CS 344 – Guide 5

- Bayesian Networks:
  - Definition: a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).
  - Be familiar with the structure of Bayesian networks:
    - Nodes represent variables – observable quantities, latent variables, unknown parameters or hypotheses.
    - Edges represent conditional dependencies.
    - Unconnected nodes represent variables that are conditionally independent of each other.
    - Directed acyclic graphs:
      - A finite directed graph with no directed cycles.
      - Consists of finitely many vertices and edges such that there is no way to start at any vertex v and follow a consistently-directed sequence of edges that eventually loops back to v again.
    - Semantics (see "factorization definition" section)
      - Factorization definition:
        - X is a Bayesian network with respect to G if its joint probability density function (with respected to a product measure) can be written as a product of the individual density functions, conditional on their parent variables.
        - p(x) = ITP (xv I Xpa{v))
          - vEV
        - where pa(v) is the set of parents of v (i.e. those vertices pointing directly to v via a single edge)

Factorization definition [edit]

X is a Bayesian network with respect to *G* if its joint probability density function (with respect to a product measure) can be written as a product of the individual density functions, conditional on their parent variables:[

$$p(x) = \prod_{v \in V} p\left(x_v \,\middle|\, x_{\mathrm{pa}(v)}\right)$$

where pa(*v*) is the set of parents of *v* (i.e. those vertices pointing directly to *v* via a single edge).

For any set of random variables, the probability of any member of a joint distribution can be calculated from conditional probabilities using the chain rule (given a topological ordering of *X*) as follows:[16]

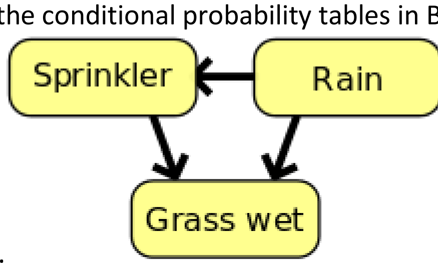
$$\mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{v=1}^{n} \mathrm{P}\left(X_v = x_v \mid X_{v+1} = x_{v+1}, \ldots, X_n = x_n\right)$$

Using the definition above, this can be written as:

$$\mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{v=1}^{n} \mathrm{P}(X_v = x_v \mid X_j = x_j \text{ for each } X_j \text{ which is a parent of } X_v)$$

The difference between the two expressions is the conditional independence of the variables from any of their non-descendants, given the values of their parent variables.

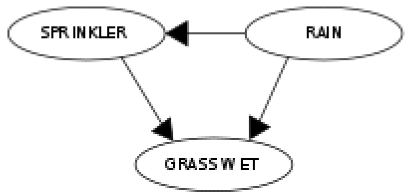
  - Compare and contrast the conditional probability tables in Bayes Networks with the full joint



probability distribution.
  - E.g. see the rain-sprinkler example:
    - How many values must it store?
      - 14 values.
    - How many would the full joint have to store?

- 2^n = 2^3 = 8

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

-
- What benefit is there to using Bayesian networks as opposed to the probabilistic mechanism discussed in the previous unit?
  - Computationally far less expensive the more variables you have as the full joint requires n-dimensions for n-variables, leading to 2^n exponential run-time.

- Scikit Learn
  - ^_^