)}$

- **New filtering in RKF :**





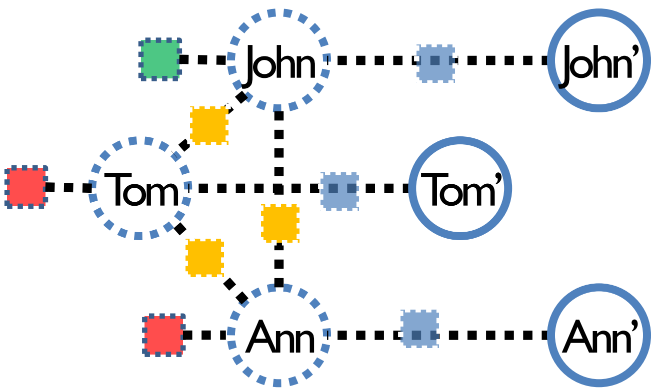
m: # of clusters

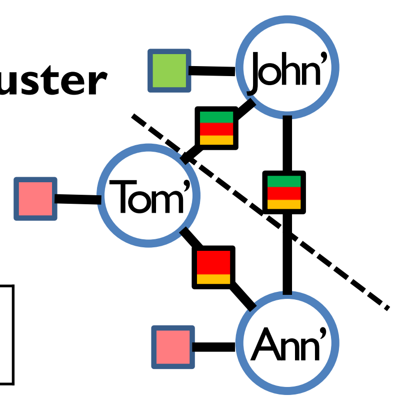
$\boxed{O(\mathbf{n} \cdot \mathrm{m}^2)}$

Key Intuition in RKF

Theorem: Two variables in a cluster continue to have **the same variance and covariances** at the next time step **if the same # of obs is made.**

- **Vanilla Kalman filter:**



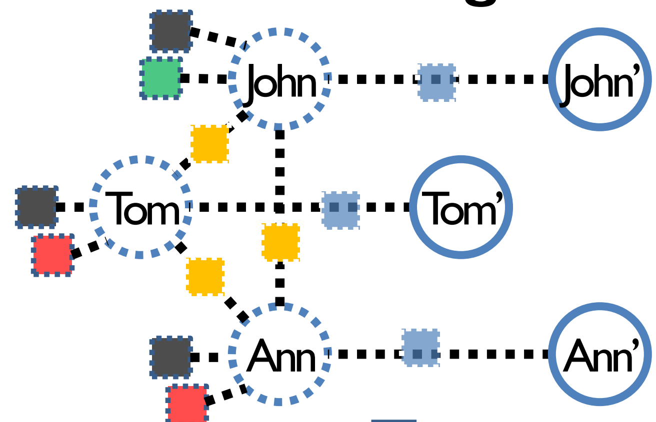


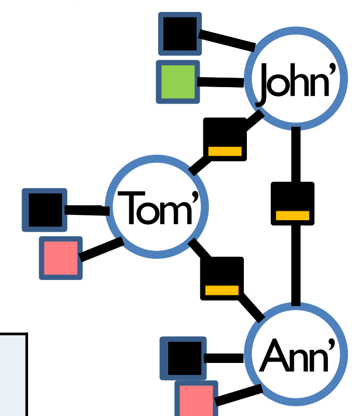
Split the cluster

n: # of variables

$O(\mathbf{n}^3)$

- **New filtering in RKF :**





m: # of clusters

$O(\mathbf{n} \cdot m^2)$

The Relational Kalman Filtering Algorithm

Marginalize all X^t in time step, t

$$\phi'(X^{t+1}) \leftarrow \int \phi_T(X^{t+1} | X^t)\phi(X^t)\, dX^t$$

Split variables from clusters given different # of obs.

$$X^{t+1}_{new} \leftarrow Split(X^{t+1}, O^{t+1})$$

Apply update models

$$\phi(X^{t+1}_{new}) \leftarrow \phi'(X^{t+1}_{new}) \cdot \phi_O(O^{t+1}|X^t)$$

Return $\phi(X^{t+1}_{new})$, go to time step, t+1

With dense observations, m << n is maintained!

The Relational Kalman Filtering Algorithm

Marginalize all X^t in time step, t

$$\phi'(X^{t+1}) \leftarrow \int \phi_T(X^{t+1} | X^t)\phi(X^t)\, dX^t$$

Next slide

Split variables from clusters given different # of obs.

$$X_{new}^{t+1} \leftarrow Split(X^{t+1}, O^{t+1})$$

With dense observations, m << n is maintained!

Apply update models

$$\phi(X_{new}^{t+1}) \leftarrow \phi'(X_{new}^{t+1}) \cdot \phi_O(O^{t+1} | X^t)$$

Return $\phi(X_{new}^{t+1})$, go to time step, t+1

Relational Kalman Filtering Algorithms

- Filtering is inference with RGMs:
  - Marginalize all state variables (x_{john}^{10}, x_{Tom}^{10} ,…) at time t.

$$\int \exp\left(-\sum_{i,j} \sum_{\substack{x\in X_{t,i} \\ y\in X_{t,j}}} \frac{(x-\beta_{RPM_{i,j}} y-\mu_{RPM_{i,j}})^2}{2\cdot\sigma^2_{RPM_{i,j}}} - \sum_{x\in X_t} \frac{(x-\mu_x)^2}{2\cdot\sigma_x^2} - \sum_{i,j} \sum_{\substack{x\in X_{t,i} \\ y\in X_{t,j}}} \frac{\left(y_{t+1}-\beta_{RTM_{i,j}} x_t\right)^2}{2\cdot\sigma^2_{RTM_{i,j}}} - \sum_{i} \sum_{x,o_x\in O_i} \frac{(x-o)^2}{2\cdot\sigma^2_{ROM_i}} \right) dX_t$$

  - Marginalize a variable $x \in X_t$ ($X'_t = X_t \backslash x$)

$$\int\int \exp\left(-Ax^2+2Bx+C\right) dx\, dX'_t = \int \frac{\sqrt{\pi}}{\sqrt{A}} \exp\left(B^2/A+C\right) dX'_t$$

  - Marginalization preserves pair-wise potentials.

$$B^2 = \sum_i \left(c^2_{t,i} \sum_{y\in X'_{t,i}} y^2 + c^2_{t+1,i} \sum_{y\in X_{t+1,i}} y^2 \right) + 2\sum_{i,j} \left(c_{t,i} c_{t,j} \sum_{\substack{y\in X'_{t,i} \\ y'\in X'_{t,j}}} yy' + c_{t,i} c_{t+1,j} \sum_{\substack{y\in X'_{t,i} \\ y'\in X'_{t+1,j}}} yy' + c_{t+1,i} c_{t+1,j} \sum_{\substack{y\in X'_{t+1,i} \\ y'\in X'_{t+1,j}}} yy' \right) + \alpha$$

  - Continue to marginalize all remaining variables.

