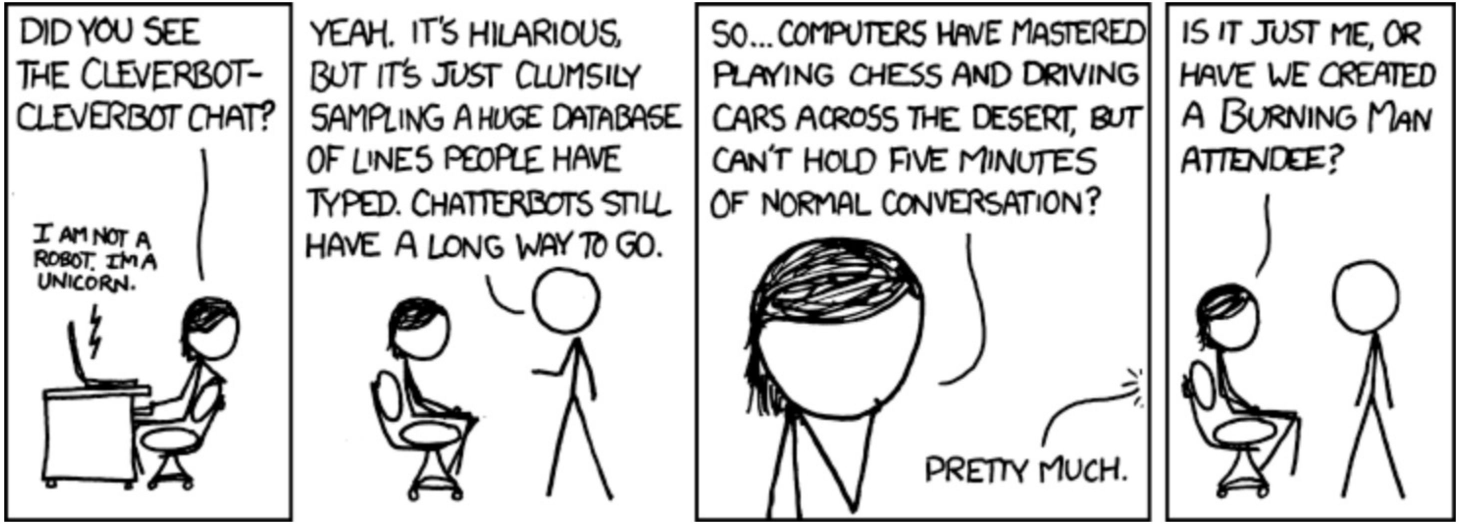
Credal Models for Uncertainty and Logic Treatment

CASSIO DE CAMPOS @ PROBABILISTIC CIRCUITS AND LOGIC
SIMONS INSTITUTE - UC BERKELEY – 16 OCTOBER 2023

EINDHOVEN UNIVERSITY OF TECHNOLOGY

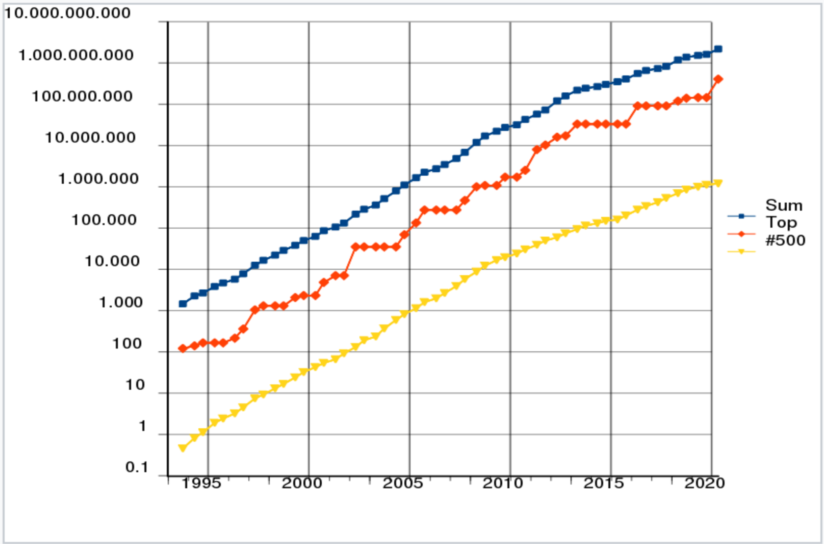
12 Years ago…



Title text: And they both react poorly to showers.

Source: xkcd

AI and Better Computing Power



Rapid growth of supercomputer performance, based on data from top500.org site. The logarithmic *y*-axis shows performance in GFLOPS.

- Combined performance of 500 largest supercomputers
- Fastest supercomputer
- Supercomputer in 500th place

Source: https://en.wikipedia.org/wiki/TOP500

Example - Deep Fakes

Source: NPO

Source: NPO, modified

Examples from 21 September 2023

https://www.bbc.com/news/technology-66866577

Game of Thrones author sues ChatGPT owner OpenAI

4 hours ago

The hit TV show Game of Thrones was based on George RR Martin's novels

By Tom Gerken & Liv McMahon
Technology reporters

US authors George RR Martin and John Grisham are suing ChatGPT-owner OpenAI over claims their copyright was infringed to train the system.

Martin is known for his fantasy series A Song of Ice and Fire, which was adapted into HBO show Game of Thrones.

Source: bbc.com

https://www.bbc.com/news/world-us-canada-66873982

Google accused of directing motorist to drive off collapsed bridge

13 hours ago

By Max Matza
BBC News

The family of a US man who drowned after driving off a collapsed bridge are claiming that he died because Google failed to update its maps.

Philip Paxson's family are suing the company over his death, alleging that Google negligently failed to show the bridge had fallen nine years earlier.

Mr Paxson died in September 2022 after attempting to drive over the damaged bridge in Hickory, North Carolina.

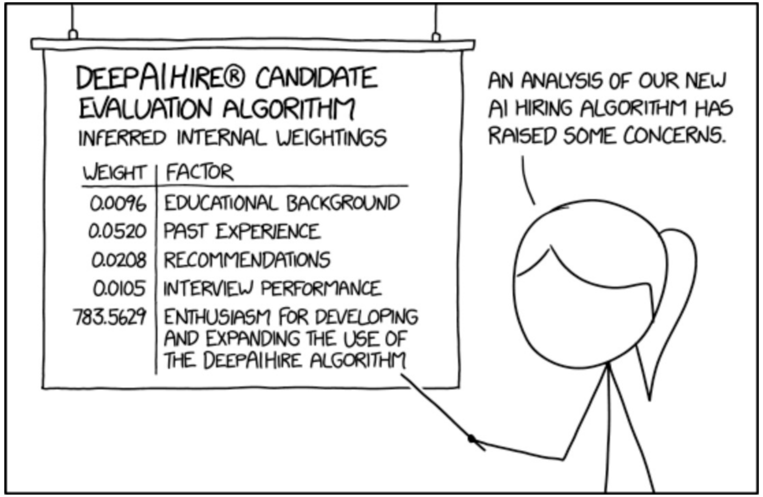
Source: bbc.com

Some Consequences

- AI research has fast growing impact in society
  - The use of AI might need stronger regulation

- We may need ways to certify and control AI systems, specially if we do not fully "understand" them

- More investment in "better" AI, specially from governments



Source: xkcd.com

TU/e

BBC News – 14 September 2023

Some companies (Google, Meta, Microsoft, SpaceX/X/Tesla) met to discuss AI.

