Fundamentals of Probabilistic Data Mining

Chapter I - Introduction

Xavier Alameda-Pineda, Thomas Hueber and Fei Zheng

with the help of Jean-Baptiste Durand

Ensimag/Inria/CNRS

Table of Today's Contents

Course Organisation

Course Content

Fundamentals of Probabilistic Data Mining (FPDM)
is structured into five chapters:

1. Basics of probabilistic models: **conditional dependence**
2. Model-based clustering and **Gaussian** mixture models
3. Probabilistic principal component analysis
4. Sequential data and hidden **Markov** models
5. Approximate inference: variational expectation-maximisation

Course Content

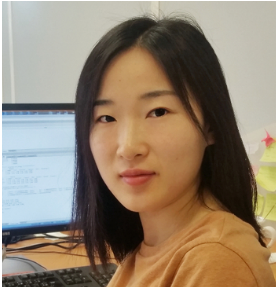
Fundamentals of Probabilistic Data Mining (FPDM)
is structured into five chapters:

1. Basics of probabilistic models: **conditional dependence**
2. Model-based clustering and **Gaussian** mixture models
3. Probabilistic principal component analysis
4. Sequential data and hidden **Markov** models
5. Approximate inference: variational expectation-maximisation

Changes

We added variational inference and reduced the length fo the first chapter. Variational methods (i.e. VAE) are more and more popular, and it is important you know about the basics.

FPDM Team: Fei Zheng



Postdoctoral Researcher
`fei.zheng@inria.fr`
Mistis Team, Inria Grenoble Rhône-Alpes

Time-serie signal processing, anomaly detection for glycemia and artificial pancreas

For FPDM:

- Teaching chapter 2.
- Pratical sessions 1 & 3.

FPDM Team: Thomas Hueber

Research Scientist [website] @thomashueber
`thomas.hueber@gipsa-lab.fr`
Co-head of CRISSP (cognitive robotics, interactive systems, speech processing), GIPSA-Lab, CNRS

Multimodal speech processing, machine learning, interactive systems

For FPDM:

- Teaching chapter 4.
- Pratical session 2.

FPDM Team: Xavier Alameda-Pineda (Xavi)

Research Scientist [webpage] @xavirema
xavier.alameda-pineda@inria.fr
Perception Team, Inria Grenoble Rhône-Alpes

Audio-visual perception, probabilistic and deep learning, human-robot interaction

For FPDM:

- Teaching chapters 1, 3 and 5.
- All pratical sessions.
- Responsible of the subject.

Calendar

When?	What?	Who?	Where?
2-Oct	Ch1: Basics	Xavi	Here
9-Oct	Ch1: Basics	Xavi	
16-Oct	Ch2: GMM	Fei	
23-Oct	Ch2: GMM	Fei	
6-Nov	TP1: GMM & EM	Fei & Xavi	
13-Nov	Ch3: PPCA	Xavi	
20-Nov	Ch4: HMM	Thomas	
27-Nov	Ch4: HMM	Thomas	
4-Dec	TP2: HMM & Speech Recognition	Thomas & Xavi	
11-Dec	Ch5: VEM	Xavi	
18-Dec	Ch5: VEM	Xavi	
8-Jan	TP3: VEM	Fei & Xavi	
22-Jan	Misc	Xavi	

Pre-requisites

- Probability theory: pdf, cdf, expectation, random vectors, correlation, conditional distributions.
- Basic distributions: Gaussian, categorical.
- Linear algebra and multivariate calculus.
- Statistics: maximum likelihood estimator and its properties.
- Constrained optimisation, Lagrange multipliers.
- Bonus: geometric approach of PCA, multiple linear regression.

Rules

- Lab work (LW) & final exam (FE).
- Grade = 0.5*LW+0.5*FE.
- Lab work:
  - Individual pre-work (before the lab).
  - Questions about pre-work will be answered only at the begining of the lab session.
  - Team (3/4 members) post-report.
  - Grade is based on the post-report.
- Final exam: authorized documents (1 sheet of A4 paper).

Support material: slides, scripts, etc. are on `http://chamilo.grenoble-inp.fr/courses/ENSIMAGWMM9AM21/`.

Motivation for Probabilistic Models

What is probabilistic data mining?

- Probabilistic means we **model** our data using probabilities.
- For example in classification, we aim to estimate the posterior probability: $P(c|x)$ for every possible class c.
- Probabilistic generative models & Bayes rule:

$$p(x, c) = p(x|c)p(c) \quad \Rightarrow \quad p(c|x) = \frac{p(x|c)p(c)}{\sum_k p(x|k)p(k)}$$

- What are all these "p"? What do they mean?

