CS6462

Probabilistic and Explainable AI

Lesson 22

Causal Inference

by Emil Vassev

March 24, 2025

Causal Concepts

Specifics:

- *standard statistical analysis*:
  - assess parameters of a distribution from samples drawn of that distribution
  - probabilities (prior and posterior) are estimated based on inference techniques working on associations among random variables
  - works well under static conditions – parameters are not dynamic
- *causal analysis*: aims to infer probabilities under dynamic conditions (parameters)

Causal and associational concepts do not mix:

- associational concept is any relationship that can be defined in terms of a joint distribution of observed variables: *correlation*, *regression*, *dependence*, *conditional independence*, *likelihood*, etc.
- causal concept is any relationship that cannot be defined in terms of a distribution function: *structural coefficients*, *spurious correlation*, *stability*, *instrumental variables*, etc.
- probability calculus is insufficient to express causal relations

Causality

Specifics:

- questions to be considered:
  - What type of evidence can we use to establish causality?
  - What is enough evidence of the existence of a causal relationship?
- *Common Cause Principle**:
  - if two observables (random variables) ***X*** and ***Y*** are statistically dependent, then there exists a variable ***Z*** that:
    - causally influences both ***X*** and ***Y***
    - explains their dependence by considering them independent when conditioned on ***Z***
  - special case: ***Z*** can coincide with ***X*** or ***Y***

H. Reichenbach. 1956. The Direction of Time. University of California Press, Berkeley, CA.

Causality (cont.)

Example:

- X is the frequency of storks
- Y the human birth rate (in European countries, these have been reported to be correlated)

1) if storks bring babies, then the correct causal graph is $X \rightarrow Y$
2) if babies attract storks, it is $X \leftarrow Y$
3) if there is some other variable that causes both, e.g., economic development Z, we have:

   $X \leftarrow Z \rightarrow Y$

   X and Y are independent, but conditioned on Z

A causal model thus contains genuinely more information than a statistical one.

Structural Causal Model

Mathematical Representation: causally sufficient model

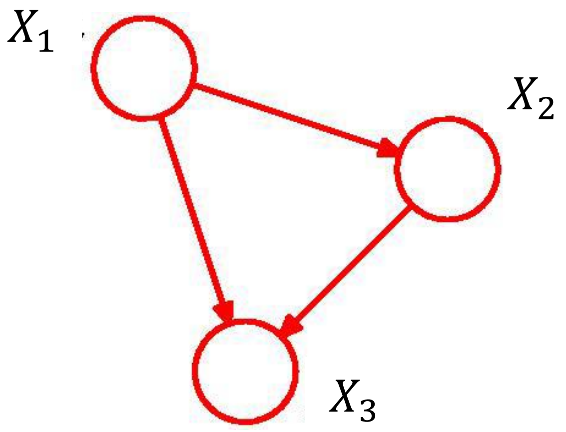
- $X = \{X_1, X_2, \ldots, X_n\}$ – set of observables (modelled as random variables)
- G – DAG, X associated with the vertices of G
- each observable X is the result of an assignment:
  $X_i = f_i(PA_i, U_i), \mathrm{i} = 1 \ldots \mathrm{n}$

  f_i - deterministic function

  PA_i - parents of X_i

  U_i - stochastic unexplained variable, noise

  $U = \{U_1, U_2, \ldots, U_n\}$ – jointly independent (requirement for causal sufficiency)
- directed edges in G represent direct causation, since parents PA_i are connected to X_i by directed edges they directly affect the assignment of X_i
- $p(X_i|PA_i)$ - general conditional distribution ensured by the noise U_i



Structural Causal Model (cont.)

Mathematical Representation (cont.): causally sufficient model

- $U = \{U_1, U_2, \dots, U_n\}$ – (information probes) if specified, then we can compute:
  - recursively X_i $\boxed{X_i = f_i(PA_i, U_i)}$
  - joint distribution $p(X)$, $X = \{X_1, X_2, \dots, X_n\}$
- $p(X)$ has structural properties inherited from the graph: it satisfies the *Causal Markov Condition* – conditioned on its parents, each X_i is independent of its non-descendants
- *interventions*: modify subset of X_i - change U_i or set f_i to a constant
- *causal (disentangled) factorization*:

$$p(X) = P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} p(X_i|PA_i) = \prod_{i=1}^{n} p(f_i(PA_i, U_i)|PA_i)$$