- ``*Congress should engage with AI to support innovation and safeguards*'', Mark Zuckerberg CEO Meta.
- ``*I think if this technology goes wrong, it can go quite wrong... we want to be vocal about that*'', Samuel Altman CEO OpenAI continues ``*We want to work with the government to prevent that from happening*''.
- ``*I think there should be a regulatory body established for overseeing AI to make sure that it does not present a danger to the public*'', Elon Musk CEO SpaceX/Tesla. And he continues ``*better that the standard is set by American companies that can work with our government to shape these models on important issues*''.

Doctors' Example

- Patient Mr. Sick has either auto-immune (A) disease or an infection (B). Without treatment he will likely die very soon. Assume these diseases are equally likely a priori.

- After studying the case in private, Dr. Imprecise tells she does not know whether it is A or B. Dr. Precise tells it is A.

Which doctor would you prefer if you were Mr. Sick?

``It's not (only) about the result, it's about how we reached it.''

The hypothetical underlying process for the diagnosis:

- Dr. Imprecise concluded the answer is in the set A,B after studying the data. She was not able to pinpoint a unique option.
- Dr. Precise told it is A after flipping a fair coin and using the outcome to choose.

After knowing the process, which doctor would you prefer if you were Mr. Sick?

Example: knowing when one does not know

Suppose there are 10 options (e.g. the digits) and image data of them. We must discover the digit in the image. What is best?

- An approach which always predicts a digit for any given image and has 90% accuracy.
- An approach which always predicts a digit for any given image with accuracy 99.9%, but is allowed to say "I do not know" in a certain amount of cases.
- An approach which some times predicts multiple digits (e.g. could not decide between a "6" and a "8") and has 99.99% accuracy (meaning the correct is within the set of predicted options).

AI must consider multiple types of uncertainty

BBC is paying us to discover the popularity of Eastenders (long running soap opera). We decide to call 10 "random" valid phone numbers.

- 4 people answered the phone and said they like it
- 1 people answered the phone and said they do not like it
- 5 people did not answer the phone

Typical approaches in AI/ML assume missing data at random, which would lead to 80% of people like Eastenders. Is that a meaningful result? Are we ok with reporting this percentage back to BBC?

AI must consider multiple types of uncertainty

BBC is paying us to discover the popularity of Eastenders (long running soap opera). We decide to call 10 “random” valid phone numbers.

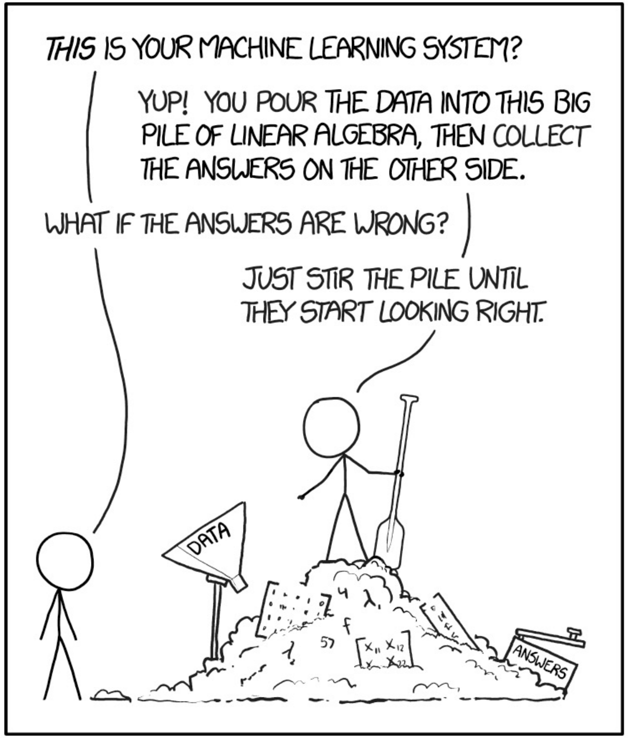
- 4 people answered the phone and said they like it
- 1 people answered the phone and said they do not like it
- 5 people did not answer the phone

Typical approaches in AI/ML assume missing data at random, which would lead to 80% of people like Eastenders. Is that a meaningful result? Are we ok with reporting this percentage back to BBC?

- Eastenders is more popular among older people
- Young people much more often do not answer the phone

Better AI?

- Desirable properties
  - Interpretability
  - Robustness
  - Explainability
  - Privacy
  - Fairness
- Usually bring benefits but do not come for free
  - More computational resources
  - More intricate solutions



Source: xkcd.com

Are we willing to pay the price for <u>trustworthy</u> AI?

TU/e

Three different levels of knowledge

- Football Match: Italy vs. Sweden
- Italy result? Win, draw or loss?

DETERMINISM

Buffon (Italy goalkeeper) is just unbeatable ... while Sweden always gets at least a goal

Italy (certainly) wins

$$\begin{matrix} P(win) \\ P(draw) \\ P(loss) \end{matrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

UNCERTAINTY

Win is two times more probable than draw, this being three times more probable than loss

$$\begin{matrix} P(win) \\ P(draw) \\ P(loss) \end{matrix} = \begin{bmatrix} .6 \\ .3 \\ .1 \end{bmatrix}$$

IMPRECISION

Win is more probable than draw, and this is more probable than loss

$$P(win) > P(draw)$$
$$P(draw) > P(loss)$$

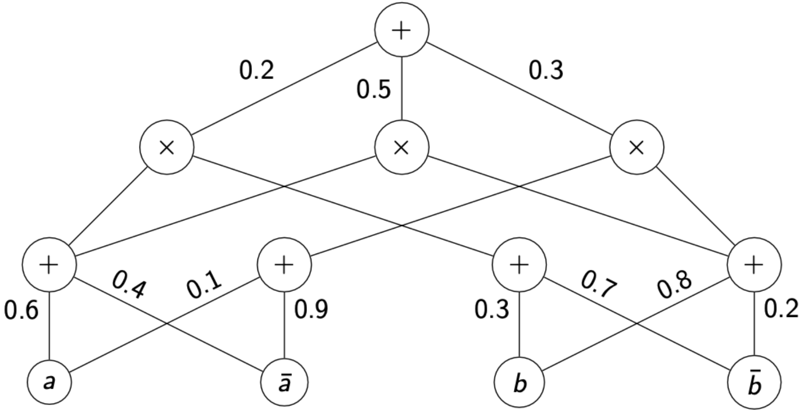
$$\begin{matrix} P(win) \\ P(draw) \\ P(loss) \end{matrix} = \begin{bmatrix} \frac{\alpha}{3} + \beta + \frac{\gamma}{2} \\ \frac{\alpha}{3} + \frac{\gamma}{2} \\ \frac{\alpha}{3} \end{bmatrix}$$