Probabilities

For **discrete variables** (i.e. measurable events are discrete):

$$p(c) = P(C = c).$$

$\rightarrow$ the probability of the random variable C to value event c.

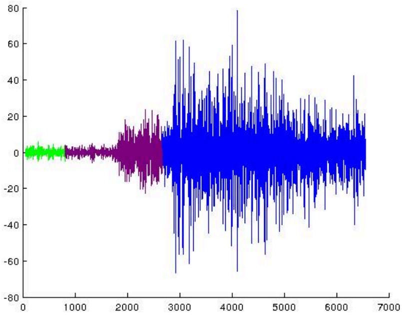
For **continuous variables** (i.e. measurable events are continuous):

$$p(x) = f_X(x) \quad \text{and} \quad p(\mathcal{X}) = P(x \in \mathcal{X}) = \int_{\mathcal{X}} f_X(x)\mathrm{d}x$$

f_X is the probability density function. Remember $P(\{x\}) = 0$.

Why probabilistic data mining?

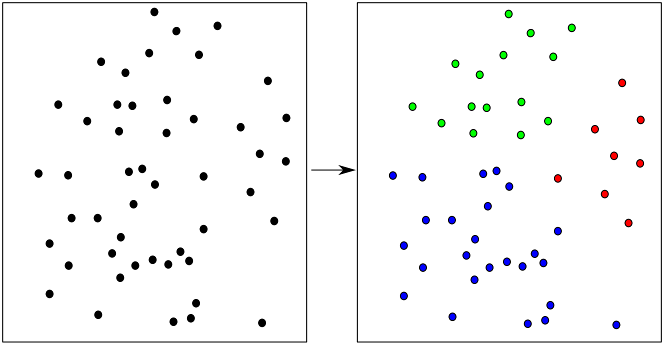
- Infer hidden variables / exploit partly missing data
- Example: clustering, image segmentation
- Incorporate particular requirements in clustering
- Model complex data (on grids, graphs, temporal, ...)
- Simulate phenomena (speech synthesis), make predictions (regime switching in time series)



Segmentation of time series with respect to the variance

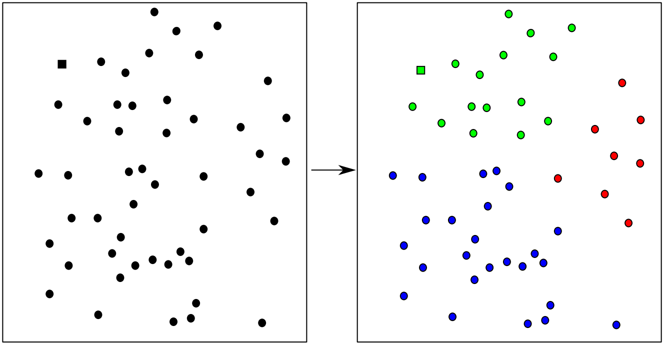
Example (I): Clustering

- Data: points $(x_j)_{j=1,\ldots,n}$ in $\mathbb{R}^d$.
- Aim: find (& predict) clusters.
- Model-based approach: let z_j be the (unknown) cluster of x_j.
  $z_i = z_j \Rightarrow x_i$ and x_j should have the same (conditional) distribution.



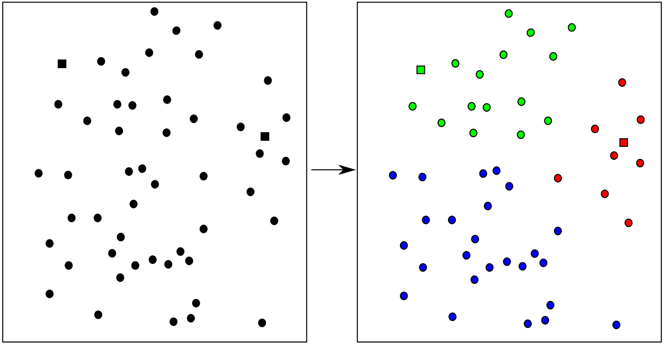
Example (I): Clustering

- Data: points $(x_j)_{j=1,\ldots,n}$ in $\mathbb{R}^d$.
- Aim: find (& predict) clusters.
- Model-based approach: let z_j be the (unknown) cluster of x_j.
  $z_i = z_j \Rightarrow x_i$ and x_j should have the same (conditional) distribution.