Relational Kalman Filtering Algorithms

- Filtering is inference with RGMs:
  - Marginalize all state variables (x_{john}^{10}, x_{Tom}^{10} ,…) at time t.

$$\int \exp\left(-\sum_{i,j}\sum_{\substack{x\in X_{t,i}\\ y\in X_{t,j}}} \frac{(x-\beta_{RPM_{i,j}} y-\mu_{RPM_{i,j}})^2}{2\cdot\sigma^2_{RPM_{i,j}}} - \sum_{x\in X_t}\frac{(x-\mu_x)^2}{2\cdot\sigma_x^2} - \sum_{i,j}\sum_{\substack{x\in X_{t,i}\\ y\in X_{t,j}}} \frac{\left(y_{t+1}-\beta_{RTM_{i,j}} x_t\right)^2}{2\cdot\sigma^2_{RTM_{i,j}}} - \sum_i \sum_{x,o_x\in O_i}\frac{(x-o)^2}{2\cdot\sigma^2_{ROM_i}} \right) dX_t$$

  - Marginalize a variable $x \in X_t$ ($X'_t = X_t \backslash x$)

$$\int\int \exp\left(\boxed{-A}x^2 + \boxed{2Bx} + C\right) dx\, dX'_t = \int \frac{\sqrt{\pi}}{\sqrt{A}} \exp\left(\boxed{B^2}/A + C\right) dX'_t$$

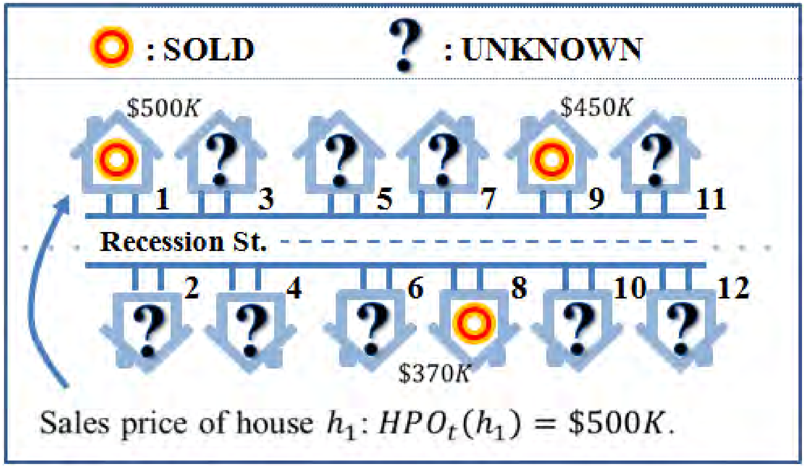
Constant   Sum of variables   Square of sum of variables

  - Marginalization preserves pair-wise potentials.

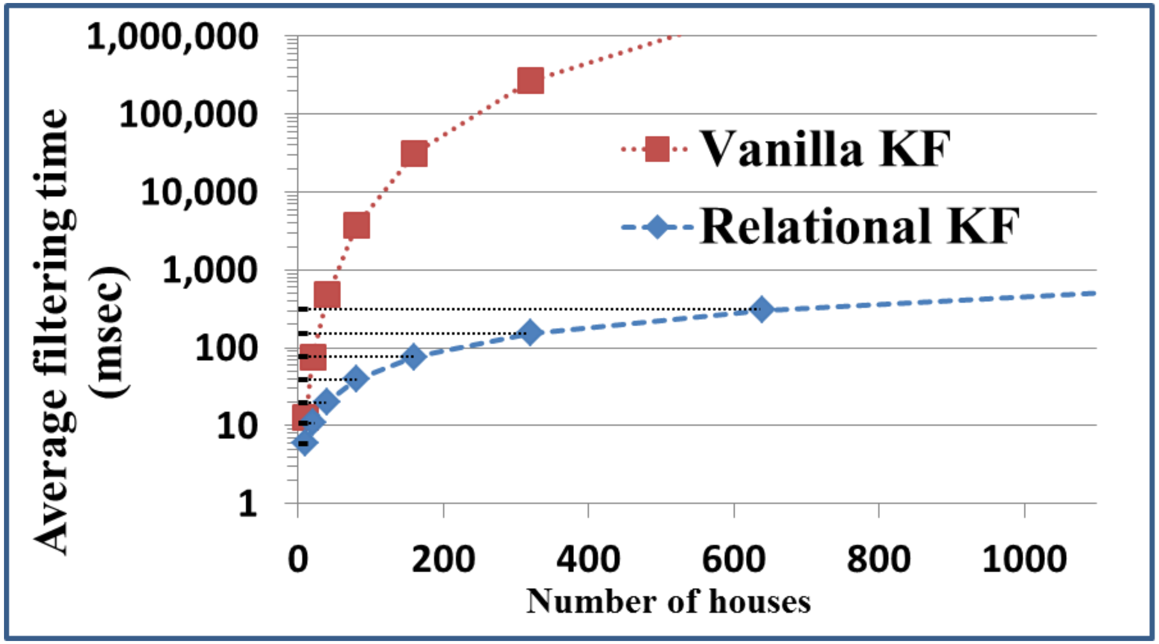
$$B^2 = \sum_i \left(c^2_{t,i} \sum_{y\in X'_{t,i}} y^2 + c^2_{t+1,i} \sum_{y\in X_{t+1,i}} y^2 \right) + 2\sum_{i,j} \left(c_{t,i}c_{t,j} \sum_{\substack{y\in X'_{t,i}\\ y'\in X'_{t,j}}} yy' + c_{t,i}c_{t+1,j} \sum_{\substack{y\in X'_{t,i}\\ y'\in X'_{t+1,j}}} yy' + c_{t+1,i}c_{t+1,j} \sum_{\substack{y\in X'_{t+1,i}\\ y'\in X'_{t+1,j}}} yy' \right) + \alpha$$

  - Continue to marginalize all remaining variables.

