Fundamentals of Probability

Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-635

9/6/2024

Independence

$$\forall x \in \mathrm{x}, y \in \mathrm{y},\ p(\mathrm{x} = x, \mathrm{y} = y) = p(\mathrm{x} = x)p(\mathrm{y} = y)$$

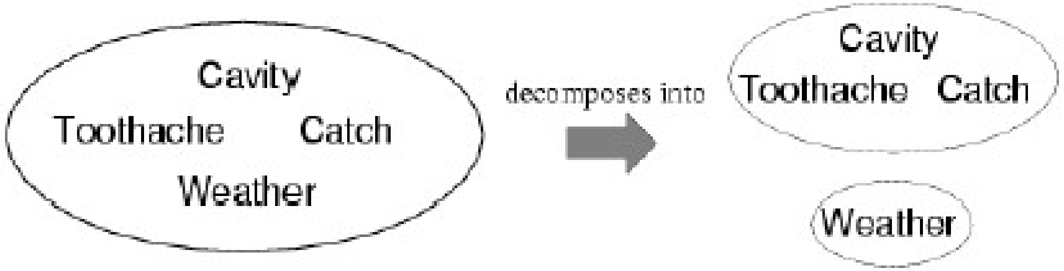
Two events are **independent**, **statistically independent**, or **stochastically independent** if occurrence of one does not affect probability of occurrence of the other

Similarly, two random variables are independent ***if*** realization of one does not affect probability distribution of other

(Absolute) Independence

A and *B* are independent iff (Note: following are equivalent)

$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$



$\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$= \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$

32 (2^3 * 4 (*Weather*)) entries reduced to 12 (2^3 + 4 (*Weather*))

Absolute independence is powerful, but rare.

For your review!

Conditional Independence

$$\forall x \in \mathrm{x}, y \in \mathrm{y}, z \in \mathrm{z},\ p(\mathrm{x}=x, \mathrm{y}=y \mid \mathrm{z}=z) = p(\mathrm{x}=x \mid \mathrm{z}=z)p(\mathrm{y}=y \mid \mathrm{z}=z)$$

Two events *x* and *y* are **conditionally independent** given a third event *z* if occurrence of *x* *and* occurrence of *y* are independent events in their conditional probability distribution given *z*

In other words, *x* and *y* are conditionally independent given *z* if and only if (***iff***), given knowledge that *z* occurs, knowledge of whether *x* occurs provides no information on likelihood of *y* occurring, and knowledge of whether *y* occurs provides no information on likelihood of *x* occurring

Conditional Independence

If I have a cavity, probability the probe catches doesn't depend on whether I have a toothache:

(1) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$

The same independence holds if I haven't got a cavity:

(2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$

Catch is conditionally independent of *Toothache* given *Cavity*:

(3) $\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)$

Equivalent statements:

$\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$

$\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$

For your review!

Conditional independence

We can now write out the full joint distribution as:

P(*Toothache, Catch, Cavity*)
= **P**(*Toothache, Catch* | *Cavity*) **P**(*Cavity*) // *product rule*
= **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*) **P**(Cavity) // *cond. ind.*

In many cases

Use of conditional independence reduces the size of the representation of the joint distribution from exponential in *n* to linear in *n*

Conditional independence

Our most basic and robust form of knowledge about uncertain environments

For your review!

Bayes, in English Please?

- What does Bayes' Formula helps to find?
  - Helps us to find:

$$P(B \mid A)$$

  - By having already known:

$$P(A \mid B)$$

$$P(\mathrm{x} \mid \mathrm{y}) = \frac{P(\mathrm{x})P(\mathrm{y} \mid \mathrm{x})}{P(\mathrm{y})}$$

Thomas Bayes, 1701-1761

Bayes, in English Please?

- What does Bayes' Formula helps to find?
  - Helps us to find:

$$P(B \mid A$$

Bayes Theorem shall return!