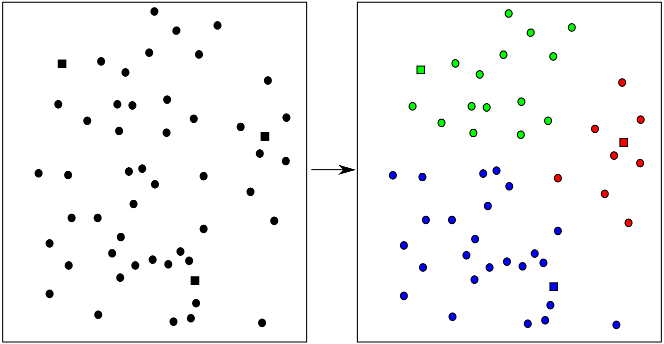
Example (I): Clustering

- Data: points $(x_j)_{j=1,\dots,n}$ in $\mathbb{R}^d$.
- Aim: find (& predict) clusters.
- Model-based approach: let z_j be the (unknown) cluster of x_j.
  $z_i = z_j \Rightarrow x_i$ and x_j should have the same (conditional) distribution.



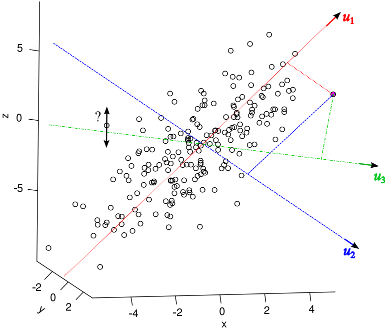
Example (I): Clustering

- Data: points $(x_j)_{j=1,\dots,n}$ in $\mathbb{R}^d$.
- Aim: find (& predict) clusters.
- Model-based approach: let z_j be the (unknown) cluster of x_j.
  $z_i = z_j \Rightarrow x_i$ and x_j should have the same (conditional) distribution.



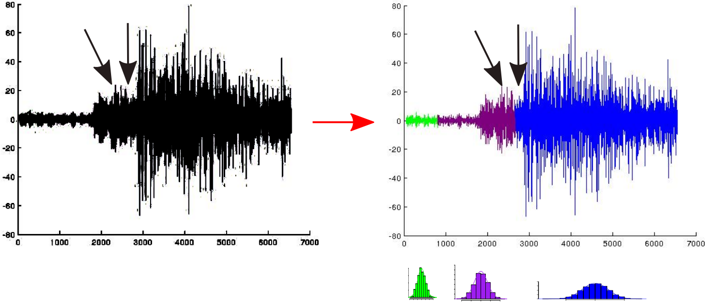
Example (II): Dimensionality reduction

- Raw data are high-dimensional descriptors.
- Difficult to mine patterns/visualize.
- Projection on the directions of maximum variance.
- What if for most or even every point x_j, some coordinates are missing?
- Probabilistic (i.e., model-based) PCA relies on a generative model to exploit partially observed / unknown data.



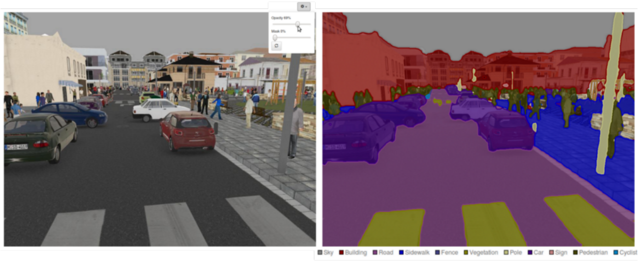
Example (III): Analysis of sequential data

- Special case of clustering with temporal dependencies
- Piecewise statistically invariant features with Markovian jumps
- **Markovian**: it depends only on a few close neighbors.



Example (IV): Inference on complex models

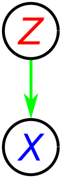
- Modelling complex phenomena requires complex models.
- Exact inference may not be possible.
- Image semantic segmentation: we must infer the class of each pixel (u, v) ($u \in \{1, \ldots, U\}$, $v \in \{1, \ldots, V\}$) of an image. The (hidden) class label at (u, v) depends on the label of its neighbours $\mathcal{N}_{u,v}$. All these cross-dependencies make exact inference intractable.
- Variational inferences provides sound mathematical formulation.



Basics & Conditional Dependency

What is a model?

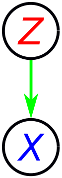
What does it mean to *model* the relationship between two variables?



- *We choose* the nature of Z & X: cont./discrete, bounded, ...
- *We choose* the dependencies, i.e. $p(x, z) = p(x|z)p(z)$.
- *We choose* the distribution $p(z)$.
- *We choose* the distribution $p(x|z)$.

What is a model?

What does it mean to *model* the relationship between two variables?



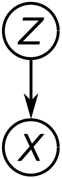
- *We choose* the nature of Z & X: cont./discrete, bounded, ...
- *We choose* the dependencies, i.e. $p(x, z) = p(x|z)p(z)$.
- *We choose* the distribution $p(z)$.
- *We choose* the distribution $p(x|z)$.