$\forall \alpha, \beta, \gamma$ such that
$\alpha > 0, \beta > 0, \gamma > 0,$
$\alpha + \beta + \gamma = 1$

TU/e

Deep Models

- Sum-Product Networks: sacrifice "interpretability" for the sake of computational efficiency; represent computations not interactions.
- Complex mixture distributions represented graphically as an arithmetic circuit.



Sum-Product Network

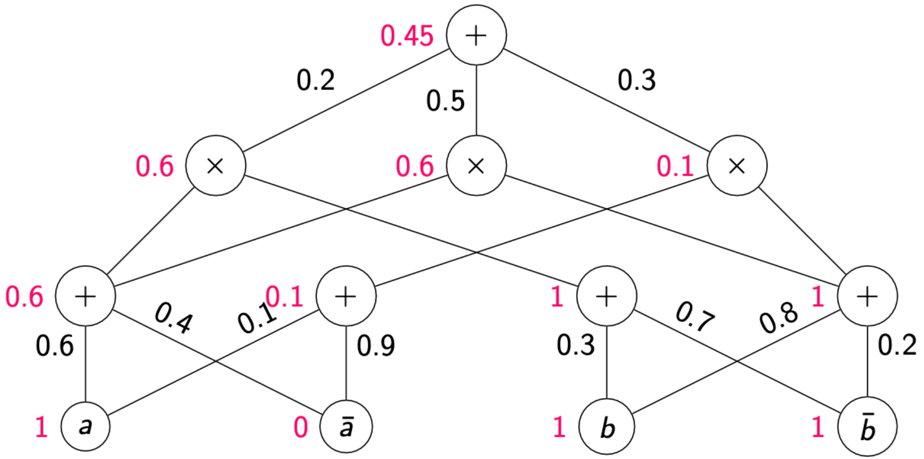
Distribution $P(X_1, \ldots, X_n)$ built by

- an indicator function over a single variable
  - $I(X = 0)$, $I(Y = 1)$ (also written $\neg x, y$),
- a weighted sum of SPNs with same domain and nonnegative weights
  - $P_3(X, Y) = 0.6 \cdot P_1(X, Y) + 0.4 \cdot P_2(X, Y)$,
- a product of SPNs with disjoint domains
  - $P_3(X, Y, Z, W) = P_1(X, Y) \cdot P_2(Z, W)$.

Evaluation (Inference)

- Propagate values bottom-up:

$P(A = a) =$

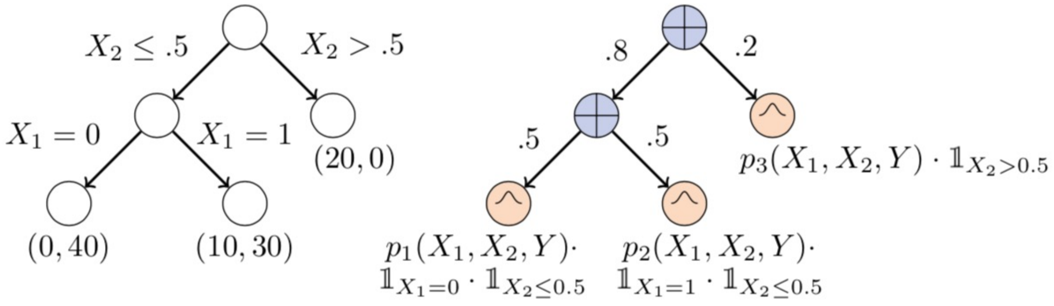


Note: takes linear time in the size of circuit!

Generative Decision Trees and Random Forests

Representation of Decision Trees as Probabilistic Circuit

- Convert each internal node to a sum node
    - Weights are given by the mass of each children
- Convert each leaf into a distribution node
    - Fit a density over the instances in each leaf

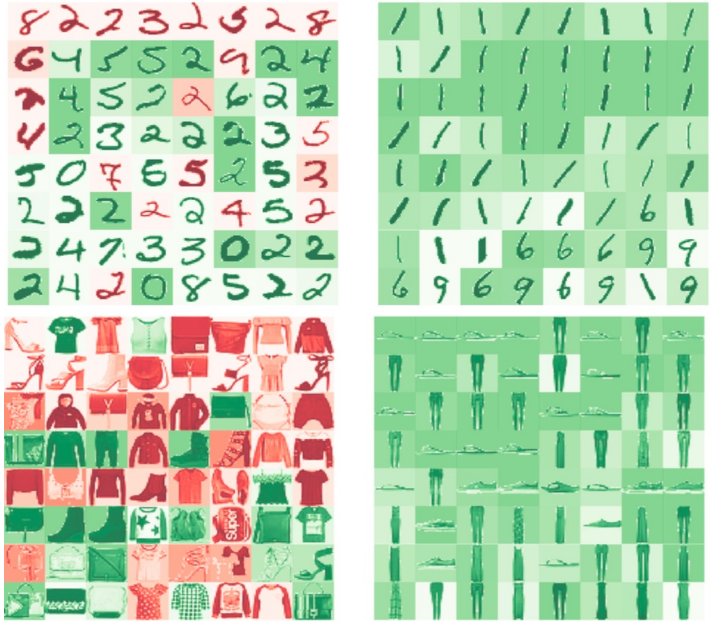


Generative Random Forests

- State of the art for tabular data
  - Probabilistic model with tractable marginals/conditionals

- Same quality of results of random forests, while better at:
  - Missing data treatment
  - Outlier detection
  - Smoothing decision boundaries
  - Robustness/adversarial training
  - Sensitivity analysis