  - By having alread own:

$$P(A \mid B)$$

$$) = \frac{P(x)P(y \mid x)}{P(y)}$$

Thomas Bayes, 1701-1761

Bayes' Rule

Product rule: $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

$\Rightarrow$ Bayes' rule: $P(a \mid b) = P(b \mid a)\, P(a) / P(b)$

or in distribution form

$$\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\, \mathbf{P}(Y) / \mathbf{P}(X) = \alpha\, \mathbf{P}(X|Y)\, \mathbf{P}(Y)$$

Causal Probability (useful for diagnostics):

P(Cause | Effect) = P(Effect|Cause) P(Cause) / P(Effect)

E.g., let *M* be meningitis, *S* be stiff neck:

$$P(m|s) = P(s|m)\, P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$$

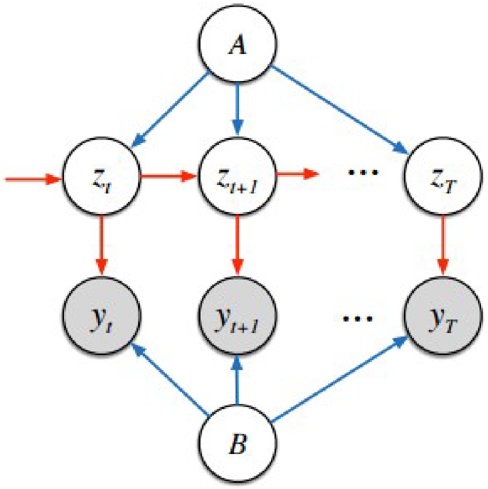
Note: posterior probability of meningitis still very small!

Summary of Probability

- Probability is rigorous formalism for *uncertain* knowledge
- Joint probability distribution specifies probability of every atomic event (in sample/event space)
- *Queries* can be answered by summing over atomic events
- For nontrivial domains, we must find ways to reduce joint probability distributional search space
  - *Independence* & *conditional independence* = your tools for reducing joint probability distribution table size

- ***Note***: These ideas/axioms apply equally well to vector/matrix variables

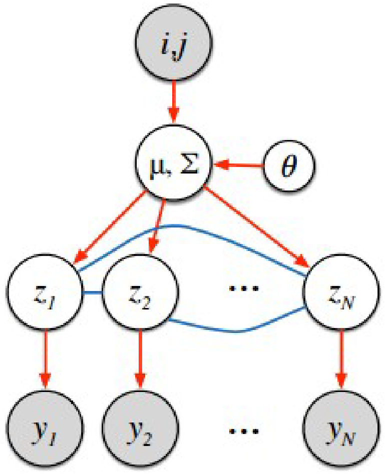
Probability allows us to build models of stochastic, data-generating processes….