Once this is set we can compute:

- The **likelihood** of x: $p(x) = \displaystyle\int_{\mathcal{Z}} p(x|z)p(z)\mathrm{d}z$.
- The **posterior** of z given x: $p(z|x) = \dfrac{p(x|z)p(z)}{p(x)}$.

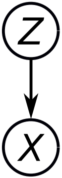
Example: Gaussian mixture model

- The nature: Z is discrete & bounded, X is 1D & continuous.
- The dependencies: $p(x, z) = p(x|z)p(z)$.



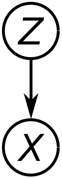
Example: Gaussian mixture model

- The nature: Z is discrete & bounded, X is 1D & continuous.
- The dependencies: $p(x, z) = p(x|z)p(z)$.
- The distribution $p(z)$, $z \in \{1, \ldots, K\}$ is categorical:



$$p(Z = k) = \pi_k, \qquad \pi_k \geq 0, \sum_{k=1}^{K} \pi_k = 1.$$

Example: Gaussian mixture model



- The nature: Z is discrete & bounded, X is 1D & continuous.
- The dependencies: $p(x,z) = p(x|z)p(z)$.
- The distribution $p(z)$, $z \in \{1,\ldots,K\}$ is categorical:

$$p(Z=k) = \pi_k, \qquad \pi_k \geq 0, \ \sum_{k=1}^{K} \pi_k = 1.$$

- The distribution $p(x|z)$ is Gaussian:

$$p(x|Z=k) = \mathcal{N}(x;\mu_k,\nu_k) = \frac{1}{\sqrt{2\pi\nu_k}} \exp\left(-\frac{(x-\mu_k)^2}{2\nu_k}\right)$$

with $\mu_k \in \mathbb{R}$ and $\nu_k > 0$, $\forall k$.

Exercise: likelihood and GMM posteriors

You've got 5 minutes to compute: $p(x)$ and $p(Z = k|x)$. Reminder:

$$p(Z = k) = \pi_k, \qquad \pi_k \geq 0, \ \sum_{k=1}^{K} \pi_k = 1.$$

$$p(x|Z = k) = \mathcal{N}(x; \mu_k, \nu_k) = \frac{1}{\sqrt{2\pi\nu_k}} \exp\Big(- \frac{(x - \mu_k)^2}{2\nu_k} \Big)$$

And we are looking for:

- The **likelihood** of x: $p(x) = \int_{\mathcal{Z}} p(x|z)p(z)\mathrm{d}z$.
- The **posterior** of z given x: $p(Z = k|x) = \dfrac{p(x|Z = k)p(Z = k)}{p(x)}$.

Hint: Just write down what things are.

Solution: likelihood and GMM posteriors

The **likelihood**: $p(x) = \int_{\mathcal{Z}} p(x|z)p(z)\mathrm{d}z$

Solution: likelihood and GMM posteriors

The **likelihood**: $p(x) = \int_{\mathcal{Z}} p(x|z)p(z)\mathrm{d}z = \sum_{k=1}^{K} p(x|Z=k)p(Z=k).$

Solution: likelihood and GMM posteriors

The **likelihood**: $p(x) = \int_{\mathcal{Z}} p(x|z)p(z)\mathrm{d}z = \sum_{k=1}^{K} p(x|Z=k)p(Z=k)$.

And then we replace:

$$p(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\nu_k}} \exp\Big(-\frac{(x-\mu_k)^2}{2\nu_k}\Big)$$

Solution: likelihood and GMM posteriors

The **likelihood**: $p(x) = \int_{\mathcal{Z}} p(x|z)p(z)\mathrm{d}z = \sum_{k=1}^{K} p(x|Z=k)p(Z=k)$.

And then we replace:

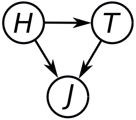
$$p(x) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\nu_k}} \exp\Big(-\frac{(x-\mu_k)^2}{2\nu_k}\Big)$$

The **posterior**:

$$p(Z=k|x) = \frac{\pi_k \frac{1}{\sqrt{2\pi\nu_k}} \exp\Big(-\frac{(x-\mu_k)^2}{2\nu_k}\Big)}{\sum_{m=1}^{K} \pi_m \frac{1}{\sqrt{2\pi\nu_m}} \exp\Big(-\frac{(x-\mu_m)^2}{2\nu_m}\Big)}$$

More on this on the GMM chapter!!!

3-variable models: full

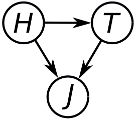


- The amount of sun H.
- The amount of apple trees T.
- The amount of juice J.

The amount of trees clearly depends on the amount of sun.
The amount of juice depends on both sun and trees.

$$p(j, t, h) = \text{You've got 3 minutes.}$$

3-variable models: full

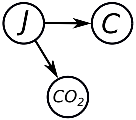


- The amount of sun H.
- The amount of apple trees T.
- The amount of juice J.

