Markov Logic

Fabian M. Suchanek

Reminder Max Sat>8

Def: Weighted Rule

A weighted rule is a rule with an associated real-valued weight.

$$weather(rain, Paris) \Rightarrow weather(rain, Berlin)[3.14]$$

We consider only rules without variables for now.

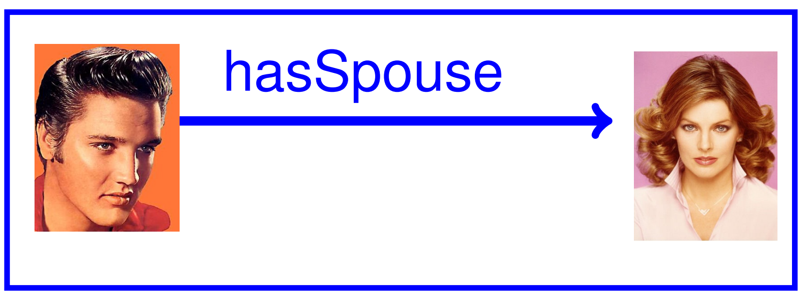
Def: Weight of a KB

Given a KB (a “possible world”) and a set of instantiated rules with weights, the weight of the KB is the sum of the weights of all true rules.

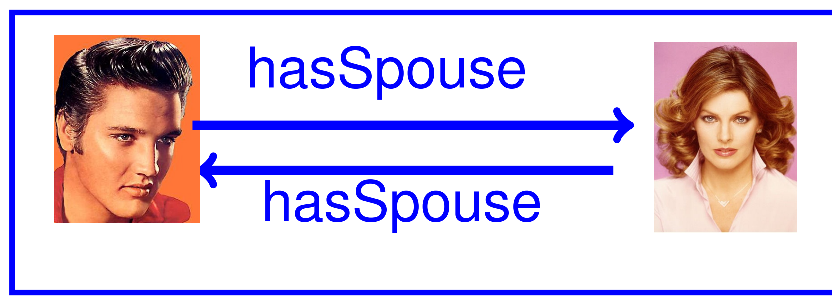
$$hasSpouse(Elvis, Priscilla) \Rightarrow hasSpouse(Priscilla, Elvis)[3]$$

$$hasSpouse(cat, dog) \Rightarrow hasSpouse(dog, cat)[2]$$

KB1



KB2



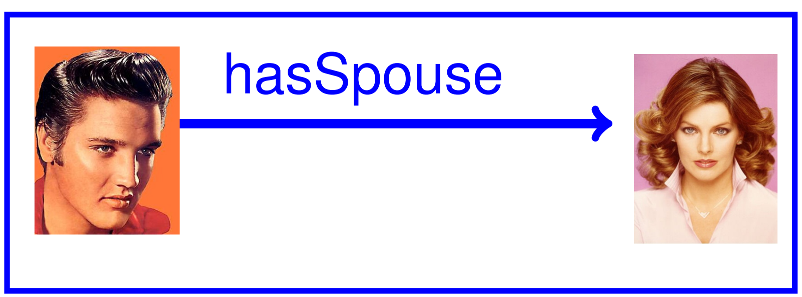
Def: Weight of a KB

Given a KB (a “possible world”) and a set of instantiated rules with weights, the weight of the KB is the sum of the weights of all true rules.

$$hasSpouse(Elvis, Priscilla) \Rightarrow hasSpouse(Priscilla, Elvis)[3]$$

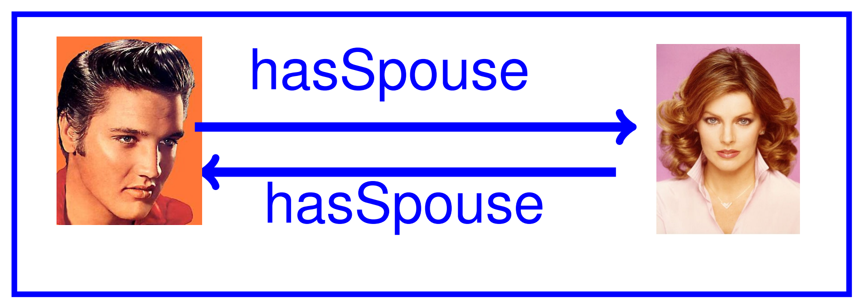
$$hasSpouse(cat, dog) \Rightarrow hasSpouse(dog, cat)[2]$$

KB1



Weight: 2

KB2



Weight: 5

Def: Weighted MAX SAT

Given a set of instantiated rules with weights, weighted MAX SAT is the problem of finding the KB with the highest weight.