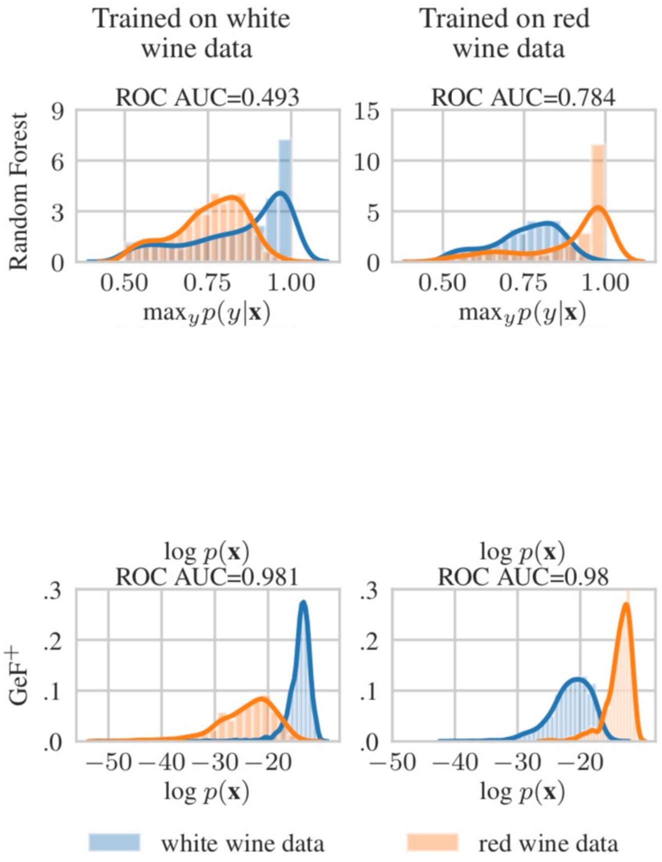
Using p(x) to know when we do not know



Samples from (Fashion-)Mnist datasets with lowest (left) and highest (right) $p(\boldsymbol{x})$ in the test set.

Using p(x) to know when we do not know

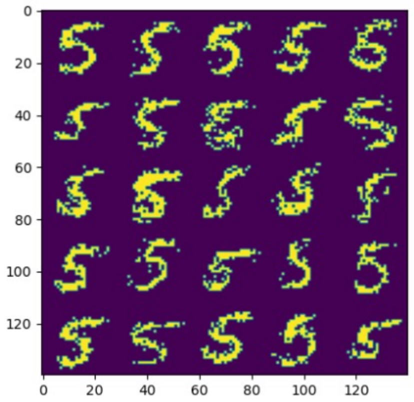
- Better than p(y|x) for outlier detection
- Can also be better for knowing when we do not know
  - E.g. Naïve Bayes classifier tends to have extreme p(y|x)



Limitations: p(x) to know when we do not know

- Imagine data has badly written 5's and 6's
  - It has many of them
  - They lie close to each other in the "space" of number images for the model in use

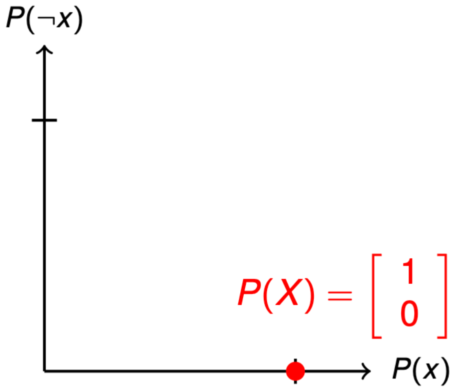
- In this case, p(x) of a new sample of interest might be very high, while there may be great uncertainty about being 5 or 6



Credal Sets over Boolean Variables

- Boolean X, values in $\mathcal{X} = \{x, \neg x\}$
- Determinism $\equiv$ degenerate mass f

  E.g., $X = x \iff P(X) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

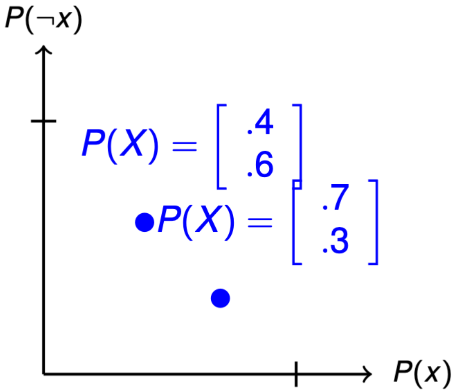


Credal Sets over Boolean Variables

- Boolean X, values in $\mathcal{X} = \{x, \neg x\}$
- Determinism $\equiv$ degenerate mass f

  E.g., $X = x \iff P(X) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- Uncertainty $\equiv$ prob mass function

  $P(X) = \begin{bmatrix} p \\ 1-p \end{bmatrix}$ with $p \in [0, 1]$



Credal Sets over Boolean Variables

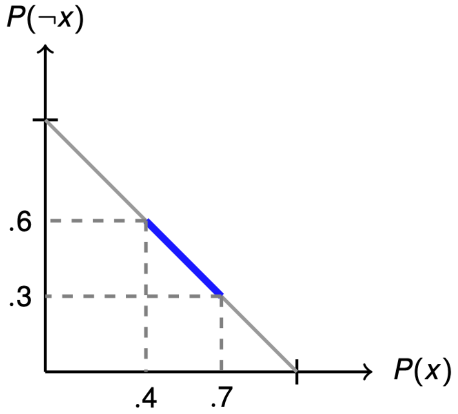
- Boolean X, values in $\mathcal{X} = \{x, \neg x\}$
- Determinism $\equiv$ degenerate mass f

  E.g., $X = x \iff P(X) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- Uncertainty $\equiv$ prob mass function

  $P(X) = \begin{bmatrix} p \\ 1-p \end{bmatrix}$ with $p \in [0, 1]$
- Imprecision credal set on the *probability simplex*

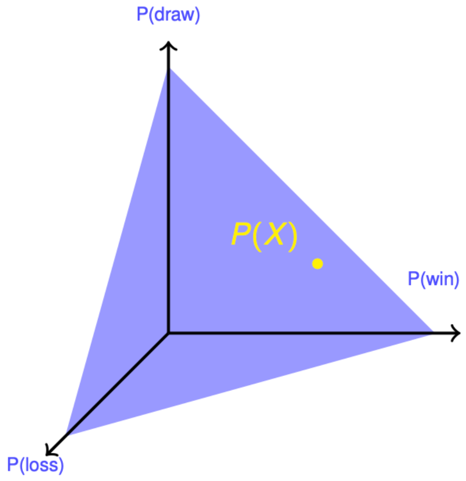
  $$K(X) \equiv \left\{ P(X) = \begin{bmatrix} p \\ 1-p \end{bmatrix} \middle| .4 \leq p \leq .7 \right\}$$
- A CS over a Boolean variable cannot have more than two vertices!

  $$\text{ext}[K(X)] = \left\{ \begin{bmatrix} .7 \\ .3 \end{bmatrix}, \begin{bmatrix} .4 \\ .6 \end{bmatrix} \right\}$$



Geometric Representation of CSs (ternary variables)