The amount of trees clearly depends on the amount of sun.
The amount of juice depends on both sun and trees.

$$p(j, t, h) = p(j|t, h)\ p(t|h)\ p(h)$$

3-variable models: two-kids

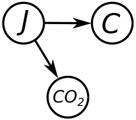


- The amount of juice J.
- The amount of cider C.
- The amount of CO_2.

Both the amount of cider and CO_2 depend on J.

$$p(co_2, c, j) = \text{You've got 3 minutes.}$$

3-variable models: two-kids

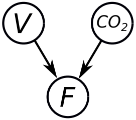


- The amount of juice J.
- The amount of cider C.
- The amount of CO_2.

Both the amount of cider and CO_2 depend on J.

$$p(co_2, c, j) = p(co_2|j)\ p(c|j)\ p(j)$$

3-variable models: two-parents

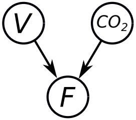


- The container volume V.
- The amount of CO_2.
- The amount of time the fan is switched on F.

F depends on the amount of CO_2 and the volume V.

$$p(f, co_2, v) = \text{You've got 3 minutes.}$$

3-variable models: two-parents

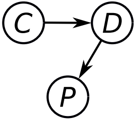


- The container volume V.
- The amount of CO_2.
- The amount of time the fan is switched on F.

F depends on the amount of CO_2 and the volume V.

$$p(f, co_2, v) = p(f|co_2, v)\ p(co_2)\ p(v)$$

3-variable models: cascaded

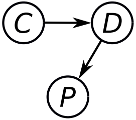


- The amount of cider C.
- How much you dance D.
- The amount of pictures on facebook F the day after.

The amount of dance D is clearly dependent on C, so is F on D.

$$p(f, d, c) = \text{You've got 3 minute.}$$

3-variable models: cascaded



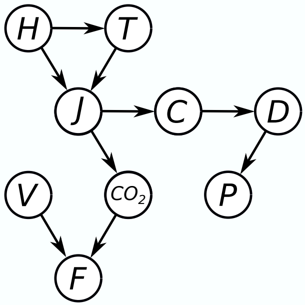
- The amount of cider C.
- How much you dance D.
- The amount of pictures on facebook F the day after.

The amount of dance D is clearly dependent on C, so is F on D.

$$p(f, d, c) = p(f|d)\ p(d|c)\ p(c)$$

n-variable models

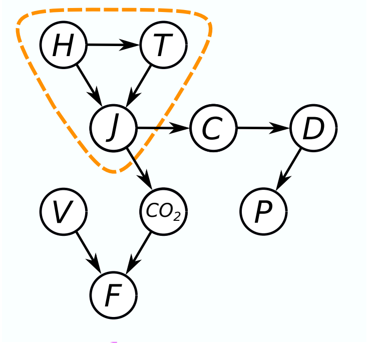
One would like to end up studying models with many variables, but first we need to understand the basics.



Four minimals schemes:

n-variable models

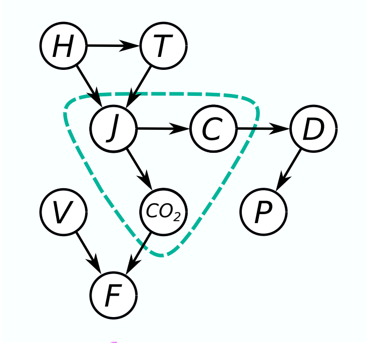
One would like to end up studying models with many variables, but first we need to understand the basics.



Four minimals schemes: Full,

n-variable models

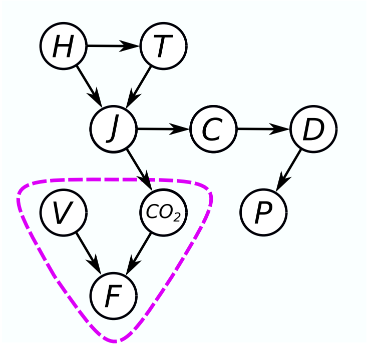
One would like to end up studying models with many variables, but first we need to understand the basics.



Four minimals schemes: Full, Two kids,

n-variable models

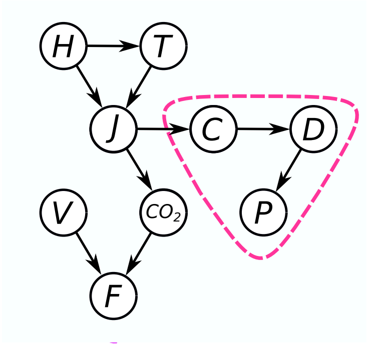
One would like to end up studying models with many variables, but first we need to understand the basics.



