UAI Formula Sheet

Fuzzy Sets

Fuzzy Set: A set whos membership function range on interval $[0, 1]$

Membership Function: Defins a set by defining the degree of membership of an element of the universe of discourse.

Frame of Cognition: A set of fuzzy sets fully covering the universe of discourse(the range of variables)
- Coverage
- Unimodality

Fuzzy Partition: A frame of cognition for which the sum of the membership values of each value of the base variable is equal to 1.

α-Cut: Is the crisp set of the values of x such that $\mu(X) \geq \alpha$

$$\alpha\text{-cut}_\mu(X) = \{\mu(x) \geq \alpha\}$$

Support: The crisp set of values $x \in X$ such that $\mu(x) > 0$

Height of a Fuzzy Set: The height of a fuzzy set f is the highest membership degree of an element x of the fuzzy set.

$$h_f(X) = \max_{x \in X} \mu_f(x)$$

Normal Fuzzy Set: A fuzzy set is normal $\iff h_f(X) = 1$

Convex Fuzzy Set: A fuzzy set is convex $\iff \forall x_1, x_2 \in X, \forall \lambda \in [0, 1]$

$$\mu_f(\lambda x_1 + (1 - \lambda)x_2) \geq \min(\mu_f(x_1), \mu_f(x_2))$$

Fuzzy Logic

Fuzzy Logic: Infinite-valued logic, with truth values is $[0..1]$ A is L Where:
- A is a linguistic variable
- L is a label of a fuzzy set

Linguistic Variable: A is $(X, T(X), U, G, M)$ Where:
- X is the universe of discourse
- $T(X)$ is the set of linguistic terms
- U is the set of values of the linguistic variable
- G is the semantic rule
- M is the mapping of the semantic rule

Fuzzy Modifiers: A fuzzy modifier is a linguistic variable that is used to modify the meaning of a linguistic variable.
- Strong Modifier - $m(a) \leq a \forall a \in [0 \dots 1]$
  - Prediction stronger, decrease the truth value
- Weak Modifier - $m(a) \geq a \forall a \in [0 \dots 1]$
  - Prediction weaker, increase the truth value

Modifiers Properties:
- $m(0) = 0$ and $m(1) = 1$
- m is a continuous function
- if m is strong, m^{-1} is weak and vice versa
- a composition of modifiers is a modifier. If both are of the same type, the result is of the same type

Handy Theorems

Chain Rule: $P(X_1, X_2) = P(X_1|X_2)P(X_2)$

Marginalization: $P(X_1) = \sum_{x \in X_2} P(X_1, X_2 = x)$

Bayes Rule: $P(X_1|X_2) = \frac{P(X_2|X_1)P(X_1)}{P(X_2)}$

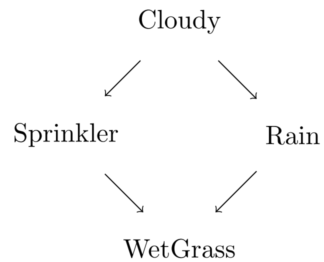
1. $P(X_1|X_2) = \frac{P(X_2|X_1) \cdot P(X_1)}{\sum_{x \in X_1} P(X_2|X_1=x) \cdot P(X_1=x)}$

2. $P(X_1|X_2, X_3) = \frac{P(X_2|X_1, X_3) \cdot P(X_1|X_3)}{P(X_2|X_3)}$

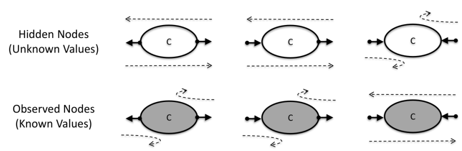
3. $P(X_1 = x|X_2) = \frac{P(X_2|X_1=x) \cdot P(X_1=x)}{P(X_2)}$

Bayesan Networks

Bayesan Network: a DAG where Nodes represent random variables and edges represent direct influence.



Conditional Independence: $P(X_1|X_2, X_3) = P(X_1|X_3)$
then X_1 and X_2 are conditionally independent given X_3



Two (sets of) nodes A and B are conditionally independent (d-separated) given C if and only if all the path from A to B are shielded by C.

Joint Distribution Factorization: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\mathrm{pa}(X_i))$

Explaining Away: describes two variable which become dependent because you observe a third one.

Markov Chains

(first Order) Markov Chain: $P(X_{t+1}|X_t, X_{t-1}, ..., X_1) = P(X_{t+1}|X_t)$

Stationary Event: $P(X_{t+1} = j|X_t = i) = p_{ij} \ \forall t$

Transition Matrix: $P = \begin{bmatrix} p_{11} & ... & p_{1n} \\ ... & ... & ... \\ p_{n1} & ... & p_{nn} \end{bmatrix} \quad \sum_{j=1}^{n} p_{ij} = 1$

Probability at step n: $P_{ij}(n) = P^n(i, j)$

Reachability: A state j is reachable from i if there exists a path from i to j

Communicability: States i and j communicate if each is reachable from the other

Absorbing State: A state i is absorbing if $p_{ii} = 1$

Transient State: A state i is transient if $\exists j$ reachable from i but not vice versa.