  - decomposes the joint distribution into conditionals corresponding to the structural assignments
- *causal learning*: seeks to exploit the fact that the joint distribution possesses a causal factorization

Independent Causal Mechanisms

Independent Causal Mechanisms (ICM) Principle:*

1) the causal generative process of system's variables is composed of autonomous modules that do not inform or influence each other

2) in the probabilistic case, the conditional distribution of each variable given its causes does not inform or influence the other variables' conditional distribution

- applied to the causal factorization (*stability* or *invariance*):

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} p(X_i|PA_i)$$

  - the factors should be independent in the sense that:

  1) changing (or intervening upon) one mechanism $p(X_i|PA_i)$ does not change the other mechanisms $p(X_j|PA_j),\ i \neq j$
  2) knowing some other mechanisms $p(X_i|PA_i)$, does not give us information about a mechanism $p(X_j|PA_j),\ i \neq j$

B. Schölkopf et al. 2012. On causal and anticausal learning. Proceedings of the 29th International Conference on Machine Learning (ICML), pp. 1255–1262.

Causal Graphical Models in Python

Example: the sprinkler.

It is a system of five variables which indicate the conditions on a certain day:

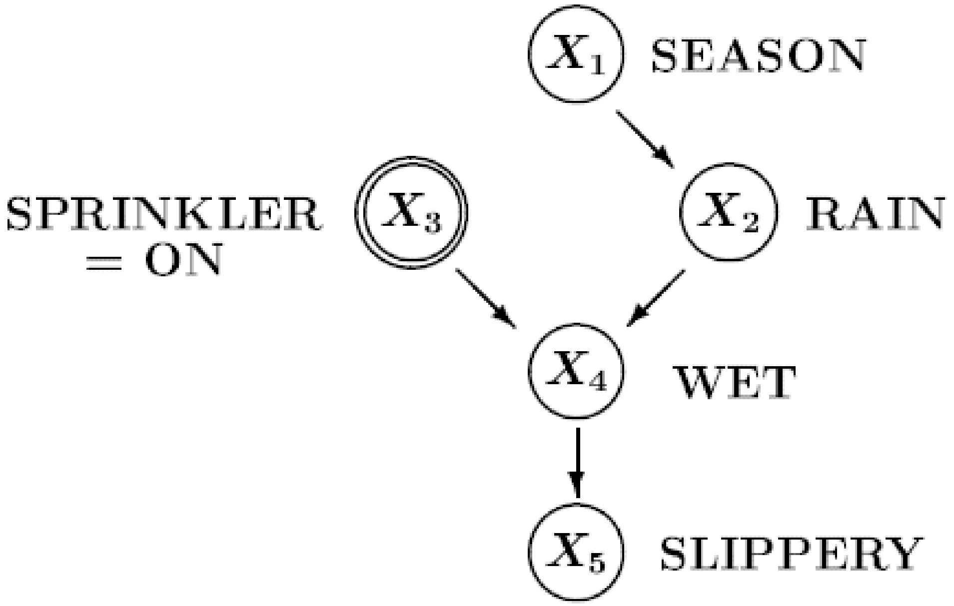
X_1 *season*: indicates the current season

X_2 *rain*: indicates whether it is raining

X_3 *sprinkler*: indicates whether our sprinkler is on

X_4 *wet*: indicates whether the ground is wet

X_5 *slippery*: indicates whether the ground is slippery



Causal Graphical Models in Python (cont.)

Implementation: causalgraphicalmodels

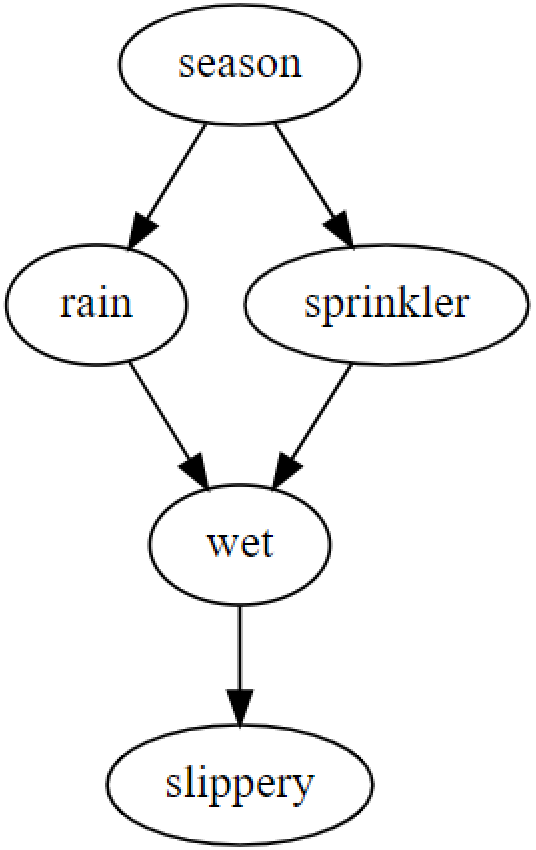
CausalGraphicalModel – creates DAG:

- *nodes*
- *edges*

```python
from causalgraphicalmodels import CausalGraphicalModel

sprinkler = CausalGraphicalModel(
    nodes=["season", "rain", "sprinkler", "wet", "slippery"],
    edges=[
        ("season", "rain"),
        ("season", "sprinkler"),
        ("rain", "wet"),
        ("sprinkler", "wet"),
        ("wet", "slippery")
    ]
)

# draw return a graphviz `dot` object, which jupyter can render
sprinkler.draw()
```



Causal Graphical Models in Python (cont.)

Implementation (cont): causalgraphicalmodels

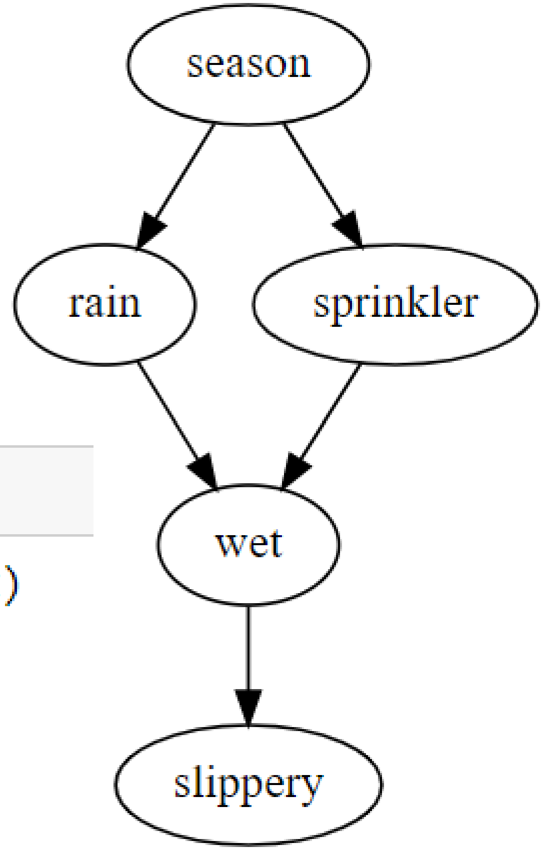
Get the joint distribution:

- *get_distribution*

$$P(X_1, X_2, X_3, X_4, X_5) = \prod_{i=1}^{5} p(X_i|PA_i)$$

```
print(sprinkler.get_distribution())
```

```
P(season)P(rain|season)P(sprinkler|season)P(wet|rain,sprinkler)P(slippery|wet)
```



Causal Graphical Models in Python (cont.)

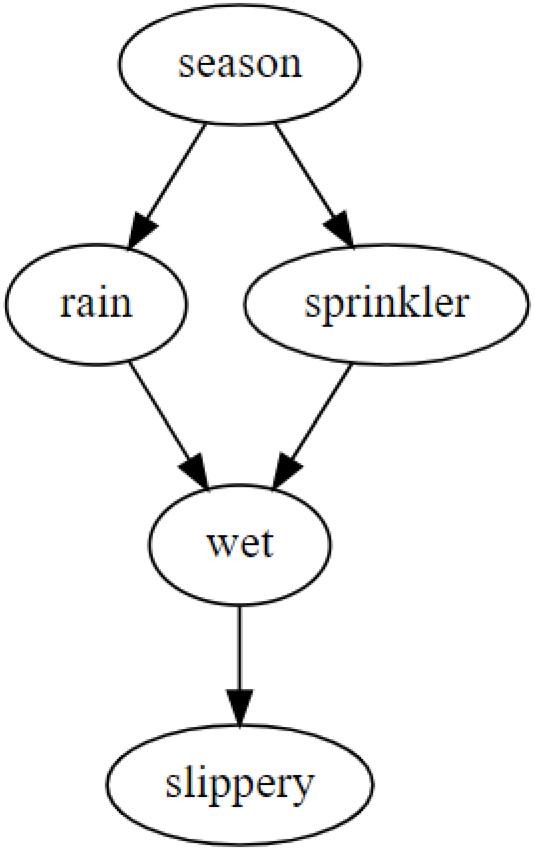
Implementation (cont): causalgraphicalmodels