Four minimals schemes: Full, Two kids, Two parents

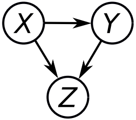
n-variable models

One would like to end up studying models with many variables, but first we need to understand the basics.



Four minimals schemes: Full, Two kids, Two parents and Cascaded.

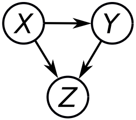
3-variable models: summary



Full all dependencies are set:

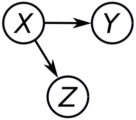
$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

3-variable models: summary



Full all dependencies are set:

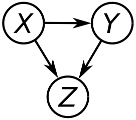
$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$



Two kids Y-Z dependency is off:

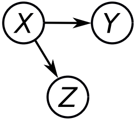
$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

3-variable models: summary



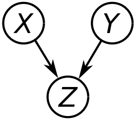
Full all dependencies are set:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$



Two kids Y-Z dependency is off:

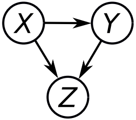
$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$



Two parents X-Y dependency is off:

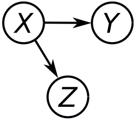
$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

3-variable models: summary



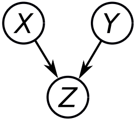
Full all dependencies are set:

$$p(x,y,z) = p(z|x,y)p(y|x)p(x)$$



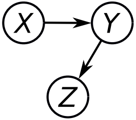
Two kids Y-Z dependency is off:

$$p(x,y,z) = p(z|x,y)p(y|x)p(x)$$



Two parents X-Y dependency is off:

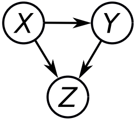
$$p(x,y,z) = p(z|x,y)p(y|x)p(x)$$



Cascaded X-Z dependency is off:

$$p(x,y,z) = p(z|x,y)p(y|x)p(x)$$

3-variable models: let's play! (I)



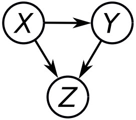
Full all dependencies are set:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation). You've got 5 minutes

Hint: use the joint probability.

3-variable models: let's play! (I)



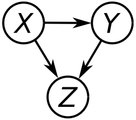
Full all dependencies are set:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)}$$

3-variable models: let's play! (I)



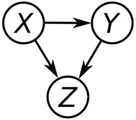
Full all dependencies are set:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(z|x, y)p(y|x)p(x)}{}$$

3-variable models: let's play! (I)



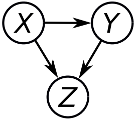
Full all dependencies are set:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(z|x, y)p(y|x)p(x)}{\int_{\mathcal{X}\times\mathcal{Y}} p(x', y', z)\mathrm{d}x'\mathrm{d}y'}$$

3-variable models: let's play! (I)



Full all dependencies are set:

$$p(x, y, z) = p(z|x, y)p(y|x)p(x)$$

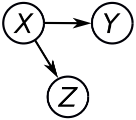
Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(z|x, y)p(y|x)p(x)}{\int_{\mathcal{X}\times\mathcal{Y}} p(x', y', z)\mathrm{d}x'\mathrm{d}y'}$$

with

$$\int_{\mathcal{X}\times\mathcal{Y}} p(x', y', z)\mathrm{d}x'\mathrm{d}y' = \int_{\mathcal{X}\times\mathcal{Y}} p(z|x', y')p(y'|x')p(x')\mathrm{d}x'\mathrm{d}y'$$

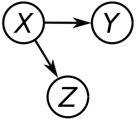
3-variable models: let's play! (II)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x).$

Let's find $p(y|z)$ in terms of the model's distributions (i.e. previous equation). You've got 5 minutes

3-variable models: let's play! (II)

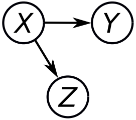


Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x).$

Let's find $p(y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(y|z) = \frac{p(y, z)}{p(z)}$$

3-variable models: let's play! (II)

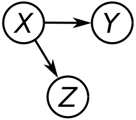


Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x).$

Let's find $p(y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(y|z) = \frac{p(y, z)}{p(z)} = \frac{\int_{\mathcal{X}} p(x, y, z)\mathrm{d}x}{\int_{\mathcal{X}\times\mathcal{Y}} p(x', y', z)\mathrm{d}x'\mathrm{d}y'}$$

3-variable models: let's play! (II)



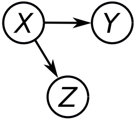
Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$.