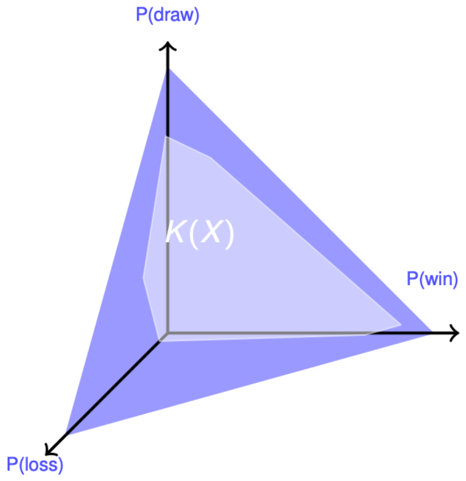
- Ternary X (e.g., $\Omega = \{\text{win,draw,loss}\}$)
- $P(X) \equiv$ point in the space (simplex)
- No bounds to $|\text{ext}[K(X)]|$
- Modelling ignorance
  - Uniform models indifference
  - Vacuous credal set
- Expert qualitative knowledge
  - Comparative judgements: win is more probable than draw, which more probable than loss
  - Qualitative judgements: adjective $\equiv$ IP statements



$$P(X) = \begin{bmatrix} .6 \\ .3 \\ .1 \end{bmatrix}$$

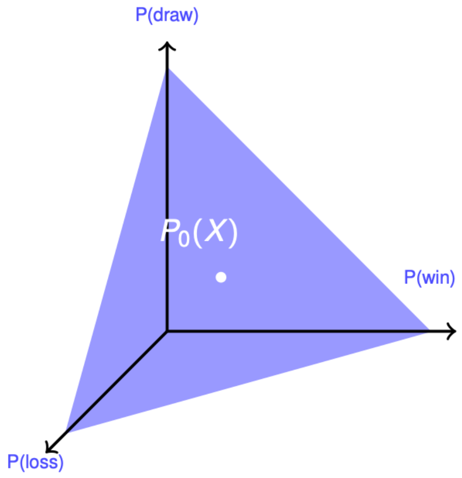
Geometric Representation of CSs (ternary variables)

- Ternary X (e.g., $\Omega = \{$win,draw,loss$\}$)
- $P(X) \equiv$ point in the space (simplex)
- No bounds to $|\mathrm{ext}[K(X)]|$
- Modelling ignorance
  - Uniform models indifference
  - Vacuous credal set
- Expert qualitative knowledge
  - Comparative judgements: win is more probable than draw, which more probable than loss
  - Qualitative judgements: adjective $\equiv$ IP statements



Geometric Representation of CSs (ternary variables)

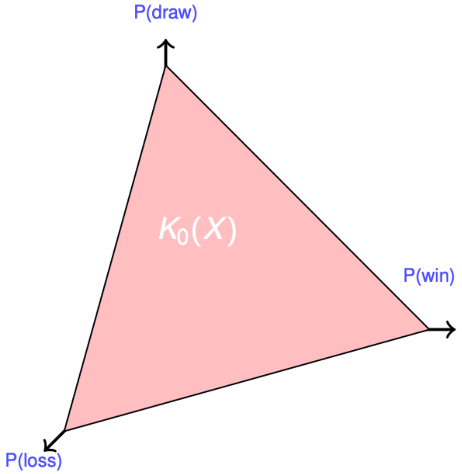
- Ternary X (e.g., $\Omega = \{$win,draw,loss$\}$)
- $P(X) \equiv$ point in the space (simplex)
- No bounds to $|\mathrm{ext}[K(X)]|$
- Modelling ignorance
  - Uniform models indifference
  - Vacuous credal set
- Expert qualitative knowledge
  - Comparative judgements: win is more probable than draw, which more probable than loss
  - Qualitative judgements: adjective $\equiv$ IP statements



$$P_0(x) = \frac{1}{|\Omega_X|}$$

Geometric Representation of CSs (ternary variables)

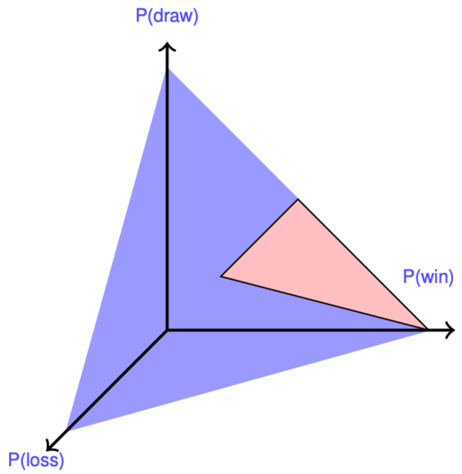
- Ternary X (e.g., $\Omega = \{\text{win,draw,loss}\}$)
- $P(X) \equiv$ point in the space (simplex)
- No bounds to $|\text{ext}[K(X)]|$
- Modelling ignorance
  - Uniform models indifference
  - Vacuous credal set
- Expert qualitative knowledge
  - Comparative judgements: win is more probable than draw, which more probable than loss
  - Qualitative judgements: adjective $\equiv$ IP statements



$$K_0(X) = \left\{ P(X) \;\middle|\; \begin{array}{l} \sum_x P(x) = 1, \\ P(x) \geq 0 \end{array} \right\}$$

Geometric Representation of CSs (ternary variables)

- Ternary X (e.g., $\Omega = \{\text{win,draw,loss}\}$)
- $P(X) \equiv$ point in the space (simplex)
- No bounds to $|\text{ext}[K(X)]|$
- Modelling ignorance
  - Uniform models indifference
  - Vacuous credal set
- Expert qualitative knowledge
  - Comparative judgements: win is more probable than draw, which more probable than loss
  - Qualitative judgements: adjective $\equiv$ IP statements



Geometric Representation of CSs (ternary variables)

- Ternary X (e.g., $\Omega = \{\text{win,draw,loss}\}$)
- $P(X) \equiv$ point in the space (simplex)
- No bounds to $|\text{ext}[K(X)]|$
- Modelling ignorance
  - Uniform models indifference
  - Vacuous credal set
- Expert qualitative knowledge
  - Comparative judgements: win is more probable than draw, which more probable than loss
  - Qualitative judgements: adjective $\equiv$ IP statements

From natural language to linear constraints on probabilities

(Walley, 1991)