(Since SAT is NP-complete, so is MAX SAT and Weighted MAX SAT. There may be multiple such worlds. We are interested in minimal worlds.)

$is(Ron, immature)$[10]

$is(Ron, immature) \wedge type(H., sorceress) \Rightarrow likes(H., Ron)$[3]

$type(Hermione, sorceress)$[4]

Def: Weighted MAX SAT

Given a set of instantiated rules with weights, weighted MAX SAT is the problem of finding the KB with the highest weight.

(Since SAT is NP-complete, so is MAX SAT and Weighted MAX SAT. There may be multiple such worlds. We are interested in minimal worlds.)

$is(Ron, immature)$[10]

$is(Ron, immature) \wedge type(H., sorceress) \Rightarrow likes(H., Ron)$[3]

$type(Hermione, sorceress)$[4]

Best world:

$is(Ron, immature)$

$type(Hermione, sorceress)$

$likes(Hermione, Ron)$

weight: 17

task>8

Task: Weighted MAX SAT

Find the KB with the highest weight:

$is(Hermione, smart)$[1]

$is(Herm., smart) \wedge is(Harry, smart) \Rightarrow likes(Herm., Harry)$[3]

$likes(Hermione, Ron) \Rightarrow \neg likes(Hermione, Harry)$[100]

$is(Harry, smart)$[10]

$likes(Hermione, Ron)$[20]

Hint: Start by satisfying disjunctions with one literal and high weight.

A probabilistic view

Let us see every ground literal as a random variable

with values T and F:

$$spouse(Harry, Ron) \longrightarrow X_3$$

Then we can draw up all possible worlds with their probabilities:

	X_1	X_2	X_3	
w_1	F	F	F	$P(w_1) = 0.1$
w_2	F	F	T	$P(w_2) = 0.3$
w_3	F	T	F	$P(w_3) = 0.01$

possible worlds with probabilities

But what are these probabilities?

Markov Random Fields

A Markov Random Field (MRF) is a set of random variables that satisfies:

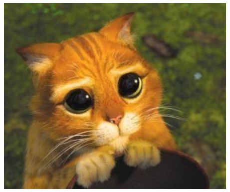
$$P(X_i|X_1, .., X_{i-1}, X_{i+1}, ..., X_n) = P(X_i|N(X_i))$$

where $N(X_i)$ is the set of random variables that are “neighbors” of X_i .

That is: The probability of X_i taking a certain value given the values of the other variables depends only on the values of the variables in the neighborhood of X_i .

Example: Markov Random Field

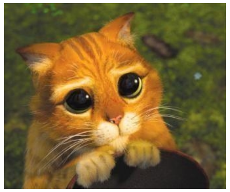
world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...



Example: Markov Random Field

world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...

X1, X2, and X3 depend on each other
(shouting is always ferociously, purring is never ferociously,
shouting is more likely to be by Shrek)



X4 and X5 depend on each other
(either both are the empty string, or X4=with and X5=pleasure)

(X1,X2,X3) and (X4,X5) are independent

Example: Markov Random Field

world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...

Neighbor sets:

$$N(X_1) = \{X_2, X_3\}$$

$$N(X_2) = \{X_1, X_3\}$$

$$N(X_3) = \{X_1, X_2\}$$

$$N(X_4) = \{X_5\}$$

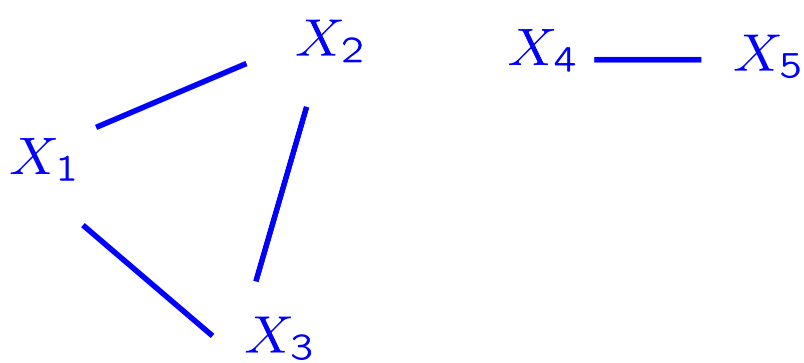
$$N(X_5) = \{X_4\}$$

Def: MRF graph

world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...