Let's find $p(y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(y|z) = \frac{p(y, z)}{p(z)} = \frac{\int_{\mathcal{X}} p(x, y, z)\mathrm{d}x}{\int_{\mathcal{X}\times\mathcal{Y}} p(x', y', z)\mathrm{d}x'\mathrm{d}y'}$$

with $\int_{\mathcal{X}} p(x, y, z)\mathrm{d}x = \int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x$

3-variable models: let's play! (II)



Two-kids $p(x,y,z) = p(z|x)p(y|x)p(x).$

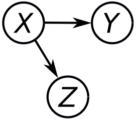
Let's find $p(y|z)$ in terms of the model's distributions
(i.e. previous equation).

$$p(y|z) = \frac{p(y,z)}{p(z)} = \frac{\int_{\mathcal{X}} p(x,y,z)\mathrm{d}x}{\int_{\mathcal{X}\times\mathcal{Y}} p(x',y',z)\mathrm{d}x'\mathrm{d}y'}$$

with $\int_{\mathcal{X}} p(x,y,z)\mathrm{d}x = \int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x$

and $\int_{\mathcal{X}\times\mathcal{Y}} p(x',y',z)\mathrm{d}x'\mathrm{d}y' = \int_{\mathcal{X}\times\mathcal{Y}} p(z|x')p(y'|x')p(x')\mathrm{d}x'\mathrm{d}y'$

3-variable models: let's play! (II)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$.

Let's find $p(y|z)$ in terms of the model's distributions
(i.e. previous equation).

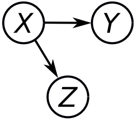
$$p(y|z) = \frac{p(y,z)}{p(z)} = \frac{\int_{\mathcal{X}} p(x,y,z)\mathrm{d}x}{\int_{\mathcal{X}\times\mathcal{Y}} p(x',y',z)\mathrm{d}x'\mathrm{d}y'}$$

with $\int_{\mathcal{X}} p(x,y,z)\mathrm{d}x = \int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x$

and $\int_{\mathcal{X}\times\mathcal{Y}} p(x',y',z)\mathrm{d}x'\mathrm{d}y' = \int_{\mathcal{X}\times\mathcal{Y}} p(z|x')p(y'|x')p(x')\mathrm{d}x'\mathrm{d}y'$

$= \int_{\mathcal{X}} p(z|x')p(x')\Big(\int_{\mathcal{Y}} p(y'|x')\mathrm{d}y'\Big)\mathrm{d}x' = \int_{\mathcal{X}} p(z|x')p(x')\mathrm{d}x'$

3-variable models: let's play! (II)



Two-kids $p(x,y,z) = p(z|x)p(y|x)p(x)$.

Let's find $p(y|z)$ in terms of the model's distributions
(i.e. previous equation).

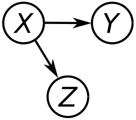
$$p(y|z) = \frac{p(y,z)}{p(z)} = \frac{\int_{\mathcal{X}} p(x,y,z)\mathrm{d}x}{\int_{\mathcal{X}\times\mathcal{Y}} p(x',y',z)\mathrm{d}x'\mathrm{d}y'} = \frac{\int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x}{\int_{\mathcal{X}} p(z|x')p(x')\mathrm{d}x'}$$

with $\int_{\mathcal{X}} p(x,y,z)\mathrm{d}x = \int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x$

and $\int_{\mathcal{X}\times\mathcal{Y}} p(x',y',z)\mathrm{d}x'\mathrm{d}y' = \int_{\mathcal{X}\times\mathcal{Y}} p(z|x')p(y'|x')p(x')\mathrm{d}x'\mathrm{d}y'$

$= \int_{\mathcal{X}} p(z|x')p(x')\Big(\int_{\mathcal{Y}} p(y'|x')\mathrm{d}y'\Big)\mathrm{d}x' = \int_{\mathcal{X}} p(z|x')p(x')\mathrm{d}x'$

3-variable models: let's play! (II)



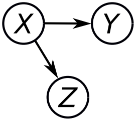
Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$.

Let's find $p(y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(y|z) = \frac{\int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x}{\int_{\mathcal{X}} p(z|x')p(x')\mathrm{d}x'}$$

Not done yet, let's find $p(y|z, x)$: You've got 2 minutes

3-variable models: let's play! (II)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$.

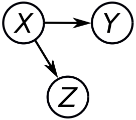
Let's find $p(y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(y|z) = \frac{\int_{\mathcal{X}} p(z|x)p(y|x)p(x)\mathrm{d}x}{\int_{\mathcal{X}} p(z|x')p(x')\mathrm{d}x'}$$

Not done yet, let's find $p(y|z, x)$:

$$p(y|z, x) = \frac{p(y, z, x)}{p(z, x)} = \frac{p(y|x)p(z|x)p(x)}{p(z|x)p(x)} = p(y|x)$$

3-variable models: let's play! (III)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$.