Experiments (Simulation)

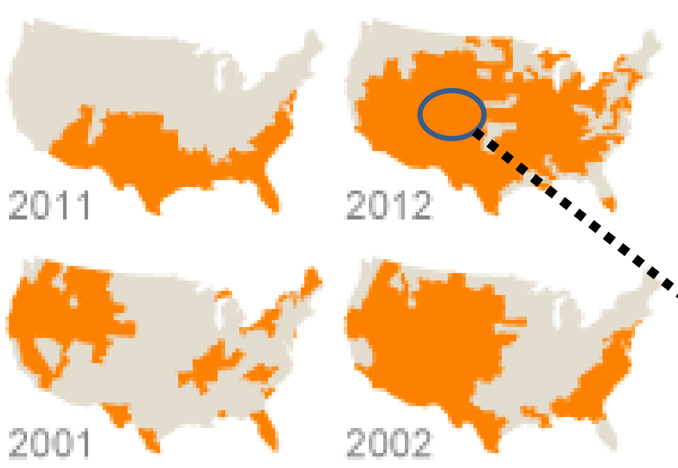


- Given: housing market example.
- Observations for:
  - Mortgage Rate
  - Sales prices for a set of houses
- Estimate the price of each house (mean and variance)

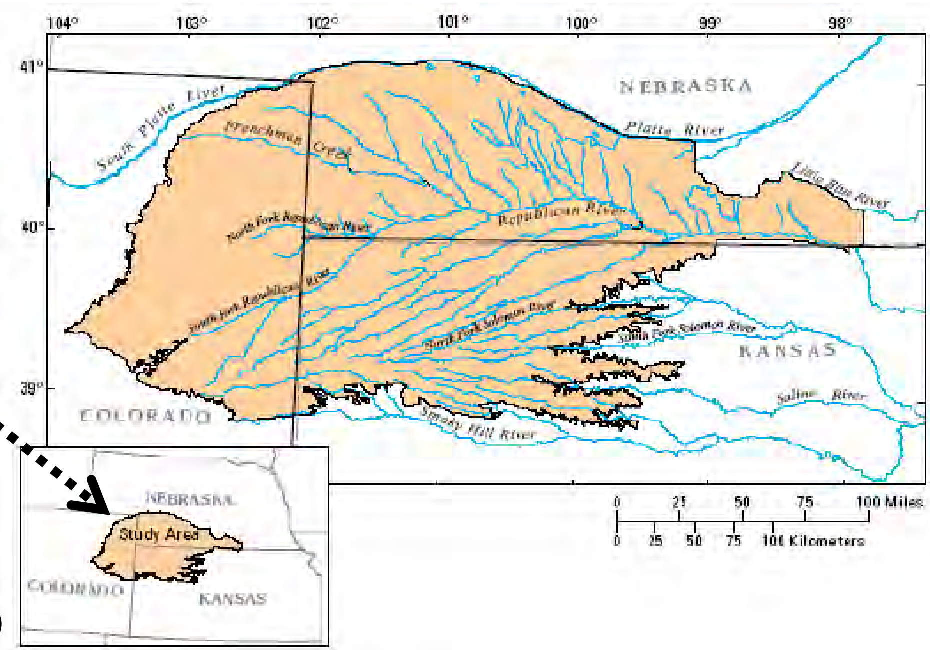


Experiments (Groundwater Models)

- Data is extracted in the largest aquifer (Ogallala) in US.
- Pumping (for farming) depletes many of water wells.
- Estimating level of groundwater is critical.

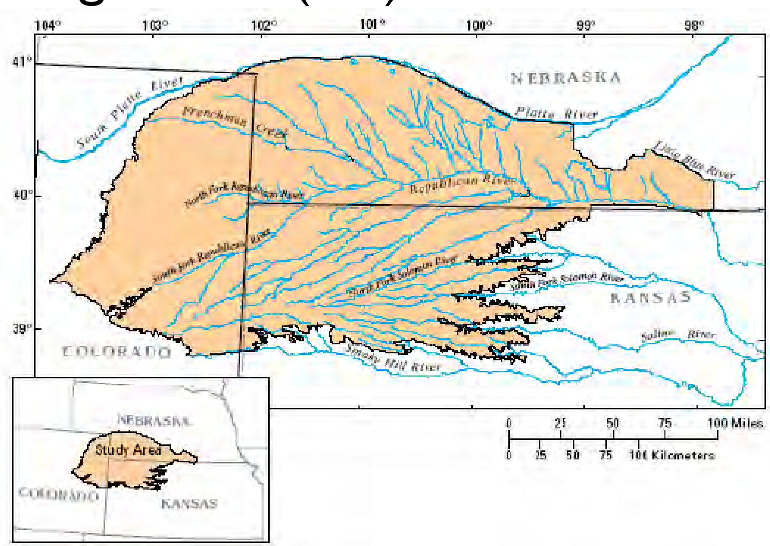


US drought maps (New York Times)



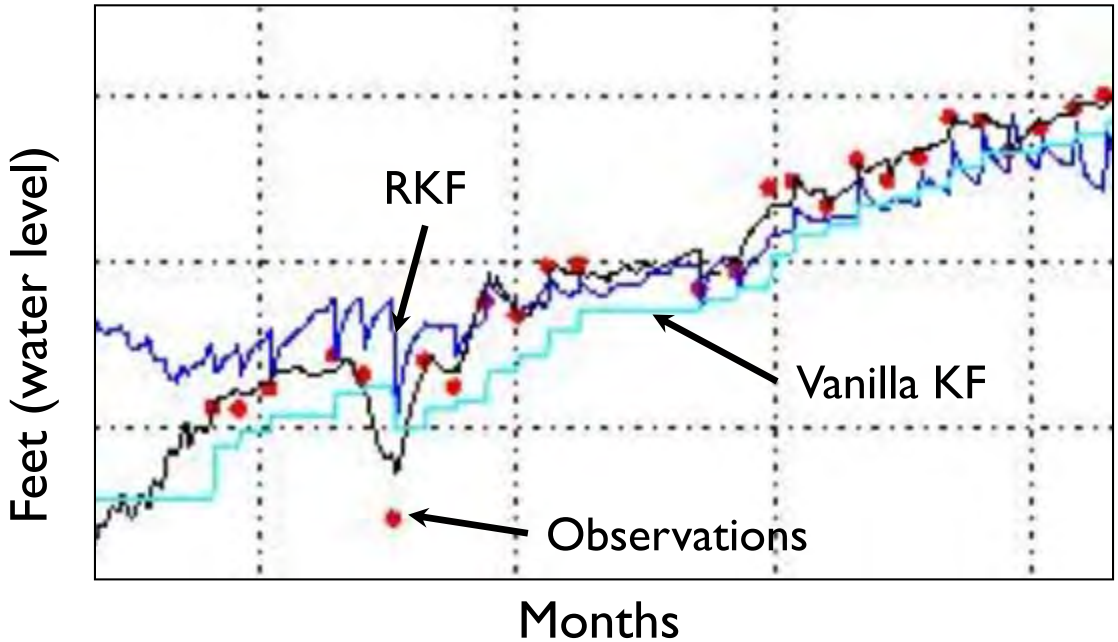
Experiments (Groundwater Models)

- Dataset
  - The model has measures (water levels) for 3078 water wells.
  - The measures span from 1918 to 2007 (about 900 months).
  - It has over 300,000 measurements.
- Cluster: 3078 wells into 10 groups.
- Train parameters using the auto regression (AR).
  - Vanilla Kalman filter
  - RKF