Recurrent State: A state i is recurrent if it is not transient.

Ergodic Markov Chain

A Markov Chain is ergodic if it is:

- Recurrent
- Aperiodic
- All states communicate

Steady State Distribution: $\pi = \lim_{n\to\infty} P^n$

Expected Transient Time: $m_{ij} = 1 + \sum_{k\neq j} p_{ik} m_{kj}$

Absorbing Markov Chain

$P = \begin{bmatrix} Q & R \\ 0 & 1 \end{bmatrix}$
Q transient states,
R absorbing states.

Expected time trans s i from trans j ij^{th} element of $(I - Q)^{-1}$

Expected time abs j from trans i ij^{th} row of $(I - Q)^{-1}R$

Hidden Markov Models

Hidden Markov Models: A HMM is (S, E, P, A, B):

- S is the set of hidden states
- E is the set of observations
- P is the distribution of the initial state
- A is the transition probability matrix
- B emission probability matrix

HMM: Viterbi Algorithm

F rom observations, compute the most likely sequence of hidden states:

$$\arg\max P(X_{1:t}|e_{1:t}) = \arg\max \frac{P(X_{1:t}, e_{1:t})}{P(e_{1:t})} = \arg\max P(X_{1:t}, e_{1:t}) \quad (1)$$

Lets apply bayesan factorization:

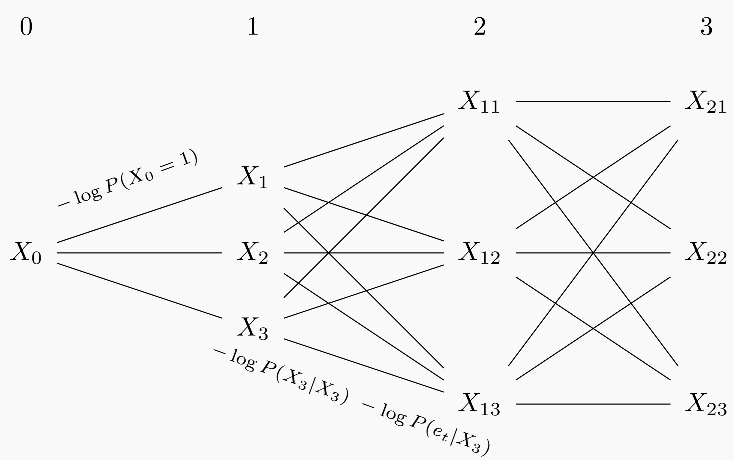
$$P(X_{1:t}, e_{1:t}) = P(X_0) \prod_{i=1}^{t} P(X_i|X_{i-1}) P(e_i|X_i) \quad (2)$$

The solution is the one that minimize:

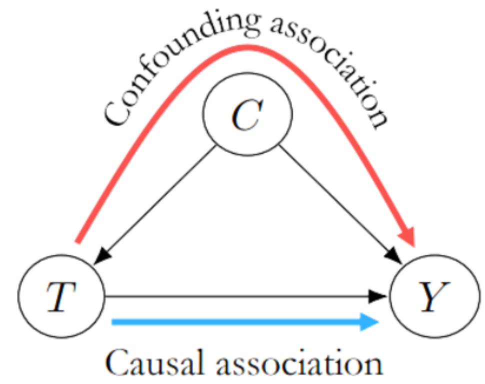
$$-\log P(X_{1:t}, e_{1:t}) = -\log P(X_0) + \sum_{i=1}^{t} (-\log P(X_i|X_{i-1}) - \log P(e_i|X_i)) \quad (3)$$

Graph Construction

Construct a graph with $1 + t \cdot n$ nodes, where n is the number of hidden states with one initial node and n nodes at time i where j^{th} node is $X_i = s_j$.
Find the shortest path from the initial node to the final node considering the weights of the edges.



Causal Inference



Individual treatment Effect (ITE) $ITE_i = Y_i(1) - Y_i(0)$

$$\begin{cases} ITE_i = 1 \text{ if there is causal association} \\ ITE_i = 0 \text{ otherwise} \end{cases}$$

Average treatment effect (ATE) $\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$
$\neq \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]$
i is the same individual!

Randomized Control Trial (RCT) randomizes subject into treatment groups to remove confounding association.
In this case,

$$ATE = \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]$$

Observational Study observes subjects in their natural environment without randomization.

Adjust for confounding Shield all the paths from T to Y In this case, even if no RCT:

$$\mathbb{E}[Y(t)|W=w] = \\ = \mathbb{E}[Y|do(T=t), W=w] \\ = \mathbb{E}[Y|T=t, W=w]$$

Backdoor adjustment example $\mathbb{E}[Y|do(T=t)] =$
$= \mathbb{E}_C \mathbb{E}[Y|t, C] = \sum_{c \in C} \mathbb{E}[Y|t, c] P(c)$