Gaussian Linear State Space Model
Kalman Filter

$$z_t \sim \mathcal{N}(z_t | A z_{t-1}, \sigma_z^2 I)$$

$$y_t \sim \mathcal{N}(y_t | B z_t, \sigma_y^2 I)$$

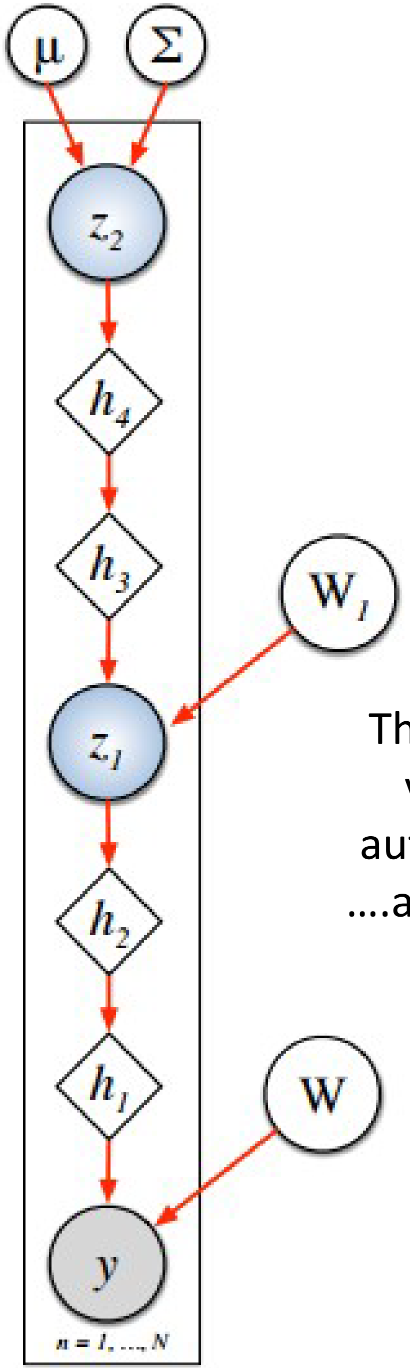


Latent Gaussian Cox Point Process

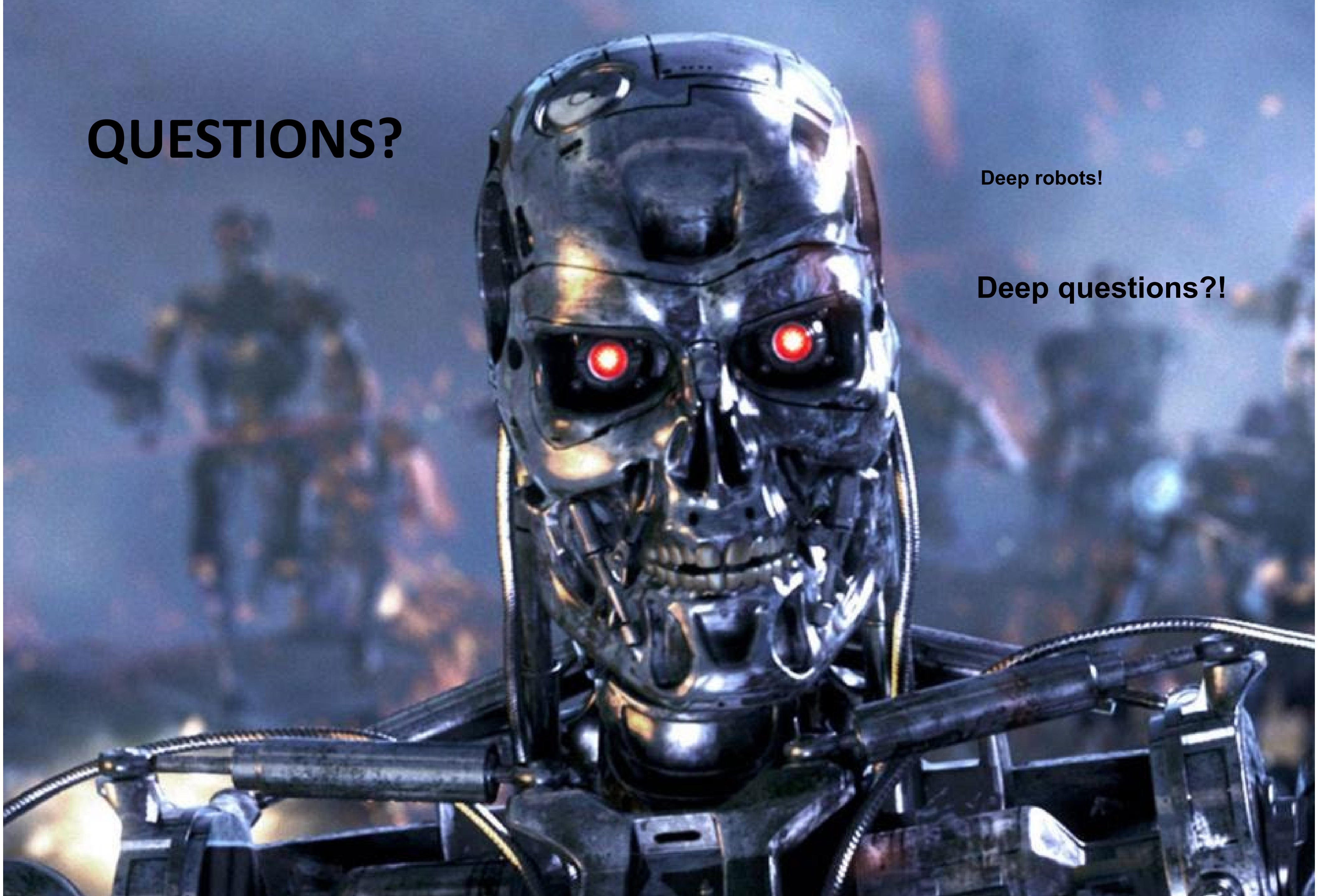
$$x \sim \mathcal{N}(x | \mu(i,j), \Sigma(i,j))$$

$$y_{ij} \sim \mathcal{P}(c \exp(x_{ij}))$$

Probabilistic graphical models (PGMs)



The venerable variational autoencoder….
….a neural PGM.


QUESTIONS?
Deep robots!
Deep questions?!