Get the independence relationships:

- *get_all_independence_relationships()*

```
sprinkler.get_all_independence_relationships()
```

```
[('slippery', 'season', {'wet'}),
 ('slippery', 'season', {'rain', 'wet'}),
 ('slippery', 'season', {'rain', 'sprinkler'}),
 ('slippery', 'season', {'sprinkler', 'wet'}),
 ('slippery', 'season', {'rain', 'sprinkler', 'wet'}),
 ('slippery', 'rain', {'wet'}),
 ('slippery', 'rain', {'sprinkler', 'wet'}),
 ('slippery', 'rain', {'season', 'wet'}),
 ('slippery', 'rain', {'season', 'sprinkler', 'wet'}),
 ('slippery', 'sprinkler', {'wet'}),
 ('slippery', 'sprinkler', {'rain', 'wet'}),
 ('slippery', 'sprinkler', {'season', 'wet'}),
 ('slippery', 'sprinkler', {'rain', 'season', 'wet'}),
 ('season', 'wet', {'rain', 'sprinkler'}),
 ('season', 'wet', {'rain', 'slippery', 'sprinkler'}),
 ('rain', 'sprinkler', {'season'})]
```



Causal Graphical Models in Python (cont.)

Implementation (cont): causalgraphicalmodels

Interventions – operations that modify X_i*:*

- *do()*

What does happen when an intervention occurs?

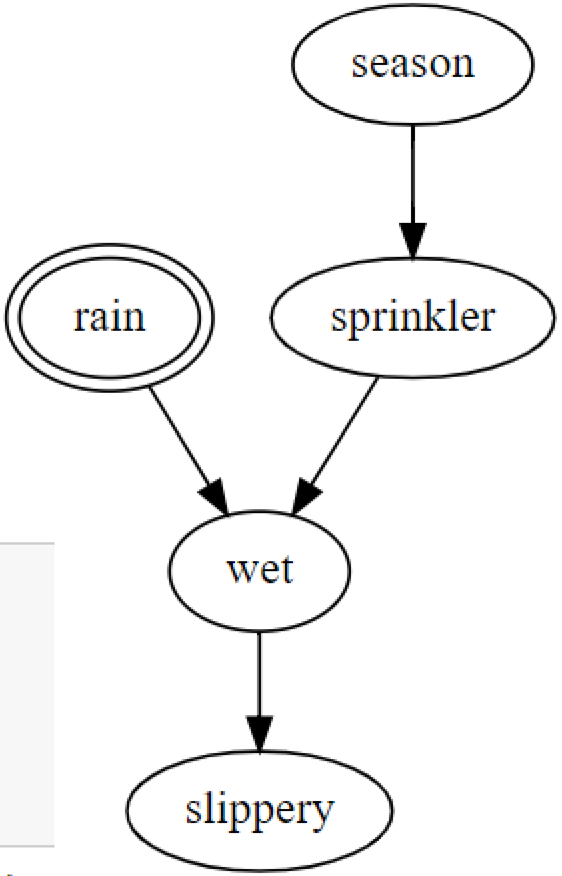
- removes the edges between the variable and it's parents

```
sprinkler_do = sprinkler.do("rain")

print(sprinkler_do.get_distribution())

sprinkler_do.draw()
```

```
P(season)P(sprinkler|season)P(wet|do(rain),sprinkler)P(slippery|wet)
```



- old:

```
P(season)P(rain|season)P(sprinkler|season)P(wet|rain,sprinkler)P(slippery|wet)
```

Summary

Causal Inference

Causal Concepts

Causality

Structural Causal Model

Independent Causal Mechanisms

Causal Graphical Models in Python

Next Lesson:

- Meta-Learning Causal Structures

Thank You!

Questions?