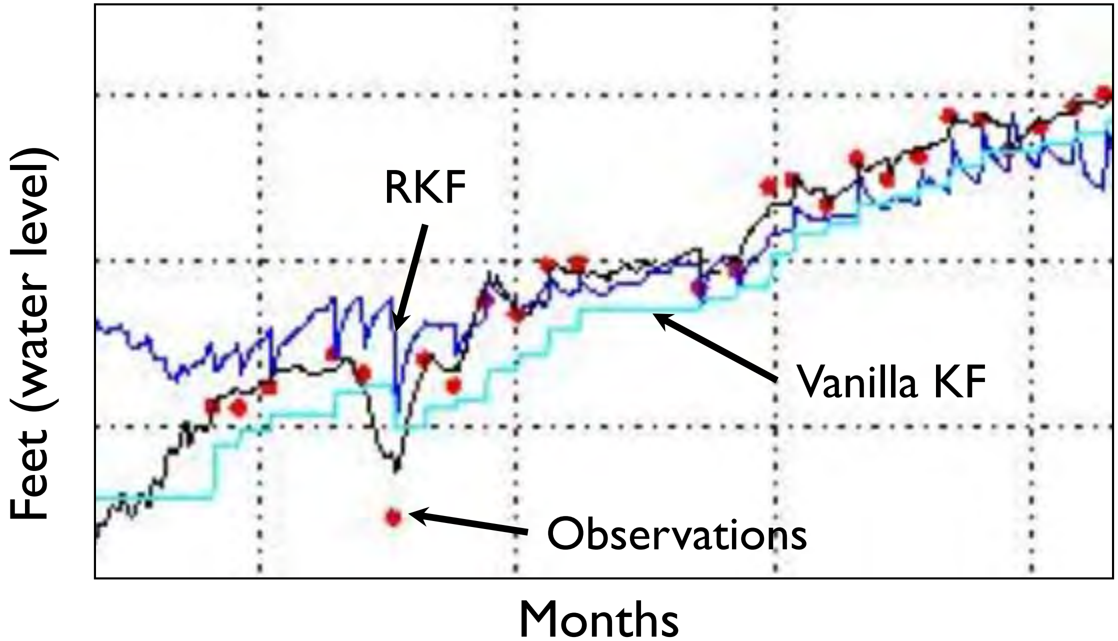
Experiments (Groundwater Models)

- Results (root-mean-square error)
  - Vanilla Kalman Filter: 4.17 feet (about 11.59 sec / filtering step).
  - RKF: 3.60 feet (about 0.60 sec / filtering step).



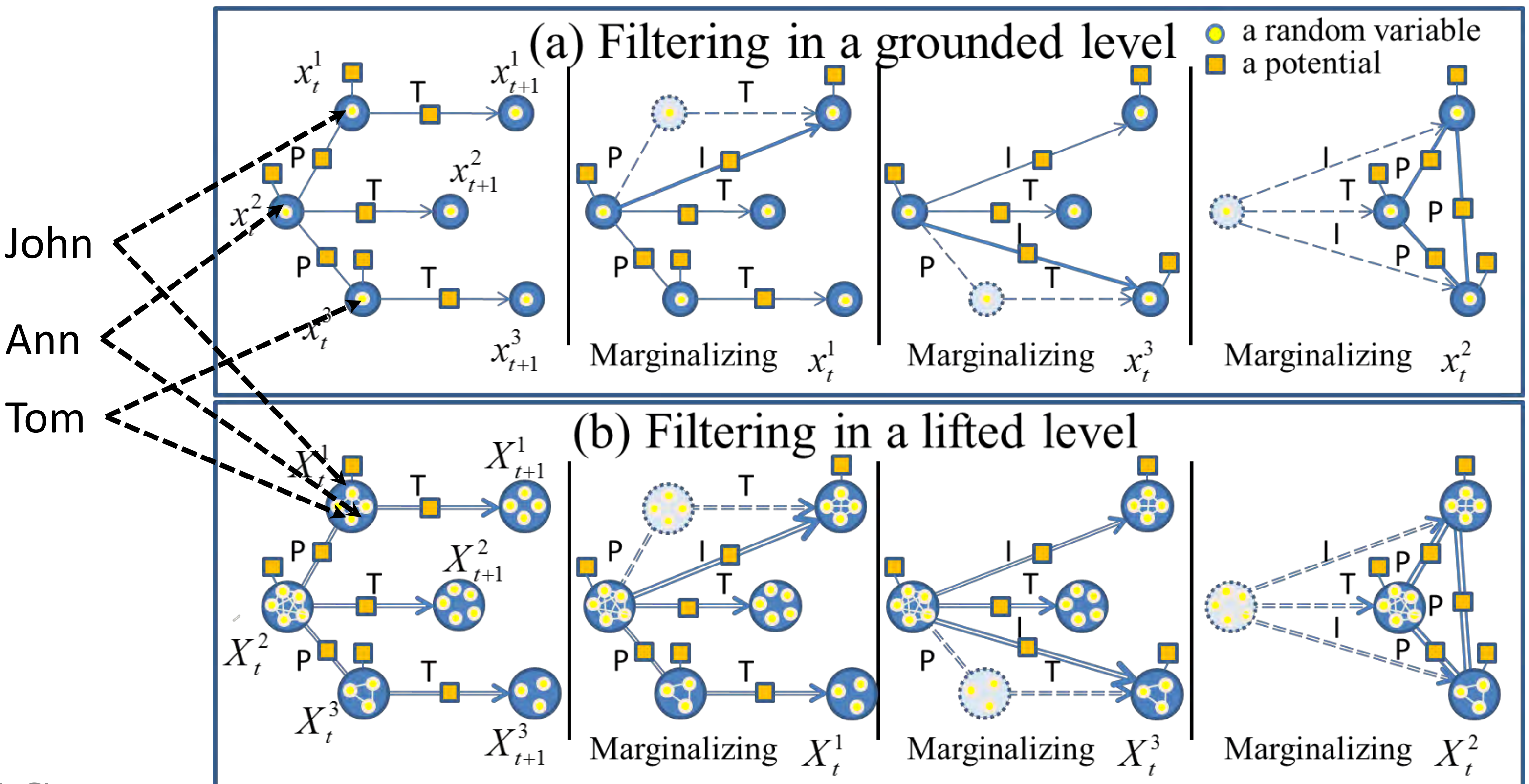
Experiments (Groundwater Models)

- Additional experiment for Vanilla KF + RKF
  - KF for coefficients of the transition model.
  - RKF for the observations model and covariance matrix (only) of the transition model.
- Result: 0.83 feet (about 11.49 sec / filtering step).



Summary of RKF

- I present a new filtering algorithm that enables linear time exact Kalman filtering in contrast to the cubic time traditional **KF**.