We depict the neighborhood as an undirected graph,
where the nodes are variables, and $X_i - X_j$ is an edge if $X_i \in N(X_j)$.

$$N(X_1) = \{X_2, X_3\}$$
$$N(X_2) = \{X_1, X_3\}$$
$$N(X_3) = \{X_1, X_2\}$$
$$N(X_4) = \{X_5\}$$
$$N(X_5) = \{X_4\}$$

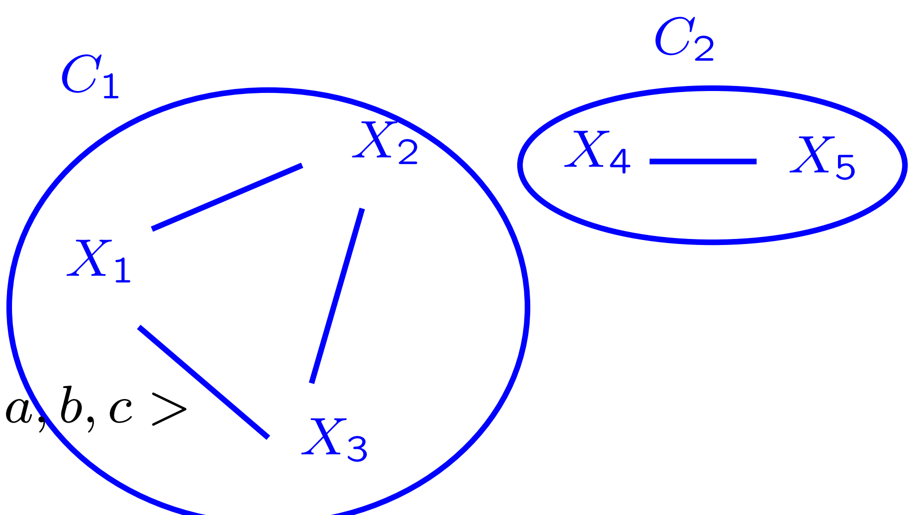


Maximal cliques

world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...

We consider the maximal cliques in the neighborhood graph.

$$C_1 = \{X_1, X_2, X_3\}$$
$$C_2 = \{X_4, X_5\}$$



We write

$$C_1(< a, b, c, d, e, f >) =< a, b, c >$$

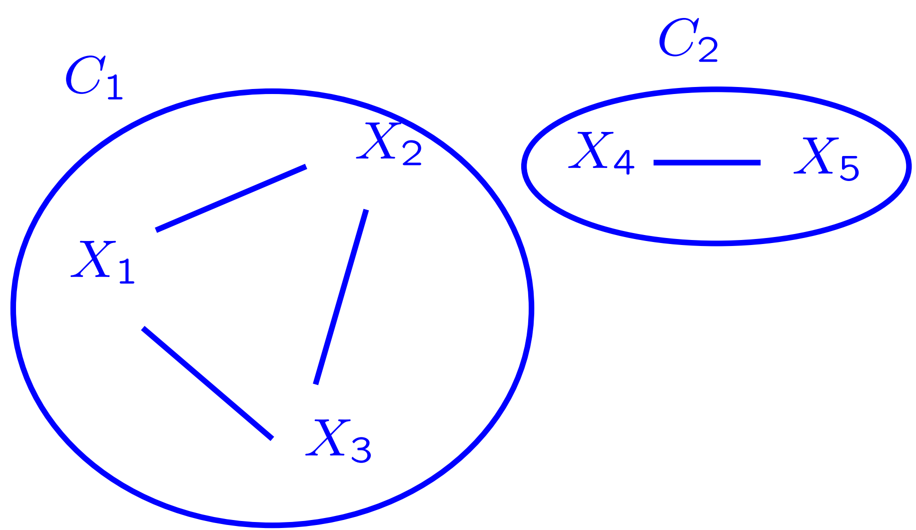
Hammersley-Clifford-Theorem

world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...

If all probabilities in an MRF are > 0, then $\exists \phi_i$ such that $P(\vec{X} = \vec{x}) = \prod_i \phi_i(C_i(\vec{x}))$.