It is now easy to see:

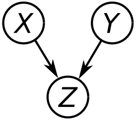
$$p(y, z|x) = p(y|x)p(z|x) \quad \text{and} \quad p(y, z) \neq p(y)p(z).$$

While Y and Z are not independent,
they are **conditionally independent given** X:

$$Y \perp\!\!\!\perp Z \mid X$$

"If we know X, Z gives no information about Y (and viceversa)."

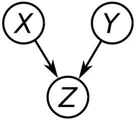
3-variable models: let's play! (IV)



Two-parents $p(x, y, z) = p(z|x, y)p(x)p(y)$.

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation). You've got 5 minutes

3-variable models: let's play! (IV)

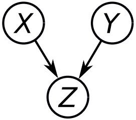


Two-parents $p(x, y, z) = p(z|x, y)p(x)p(y)$.

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)}$$

3-variable models: let's play! (IV)

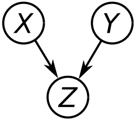


Two-parents $p(x, y, z) = p(z|x, y)p(x)p(y)$.

Let's find $p(x, y|z)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(z|x, y)p(x)p(y)}{\int_{\mathcal{X}\times\mathcal{Y}} p(z|x', y')p(x')p(y')\mathrm{d}x'\mathrm{d}y'}.$$

3-variable models: let's play! (IV)



Two-parents $p(x, y, z) = p(z|x, y)p(x)p(y)$.

Let's find $p(x, y|z)$ in terms of the model's distributions
(i.e. previous equation).

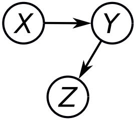
$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(z|x, y)p(x)p(y)}{\int_{\mathcal{X}\times\mathcal{Y}} p(z|x', y')p(x')p(y')\mathrm{d}x'\mathrm{d}y'}.$$

We can also compute:

$$p(x, y) = p(x)p(y).$$

X and Y are independent, but not conditionally independent given Z.

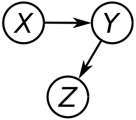
3-variable models: let's play! (V)



Cascaded $p(x, y, z) = p(z|y)p(y|x)p(x)$.

Let's find $p(x, z|y)$ in terms of the model's distributions (i.e. previous equation). You've got 5 minutes

3-variable models: let's play! (V)

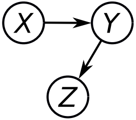


Cascaded $p(x, y, z) = p(z|y)p(y|x)p(x)$.

Let's find $p(x, z|y)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)}$$

3-variable models: let's play! (V)

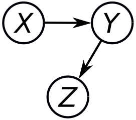


Cascaded $p(x, y, z) = p(z|y)p(y|x)p(x).$

Let's find $p(x, z|y)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(z|y)p(y|x)p(x)}{p(y)}$$

3-variable models: let's play! (V)



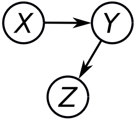
Cascaded $p(x, y, z) = p(z|y)p(y|x)p(x)$.

Let's find $p(x, z|y)$ in terms of the model's distributions (i.e. previous equation).

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(z|y)p(y|x)p(x)}{p(y)} = p(z|y)p(x|y).$$

[$p(y)$ can be computed by marginalising $p(y|x)p(x)$ w.r.t. X]

3-variable models: let's play! (V)



Cascaded $p(x, y, z) = p(z|y)p(y|x)p(x).$

Let's find $p(x, z|y)$ in terms of the model's distributions
(i.e. previous equation).

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(z|y)p(y|x)p(x)}{p(y)} = p(z|y)p(x|y).$$

[$p(y)$ can be computed by marginalising $p(y|x)p(x)$ w.r.t. X]

This expression is semantically more close to "conditional independence".

One can easily see that X and Z are not independent.

Conditional Independence: wrap up

Definition

Let X, Y, and Z be random variables, we say that X and Y are **conditionally independent** given Z, and write $X \perp\!\!\!\perp Y \mid Z$, iff one of the following equivalent expressions holds:

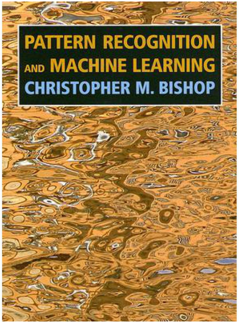
- $p(x, y|z) = p(x|z)p(y|z)$
- $p(x|y, z) = p(x|z)$
- $p(y|x, z) = p(y|z)$

Proof of the equivalence: In class or let as exercise.

References

There is **a lot** of bibliography on probabilistic graphical models.

I strongly suggest the following book:



Pattern Recognition and Machine Learning, from Christopher M. Bishop (Springer)

The concepts discussed in FPDM correspond to different parts of Ch. 2, 8, 9, 10, 12, 13.

Next session is in a different class-room!