Exchangeable Random Variables (RVs)

Exchangeable RVs: a set of RVs, which are interchangeable among others.

$$P(x_1, \cdots, x_n) = P\left(x_{\pi(1)}, \cdots, x_{\pi(n)}\right), \quad \pi\text{: a permutation}$$

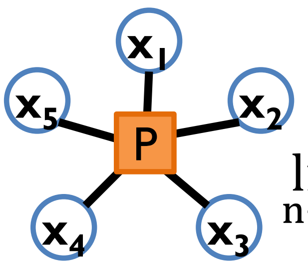
- Exchangeability is already exploited and utilized in many applications such as image & video retrieval and network analysis.
- Examples
  - Image & video matching: exchangeable image features
  - Econometrics: a set of exchangeable portfolio (in risk analysis)
  - The Netflix prize: groups of users & groups of movies

De Finetti's theorem and Dirichlet Process

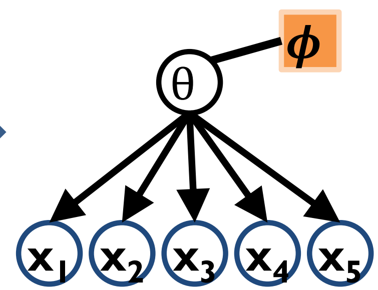
Exchangeable RVs: a set of RVs, which are interchangeable among others.

$$P(x_1, \cdots, x_n) = P\left(x_{\pi(1)}, \cdots, x_{\pi(n)}\right), \quad \pi\text{: a permutation}$$

[de Finetti, 1931] shows that any joint probability of **infinite, exchangeable,** binary RVs can be represented by **a mixture of iid RVs**.



$$\lim_{n\to\infty} P(x_1, \cdots, x_n)$$



$$\int_0^1 \theta^t (1-\theta)^{n-t} \phi(\theta) d\theta$$

when $t = \sum_i x_i$ and

$\phi(\theta)$ is a prior

[Carbonetto, Kisynski, de Freitas, Poole, 2005] uses the Dirichlet process to model infinite, exchangeable discrete variables.

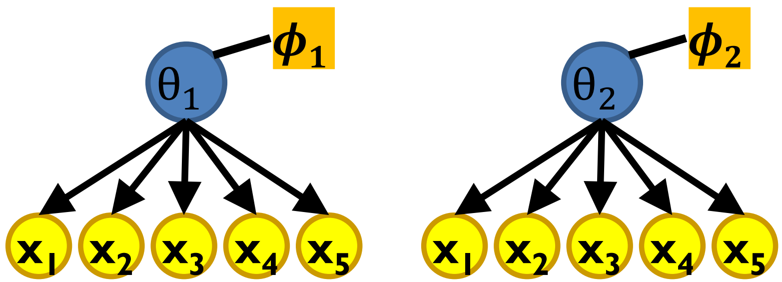
Research issues: continuous RVs, finite RVs (e.g., errors)

Relational Variational Inference (UAI-12)

The Dirichlet process does not have a closed-form

An inference issue: the product of Dirichlets is difficult to handle.

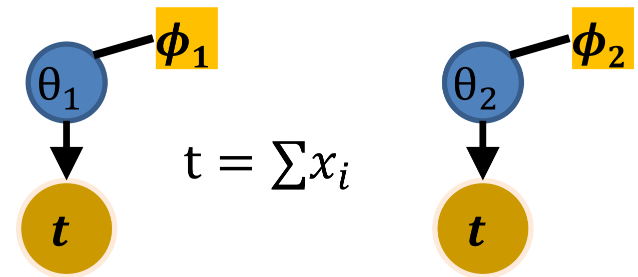
[Carbonetto, et al., 2005]



$$P_1(x_1, \cdots, x_n) = \int_0^1 \theta_1^t (1-\theta_1)^{n-t} \phi_1(\theta_1) d\theta_1$$

$$\theta_1^t(1-\theta_1)^{n-t} \cdot \theta_2^t(1-\theta_2)^{n-t} \neq \boldsymbol{\theta_3^t(1-\theta_3)^{n-t}}$$

This work



$$P_1(t) = \binom{n}{t} \int_0^1 \theta_1^t (1-\theta_1)^{n-t} \phi_1(\theta_1) d\theta_1$$

$$= \int_0^1 \mathbf{f}_{\mathbf{Binomial}}(\mathbf{t}; \mathbf{n}, \boldsymbol{\theta}_1)\, \phi_1(\theta_1) d\theta_1$$

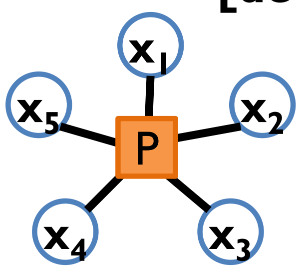
Gaussian approx. solves the issue!

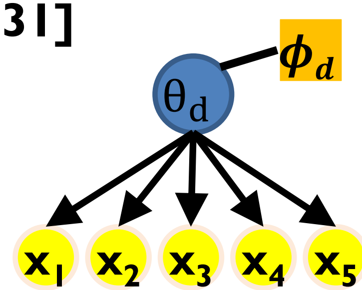
Relational Variational Inference (UAI-12)

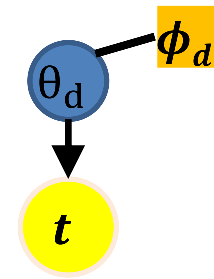
Mixture of Gaussians for Relational Hybrid Models

For **binary** exchangeable RVs.
[de Finetti, 1931]

This work: variational RHMs







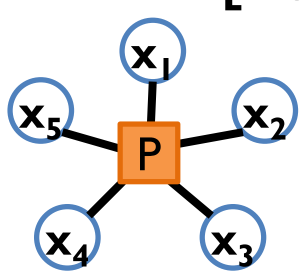
$t = \sum x_i$

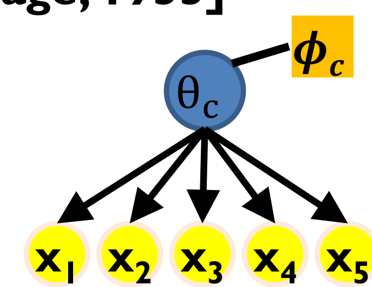
Input: $P_1(x_1, \cdots, x_n)$

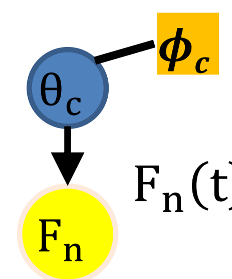
A mixture of iid RVs

A mixture of Binomial ≈
A mixture of Gaussians (MoGs)

For **continuous** exchangeable RVs.
[Hewitt & Savage, 1955]