$C_1 = \{X_1, X_2, X_3\}$

$C_2 = \{X_4, X_5\}$



>example

Hammersley-Clifford-Theorem

world	X1	X2	X3	X4	X5	probability
w1:	Shrek	shouts	ferociously	with	pleasure	P(w1)=0.1
w2:	Shrek	shouts	ferociously	ϵ	ϵ	P(w2)=0.1
w3:	Shrek	purrs	ϵ	with	pleasure	P(w3)=0.02
w4:	Shrek	purrs	ϵ	ϵ	ϵ	P(w4)=0.02
w5:	Puss	shouts	ferociously	with	pleasure	P(w5)=0.01
...	...					...

If all probabilities in an MRF are > 0, then $\exists \phi_i$ such that $P(\vec{X} = \vec{x}) = \prod_i \phi_i(C_i(\vec{x}))$.

$$\phi_1(x_1, x_2, x_3) = (x_1 = Shrek \wedge x_2 = shouts \wedge x_3 = ferociously) \vee (x_1 = Puss \wedge x_2 = purrs \wedge x_3 = \epsilon) ? 0.1 : 0.01$$

$$\phi_2(x_4, x_5) = (x_4 = with \wedge x_5 = pleasure) \vee (x_4 = x_5 = \epsilon) ? 1.0 : 0.02$$

(example might not sum to 1)

Back to the logical worlds

world	likes(H., Ron)	spouse(Harry,Ron)	spouse(Ron,Harry)	probability
w1:	T	F	F	P(w1)=0.1
w2:	T	F	F	P(w2)=0.1
...				...

Rule 42: $\underbrace{spouse(Harry, Ron)}_{X_2} \Rightarrow \underbrace{spouse(Ron, Harry)}_{X_3}$[5]

Let's define $\phi_i(x_k, ..., x_l) = exp($rule i satisfied with $x_k, ..., x_l$? w_i : 0)

$$\phi_{42}(x_2, x_3) = exp(x_2 = F \vee x_3 = T ? 5 : 0)$$

$\phi_{42}(F, T) = e^5$ $\phi_{42}(T, F) = 1$

>example

Example: Markov Logic Network

Original Weighted MAX SAT problem:

$is(Ron, immature)$[10]

$is(Ron, immature) \wedge type(H., sorceress) \Rightarrow likes(H., Ron)$[3]

$type(Hermione, sorceress)$[4]

Assignment to variables:

$is(Ron, immature) \rightarrow X_1$

$likes(Hermione, Ron) \rightarrow X_2$

$type(Hermione, sorceress) \rightarrow X_3$

Factors:

$$\phi_1(x_1) = exp(x_1 = T?10 : 0) = \begin{cases} e^{10} \text{ if } x_1 = T \\ 0 \text{ else} \end{cases}$$

$$\phi_2(x_1, x_2, x_3) = exp(x_1 \wedge x_2 \wedge \neg x_3?0 : 3)$$

$$\phi_3(x_3) = exp(x_3 = T?4 : 0)$$

$$P(x_1, x_2, x_3) = \phi_1(x_1) \times \phi_2(x_1, x_2, x_3) \times \phi_3(x_3)$$

Example: Markov Logic Network

Original Weighted MAX SAT problem:

$is(Ron, immature)$[10]

$is(Ron, immature) \wedge type(H., sorceress) \Rightarrow likes(H., Ron)$[3]

$type(Hermione, sorceress)$[4]

Assignment to variables:

$is(Ron, immature) \rightarrow X_1$

$likes(Hermione, Ron) \rightarrow X_2$

$type(Hermione, sorceress) \rightarrow X_3$

Factors:

$\phi_1(x_1) = exp(x_1 = T?10 : 0) = \begin{cases} e^{10} & \text{if } x_1 = T \\ 0 & \text{else} \end{cases}$