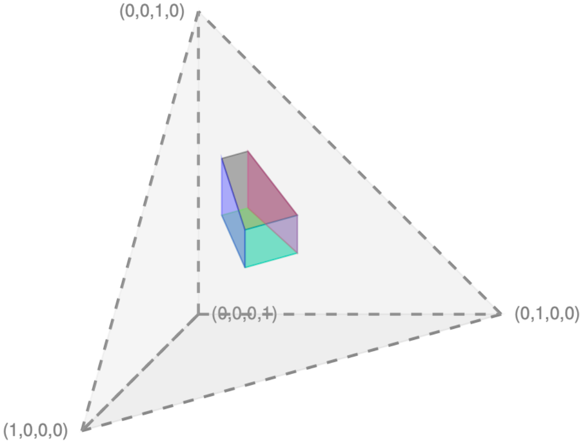
extremely probable $P(x) \geq 0.98$
very high probability $P(x) \geq 0.9$
highly probable $P(x) \geq 0.85$
very probable $P(x) \geq 0.75$
has a very good chance $P(x) \geq 0.65$
quite probable $P(x) \geq 0.6$
probable $P(x) \geq 0.5$
has a good chance $0.4 \leq P(x) \leq 0.85$
is improbable (unlikely) $P(x) \leq 0.5$
is somewhat unlikely $P(x) \leq 0.4$
is very unlikely $P(x) \leq 0.25$
has little chance $P(x) \leq 0.2$
is highly improbable $P(x) \leq 0.15$
is has very low probability $P(x) \leq 0.1$
is extremely unlikely $P(x) \leq 0.02$

Multivariate credal sets

- Two Boolean variables: **S**moker, Lung **C**ancer
- 8 “Bayesian” physicians, each assessing $P_j(S, C)$

$K(S,C) = \mathrm{CH}\{P_j(S,C)\}_{j=1}^{8}$

j	$P_j(s,c)$	$P_j(s,\bar{c})$	$P_j(\bar{s},c)$	$P_j(\bar{s},\bar{c})$
1	1/8	1/8	3/8	3/8
2	1/8	1/8	9/16	3/16
3	3/16	1/16	3/8	3/8
4	3/16	1/16	9/16	3/16
5	1/4	1/4	1/4	1/4
6	1/4	1/4	3/8	1/8
7	3/8	1/8	1/4	1/4
8	3/8	1/8	3/8	1/8



Independence concepts for credal sets

Stochastic independence/irrelevance (precise case)

- X and Y stochastically independent: $P(x, y) = P(x)P(y)$
- Y stochastically irrelevant to X: $P(X|y) = P(X)$
- independence $\equiv$ irrelevance

Strong independence (imprecise case)

- X and Y strongly independent according to $K(X, Y)$ iff stochastic independence for each $P(X, Y) \in \mathrm{ext}[K(X, Y)]$
- Equivalently, Y strongly irrelevant to X, i.e., $P(X|y) = P(X)$ for each $P(X, Y) \in \mathrm{ext}[K(X, Y)]$

Epistemic irrelevance (imprecise case)

- Y epistemically irrelevant to X according to $K(X, Y)$ iff $K(X|y) = K(X)$ for each $y \in \Omega_Y$
- Asymmetric! Simmetrization defined epistemic independence

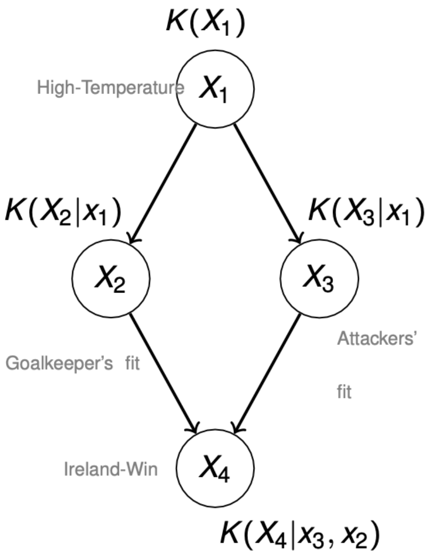
Basic operations with **strong** credal sets

	PRECISE Mass functions	IMPRECISE Credal sets
Joint	$P(X,Y)$	$K(X,Y)$
Marginalization	$P(X)$ s.t. $p(x) = \sum_y p(x,y)$	$K(X) = \left\{ P(X) \,\middle\vert\, \begin{array}{l} P(x) = \sum_y P(x,y) \\ P(X,Y) \in K(X,Y) \end{array} \right\}$
Conditioning	$P(X\|y)$ s.t. $p(x\|y) = \frac{P(x,y)}{\sum_y P(x,y)}$	$K(X\|y) = \left\{ P(X\|y) \,\middle\vert\, \begin{array}{l} P(x\|y) = \frac{P(x,y)}{\sum_y P(x,y)} \\ P(X,Y) \in K(X,Y) \end{array} \right\}$
Combination	$P(x,y) = P(x\|y)P(y)$	$K(X\|Y) \otimes K(Y) = \left\{ P(X,Y) \,\middle\vert\, \begin{array}{l} P(x,y)=P(x\|y)P(y) \\ P(X\|y) \in K(X\|y) \\ P(Y) \in K(Y) \end{array} \right\}$

Operationally, computations can be done on the extreme points only

Credal networks

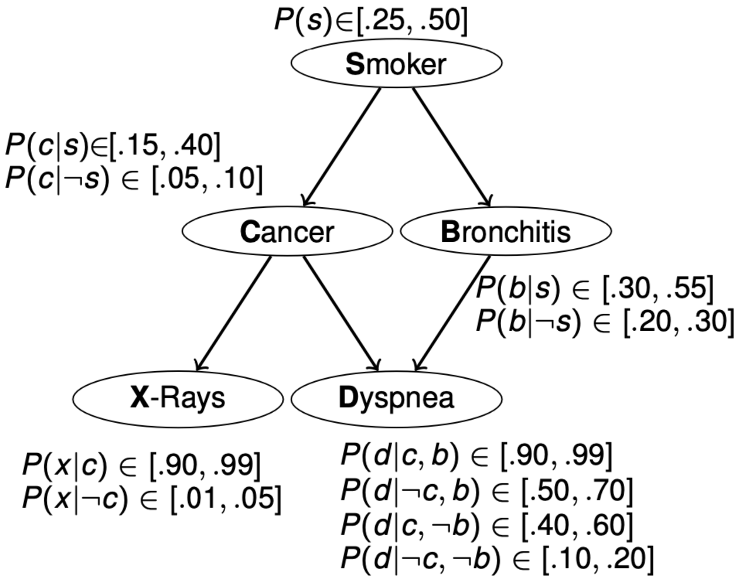
- Generalization of BNs to imprecise probabilities
- Credal sets instead of prob mass functions
  $\{P(X_i|\mathrm{pa}(X_i))\} \Rightarrow \{K(X_i|\mathrm{pa}(X_i))\}$
- Strong (instead of stochastic) independence
- Convex set of joint mass functions
  $K(X_1,\ldots,X_n) = \mathrm{CH}\left\{P(X_1,\ldots,X_n)\right\}$
  $P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i|\mathrm{pa}(X_i))$ $\forall P(X_j|\mathrm{pa}(X_j)) \in K(X_j|\mathrm{pa}(X_j))$ $\forall i = 1,\ldots,n$ $\forall \mathrm{pa}(X_i)$
- Every conditional mass function takes values in its credal set independently of the others
  CN $\equiv$ (exponential) number of BNs