$F_n(t) = \sum 1\{x_i \leq t\}$

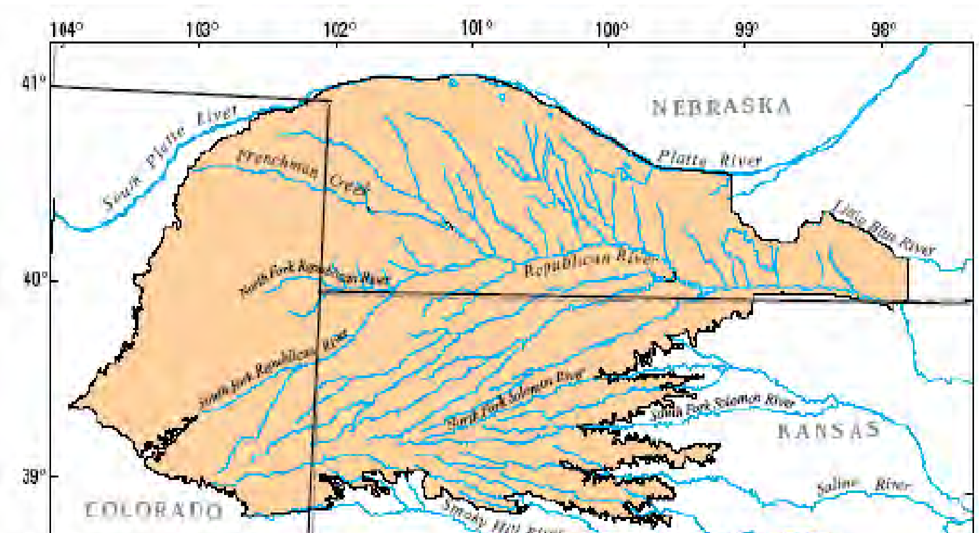
Input: $P_c(x_1, \cdots, x_n)$

A mixture of iid RVs

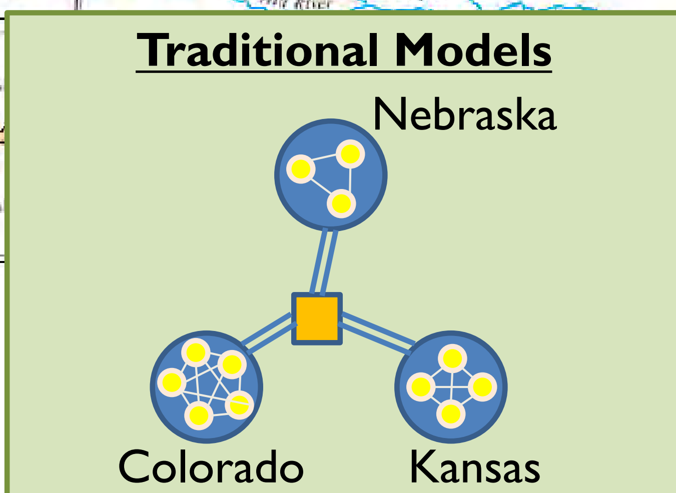
A mixture of pdfs
≈ A mixture of MoGs (KDE)

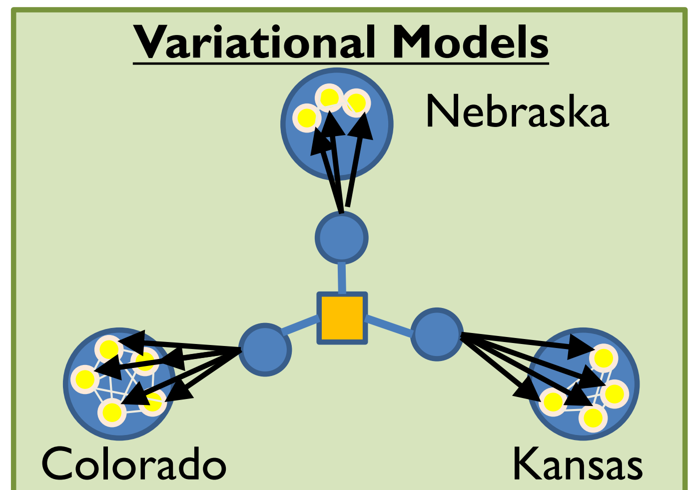
Relational Variational Inference (UAI-12)

Experiments at Groundwater



- ❑ Water level data over 3,078 wells from 1918 to 2000
- ❑ Clustering wells into 10 groups of (approx.) exchangeable RVs
- ❑ Learning a mixture of cdf (MoGs) for each group (training)
- ❑ Calculating the conditional probability of a set of RVs (testing)

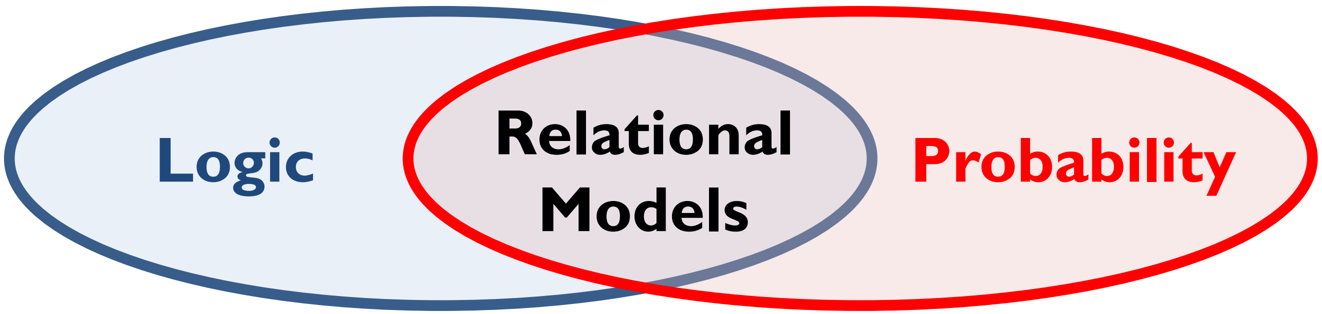




Inference time: **37.9 secs** **0.3 secs** (w/ similar precision)

Conclusion: The Big Picture: Towards Human-Level AI

- Human-like Knowledge Representation: **Logic + Probability**:



- Intuitively speaking:
  - Relational Models ≈ Probabilistic First-Order Logic (FOL)

    ≈ Probabilistic Relational Calculus (Relational DB).
  - Human-level knowledge representation

    E.g., A (owner) files Bankruptcy→ A's house is foreclosed.

Contents

Kaggle: Online Machine Learning Playground

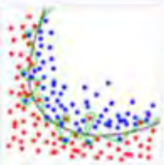
- Kaggle provides an online platform to learn and compete for several machine learning problems. (https://www.kaggle.com/competitions)
  - When there is a host who has a machine learning problem to solve
  - E.g., GE want to optimize the flight routes given an origin and destination and traffic and weather conditions. ($220K)
  - Data scientists compete to solve the problems.
  - Your submission will be evaluated immediately and posted online. https://www.kaggle.com/c/titanic-gettingStarted/leaderboard

Kaggle: Online Machine Learning Playground

Some of interesting datasets

Predict HIV Progression

This contest requires competitors to predict the likelihood that an HIV patient's infection will become less severe, given a small dataset and limited clinical information.



Don't Overfit!

With nearly as many variables as training cases, what are the best techniques to avoid disaster?