$\phi_2(x_1, x_2, x_3) = exp(x_1 \wedge x_2 \wedge \neg x_3?0 : 3)$

$\phi_3(x_3) = exp(x_3 = T?4 : 0)$

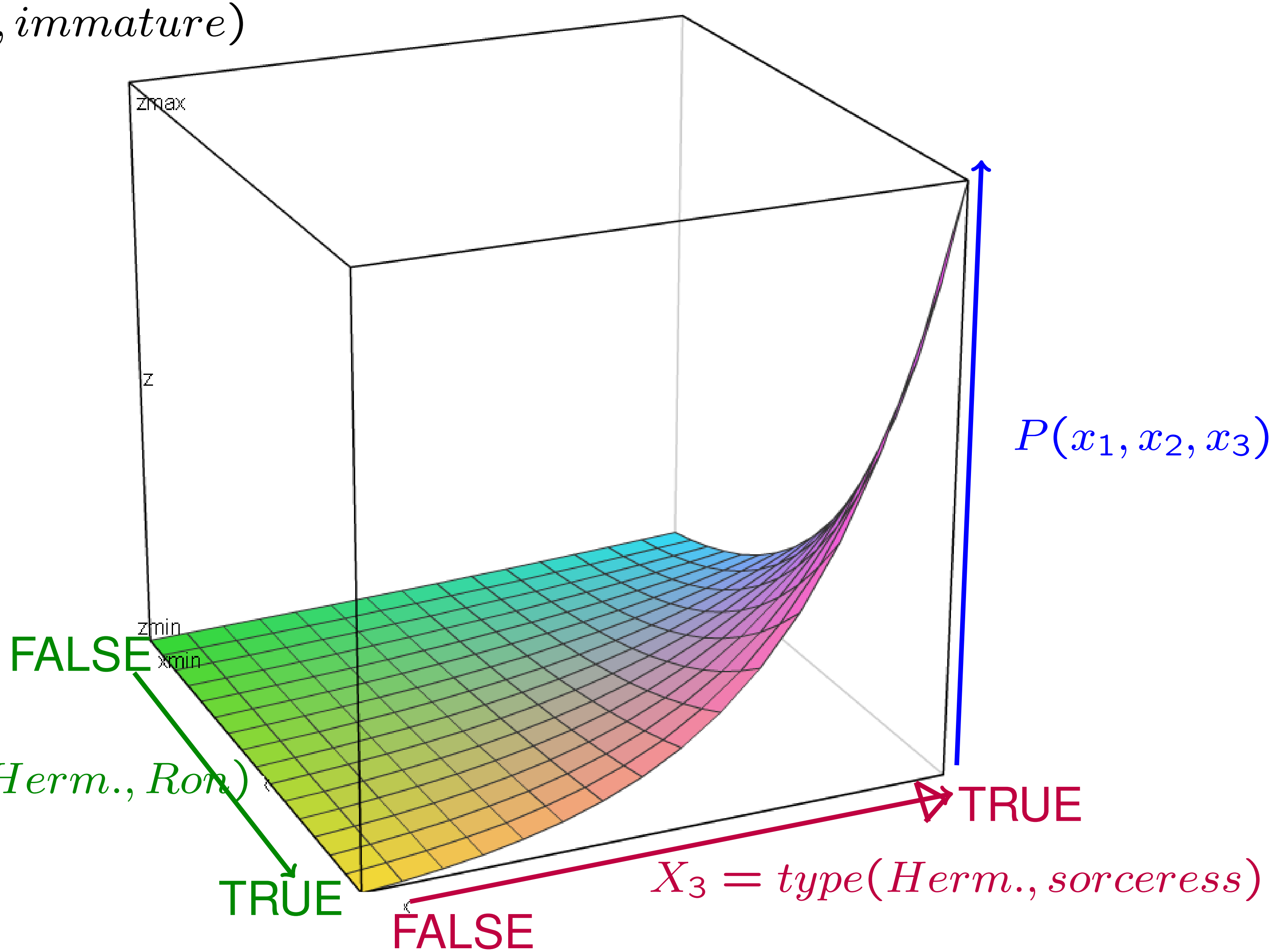
A world becomes $exp(w_i)$ more likely if formula i is satisfied.

$$P(x_1, x_2, x_3) = \phi_1(x_1) \times \phi_2(x_1, x_2, x_3) \times \phi_3(x_3)$$

Example: Markov Logic Network

$X_1 = is(Ron, immature)$

$=$TRUE



LivePhysics Plotter

Normalization still missing

Factors as continuums

>Normalization

$$P(x_1, x_2, x_3) = \phi_1(x_1) \times \phi_2(x_1, x_2, x_3) \times \phi_3(x_3)$$

Normalization

To obtain a value in [0,1], we normalize by Z

$$P(\vec{X} = \vec{x}) = \frac{1}{Z} \prod_i \phi_i(C_i(\vec{x}))$$

where Z is simply the sum over the products for all possible sequences of values $\vec{x'}$:

$$Z = \sum_{\vec{x'}} \prod_i \phi_i(C_i(\vec{x'}))$$

Yes, run over all possible sequences $\vec{x'} = < x'_1, ..., x'_n >$ and sum up the products!