E.g., $K(X_1)$ defined by constraint $P(x_1) > .7$, very likely to be warm

Simple Example of Credal network

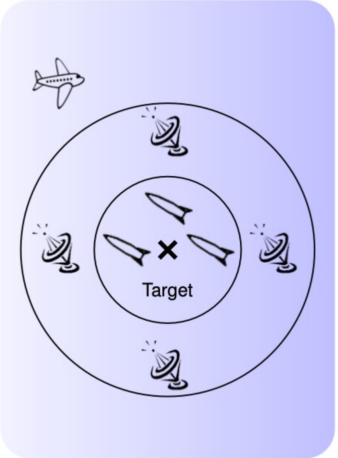
- Five Boolean variables
- Conditional independence relations given by a DAG
- The strong extension $K(S, C, B, X, D) =$



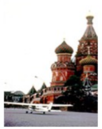
$$\mathrm{CH}\left\{ P(S, C, B, X, D) \;\middle|\; \begin{array}{l} P(s, c, b, x, d) = P(s)P(c|s)P(b|s)P(x|c)P(d|c, b) \\ P(S) \in K(S), P(C|s) \in K(C|s), \ldots \end{array} \right\}$$

No-fly zones surveyed by the Swiss Air Force

- Around important potential targets (eg. WEF, dams, nuke plants)
- Twofold circle wraps the target
  - External no-fly zone (sensors)
  - Internal no-fly zone (anti-air units)
- An aircraft (intruder) enters the zone
- Its presence, speed, height, . . . revealed by the sensors
- A team of military experts decides what the intruder intends to do



renegade



provocateur

damaged

erroneous

Difficult identification task for the experts
sensors reliabilities affected by geo/meteo conditions

Decision Making with CSs

- Most probable state x^* of X?
- Precise knowledge $P(X)$

  $x^* = \arg\max_{x \in \Omega_X} P(x)$

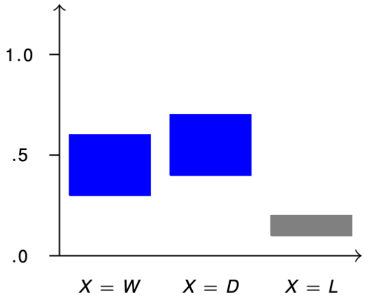
  (with 0/1 utilities)
- Imprecise knowledge $K(X)$?
- Compute lower/upper probs

  and obtain (set of) optimal states:

  $\Omega_X^* = \left\{ x \middle| \nexists x' \text{ s.t. } \underline{P}(x') > \overline{P}(x) \right\}$

  this is interval dominance
- More informative criterion: maximality

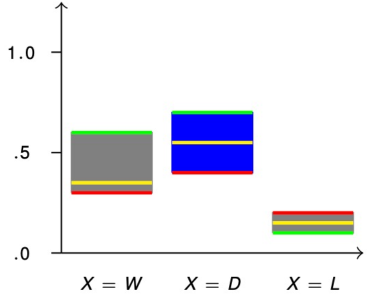
  $\left\{ x \middle| \nexists x' \text{ s.t. } P(x') > P(x) \forall P(X) \in K(X) \right\}$



$P(X) \in \begin{matrix} [.3, .6] \\ [.4, .7] \\ [.1, .2] \end{matrix}$

Decision Making with CSs

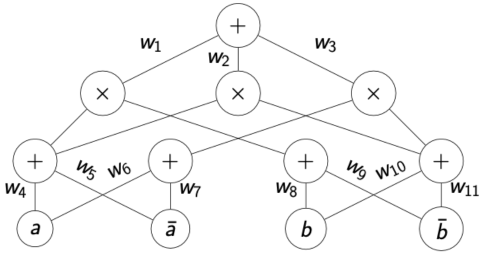
- Most probable state x^* of X?
- Precise knowledge $P(X)$
  $x^* = \arg\max_{x \in \Omega_X} P(x)$
  (with 0/1 utilities)
- Imprecise knowledge $K(X)$?
- Compute lower/upper probs
  and obtain (set of) optimal states:
  $\Omega_X^* = \left\{ x \middle| \nexists x' \text{ s.t. } \underline{P}(x') > \overline{P}(x) \right\}$
  this is interval dominance
- More informative criterion: maximality
  $\left\{ x \middle| \nexists x' \text{ s.t. } P(x') > P(x) \forall P(X) \in K(X) \right\}$



$P(X) \in \begin{matrix} [.3, .6] \\ [.4, .7] \\ [.1, .2] \end{matrix}$

Credal Sum-Product Networks

- Robustify SPNs by allowing weights to vary inside sets (for instance, towards sensitivity analisys on SPN's inference).
- Class of tractable imprecise graphical models.



$$
\left\{ \begin{array}{ll}
(w_1, w_2, w_3) \in \mathrm{CH}(& [0.28, 0.45, 0.27], \\
[0.18, 0.55, .27], & [0.18, 0.45, 0.37]), \\
0.54 \leq w_4 \leq 0.64, & 0.36 \leq w_5 \leq 0.46, \\
0.09 \leq w_6 \leq 0.19, & 0.81 \leq w_7 \leq 0.91, \\
0.27 \leq w_8 \leq 0.37, & 0.63 \leq w_9 \leq 0.73, \\
0.72 \leq w_{10} \leq 0.82, & 0.18 \leq w_{11} \leq 0.28, \\
w_4 + w_5 = 1, & w_6 + w_7 = 1, \\
w_8 + w_9 = 1, & w_{10} + w_{11} = 1
\end{array} \right\}
$$

Attack on/Sensitivity of Parameters (wrt predictions)

Sensitivity analysis

Perturb the model parameters until the predicted class changes.
(Can be also done as a perturbation of the data.)