Titanic: Machine Learning from Disaster

Predict survival on the Titanic (with tutorials in Excel, Python and an introduction to Random Forests)

Facial Keypoints Detection

Detect the location of keypoints on face images

AMS 2013-2014 Solar Energy Prediction Contest

Forecast daily solar energy with an ensemble of weather models

Accelerometer Biometric Competition

Recognize users of mobile devices from accelerometer data

CHALEARN Gesture Challenge

Develop a Gesture Recognizer for Microsoft Kinect (TM)

Algorithmic Trading Challenge

Develop new models to accurately predict the market response to large trades.

Predicting Parkinson's Disease Progression with Smartphone Data

Can we objectively measure the symptoms of Parkinson's disease with a smartphone? We have the data to find out!

Million Song Dataset Challenge

Predict which songs a user will listen to.

Kaggle: Online Machine Learning Playground

Some of interesting datasets

- Datasets from top machine learning conferences
  - KDD - Author-Paper Identification Challenge
  - ICDM - Personalize Expedia Hotel Searches
  - NIPS, ICML - Multi-label Bird Species Classification

- Datasets from companies to recruit data scientists
  - Amazon - Employee Access Challenge
  - Facebook - Keyword Extraction
  - Yelp - How many "useful" votes will a Yelp review receive?

Kaggle: Online Machine Learning Playground

- Sometimes, winners posts their winning strategies.
  - Dogs vs Cats: Convolutional Neural Network
  - Titanic: Random Forests
  - http://trevorstephens.com/post/72916401642/titanic-getting-started-with-r

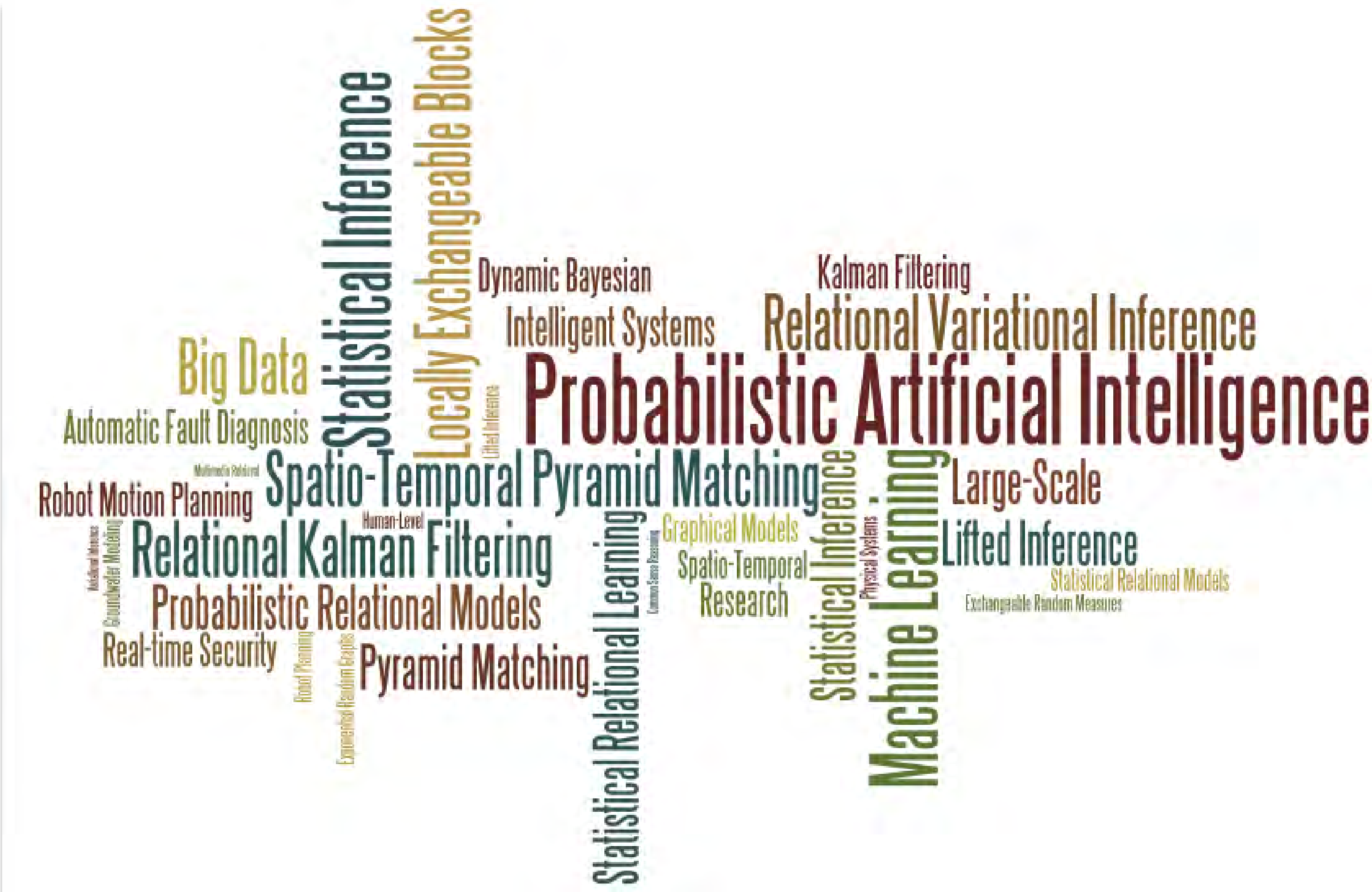
- Some teams are ranked from Machine Learning class at UNIST
  - https://www.kaggle.com/users/146714/yunseong-hwang
  - https://www.kaggle.com/users/147300/jongho-kim-at-unist

Kaggle: Online Machine Learning Playground

- Kaggle also provides links for machine learning library
  https://www.kaggle.com/wiki/Algorithms

  - Linear Regression
  - Logistic Regression
  - K Nearest Neighbors
  - Support Vector Machines
  - Random Forests - has performed very well on many competitions
  - Neural Networks
  - Ensembling
  - Elastic Net

Thank you!



If you are interested in our Probabilistic Artificial Intelligence Lab, please send an e-mail to jaesik@unist.ac.kr

Ongoing Research

- Extend RKF into non-linear systems
  - E.g., Relational Rao-Blackwellized Particle Filter.
- Learning and inference with large-scale models
  - Apply the principles of RKF and several applications.