This ensures $P(\vec{X} = \vec{x}) \in [0, 1]$.

Def: Markov Logic Network

A Markov Logic Network (MLN) for a set of weighted rules (or weighted propositional formulae) is a Markov Random Network, where the variables are the positive literals and the factors are

$$\phi_i(x_k, ..., x_l) = exp(\text{rule } i \text{ satisfied with } x_k, ..., x_l \text{ ? } w_i : 0)$$

literals involved in rule i

weight of rule i

>end

Quantified formulas

A Markov Logic Network (MLN) for a set of weighted rules (or weighted propositional formulae) is a Markov Random Network, where the variables are the positive literals and the factors are

$$\phi_i(x_k, ..., x_l) = exp(\text{rule } i \text{ satisfied with } x_k, ..., x_l \text{ ? } w_i : 0)$$

Formulas with universally quantified variables such as

$spouse(X, Y) \Rightarrow spouse(Y, X)$[5]

have to be replaced by all of their ground instances:

$spouse(Ron, Hermione) \Rightarrow spouse(Hermione, Ron)$[5]

$spouse(Ron, Harry) \Rightarrow spouse(Harry, Ron)$[5]

...

>MAX SAT

Markov Logic and Weighted MAX SAT

$$P(\vec{X} = \vec{x}) = \frac{1}{Z} \prod_i \phi_i(C_i(\vec{x}))$$

Let us find the most likely world:

$$argmax_{\vec{x}} \frac{1}{Z} \prod_i \phi_i(C_i(\vec{x}))$$

$$argmax_{\vec{x}} \prod_i \phi_i(C_i(\vec{x}))$$

$$argmax_{\vec{x}} \prod_i exp(\text{rule } i \text{ satisfied ? } w_i : 0)$$

$$argmax_{\vec{x}} exp(\sum_i \text{rule } i \text{ satisfied ? } w_i : 0)$$

$$argmax_{\vec{x}} \sum_{i, ruleisat} w_i$$

This is Weighted MAX SAT

What MLNs can do

MLNs can

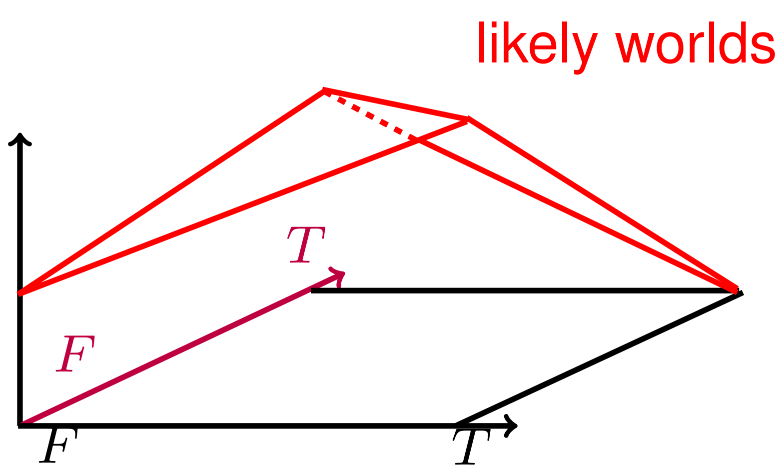
- compute the most likely world

$$argmax_{\vec{x}} P(\vec{X} = \vec{x})$$

- compute the marginals

$$P(X_3 = T)$$

- model the distribution of probabilities



Appendix: MRFs and Markov Chains

Markov Chains:

$$P(X_i|X_1, ..., X_{i-1}) = P(X_i|X_{i-1})$$

The probability of $X = v$ depends only on the value of the predecessor of X.

Markov Random Fields:

$$P(X_i|X_1, .., X_{i-1}, X_{i+1}, ..., X_n) = P(X_i|N(X_i))$$

The probability of $X = v$ depends only on the value of the neighbors of X, but neighborhood is symmetric, and the probability is also conditioned on the “future” X_i.

References

Markov Logic Networks