- Machine Learning algorithms: e.g.:
  - Learning Relational Kalman Filtering
  - Learning Exponential Random Graph Models
  - Best Predictive Generalized Linear Mixed Model using LASSO

New Result in RKF

Theorem: Two variables in a cluster continue to have **the same variance and covariances** at the next time step **if the same # of obs is made.**

E.g. x and x' in X_i have the same variances and covariances with different means:

x

$$\int \exp\left(-\sum_{i,j} \sum_{\substack{x \in X_{t,i} \\ y \in X_{t,j}}} \frac{(x - \beta_{RPM_{i,j}} y - \mu_{RPM_{i,j}})^2}{2 \cdot \sigma^2_{RPM_{i,j}}} - \frac{\boxed{(x-\mu_x)^2}}{2 \cdot \sigma_x^2} \sum_{i,j} \sum_{\substack{x \in X_{t,i} \\ y \in X_{t,j}}} \frac{\left(y_{t+1} - \beta_{RTM_{i,j}} x_t\right)^2}{2 \cdot \sigma^2_{RTM_{i,j}}} - \sum_i \sum_{x, o_x \in O_i} \frac{(x-o)^2}{2 \cdot \sigma^2_{ROM_i}} \right) dX_t$$

$$B = \sum_i \left(\sum_{y \in X'_{t,i}} c_{t,i} y + \sum_{y \in X_{t+1,i}} c_{t+1,i} y \right) + c_\mu \boxed{\mu_x} + c$$

x'

$$\int \exp\left(-\sum_{i,j} \sum_{\substack{x \in X_{t,i} \\ y \in X_{t,j}}} \frac{(x - \beta_{RPM_{i,j}} y - \mu_{RPM_{i,j}})^2}{2 \cdot \sigma^2_{RPM_{i,j}}} - \frac{\boxed{(x'-\mu_{x'})^2}}{2 \cdot \sigma_x^2} - \sum_{i,j} \sum_{\substack{x \in X_{t,i} \\ y \in X_{t,j}}} \frac{\left(y_{t+1} - \beta_{RTM_{i,j}} x_t\right)^2}{2 \cdot \sigma^2_{RTM_{i,j}}} - \sum_i \sum_{x, o_x \in O_i} \frac{(x-o)^2}{2 \cdot \sigma^2_{ROM_i}} \right) dX_t$$

$$B = \sum_i \left(\sum_{y \in X'_{t,i}} c_{t,i} y + \sum_{y \in X_{t+1,i}} c_{t+1,i} y \right) + c_\mu \boxed{\mu_x} + c$$

New Result in RKF

Theorem: Two variables in a cluster continue to have **the same variance and covariances** at the next time step **if the same # of obs is made.**

E.g. x and x' in X_i have the same variances and covariances with different means:

X

$$\int \exp\left(-\sum_{i,j}\sum_{\substack{x\in X_{t,i}\\ y\in X_{t,j}}} \frac{(x-\beta_{RPM_{i,j}}\, y-\mu_{RPM_{i,j}})^2}{2\cdot\sigma^2_{RPM_{i,j}}} - \frac{\boxed{(x-\mu_x)^2}}{2\cdot\sigma_x^2} \sum_{i,j}\sum_{\substack{x\in X_{t,i}\\ y\in X_{t,j}}} \frac{(y_{t+1}-\beta_{RTM_{i,j}}\, x_t)^2}{2\cdot\sigma^2_{RTM_{i,j}}} - \sum_i \sum_{x,o_x\in O_i} \frac{(x-o)^2}{2\cdot\sigma^2_{ROM_i}} \right) dX_t$$

$$B^2 = \sum_i \left(c^2_{t,i} \sum_{y\in X'_{t,i}} y^2 + c^2_{t+1,i} \sum_{y\in X_{t+1,i}} y^2 \right) + 2\sum_{i,j}\left(c_{t,i}c_{t,j} \sum_{\substack{y\in X'_{t,i}\\ y'\in X'_{t,j}}} yy' + c_{t,i}c_{t+1,j} \sum_{\substack{y\in X'_{t,i}\\ y'\in X'_{t+1,j}}} yy' + c_{t+1,i}c_{t+1,j} \sum_{\substack{y\in X'_{t+1,i}\\ y'\in X'_{t+1,j}}} yy' \right) + \boxed{\alpha}$$

Variances & Covariances

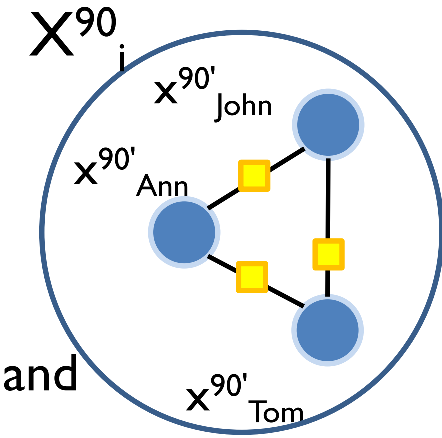
X'

$$\int \exp\left(-\sum_{i,j}\sum_{\substack{x\in X_{t,i}\\ y\in X_{t,j}}} \frac{(x-\beta_{RPM_{i,j}}\, y-\mu_{RPM_{i,j}})^2}{2\cdot\sigma^2_{RPM_{i,j}}} - \frac{\boxed{(x'-\mu_{x'})^2}}{2\cdot\sigma_x^2} - \sum_{i,j}\sum_{\substack{x\in X_{t,i}\\ y\in X_{t,j}}} \frac{(y_{t+1}-\beta_{RTM_{i,j}}\, x_t)^2}{2\cdot\sigma^2_{RTM_{i,j}}} - \sum_i \sum_{x,o_x\in O_i} \frac{(x-o)^2}{2\cdot\sigma^2_{ROM_i}} \right) dX_t$$

$$B^2 = \sum_i \left(c^2_{t,i} \sum_{y\in X'_{t,i}} y^2 + c^2_{t+1,i} \sum_{y\in X_{t+1,i}} y^2 \right) + 2\sum_{i,j}\left(c_{t,i}c_{t,j} \sum_{\substack{y\in X'_{t,i}\\ y'\in X'_{t,j}}} yy' + c_{t,i}c_{t+1,j} \sum_{\substack{y\in X'_{t,i}\\ y'\in X'_{t+1,j}}} yy' + c_{t+1,i}c_{t+1,j} \sum_{\substack{y\in X'_{t+1,i}\\ y'\in X'_{t+1,j}}} yy' \right) + \boxed{\alpha'}$$

Variances & Covariances

Relational Gaussian Models (UAI-10)



- Definitions
  - X^t is a disjoint union of m clusters of state variables X^t_i.

    $X^t = \cup_i X^t_i$, e.g., $X^{90}_i = \{x^{90'}_{Ann}, x^{90'}_{John}, x^{90'}_{Tom}\}$.
  - Any two state variables in a cluster have the same variance and covariances.

    For $x, x' \in X^t_i$, $\sigma^2_{x,x} = \sigma^2_{x',x'}$ and for any y, $\sigma^2_{x,y} = \sigma^2_{x',y}$
  - Property: any multivariate Gaussian of X_t can be represented as a product of pairwise potentials. (UAI-10)

$$P(X_t) \propto \prod_{i,j} \prod_{\substack{x \in X_{t,i} \\ y \in X_{y,j}}} \phi_{PRM_{i,j}}(x,y) \prod_{x \in X_t} \phi_\mu(x)$$