ϵ-contamination of a vector of parameters $\boldsymbol{w}$

$$C_{\boldsymbol{w},\epsilon} = \{(1-\epsilon)\boldsymbol{w} + \epsilon\boldsymbol{v} \colon v_j \geq 0, \sum v_j = 1\}$$

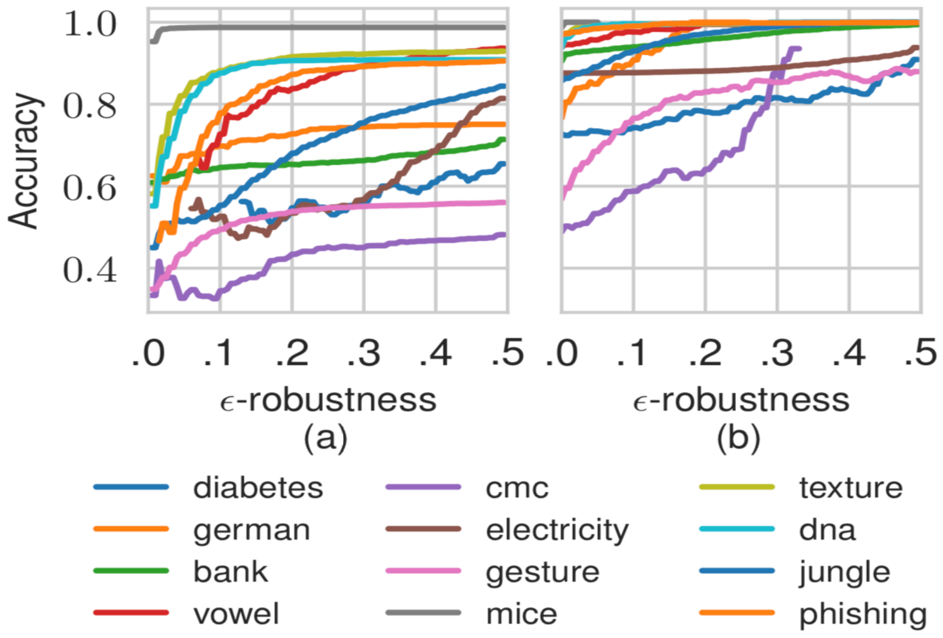
ϵ-robustness

The largest ϵ for which all parameters in $C_{w,\epsilon}$ yield the same classification.

$$\forall y' \neq y \colon \max_{\boldsymbol{w} \in C_{\boldsymbol{w},\epsilon}} \mathbb{E}_{\boldsymbol{w}} [\mathbb{I}(Y = y') - \mathbb{I}(Y = y) \mid \boldsymbol{x}] < 0$$

TU/e

Robust Classification: ϵ-robustness correlates to accuracy



Accuracy of predictions with ϵ-robustness (a) below and (b) above different thresholds for 12 OpenML datasets.

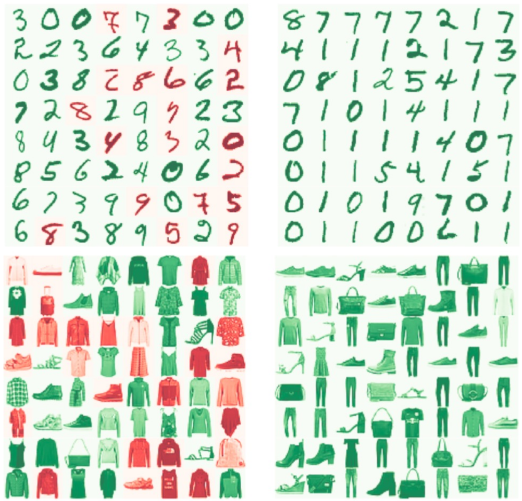
Conformal predictions

Rejection rule

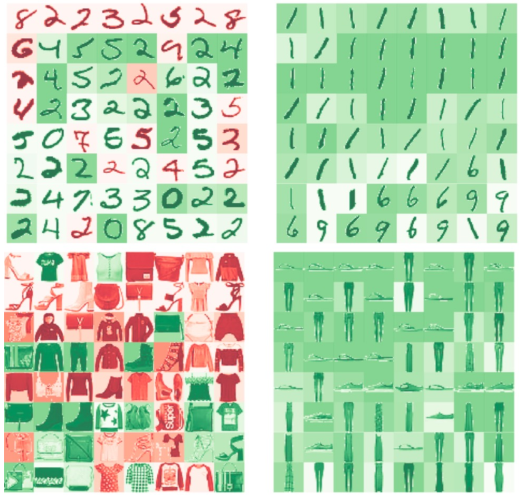
TU/e

Robust Classification

ϵ-robustness differs substantially from $p(x)$

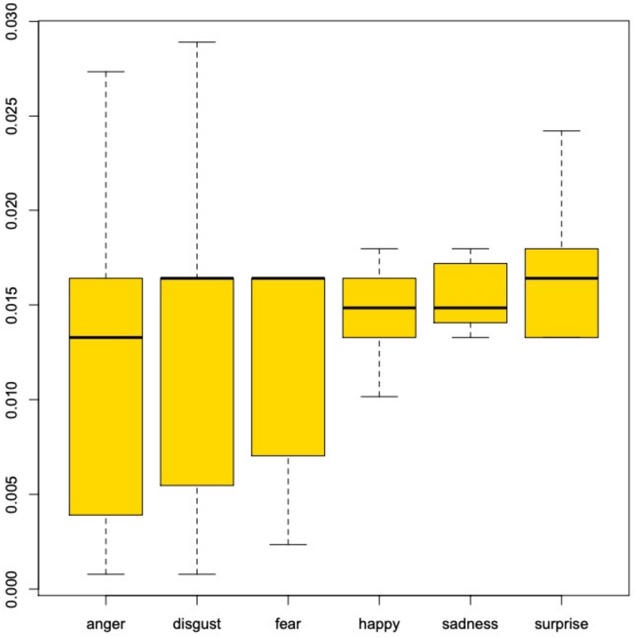


Samples from (Fashion-)Mnist datasets with lowest (left) and highest (right) ϵ-robustness in the test set.

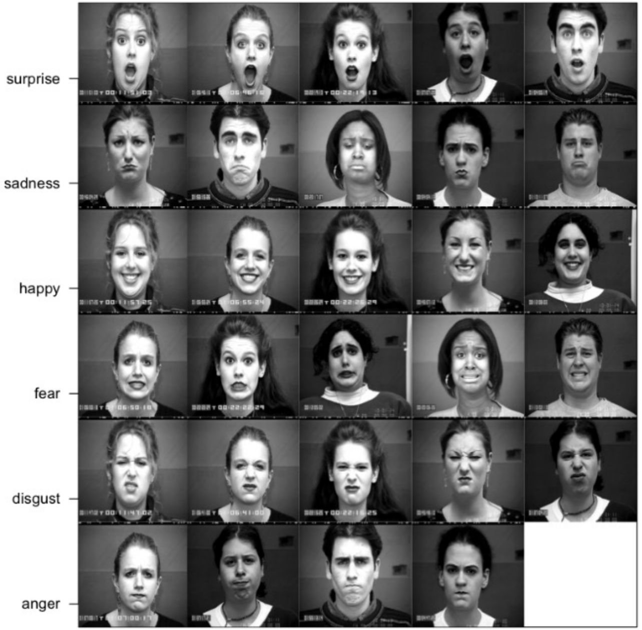


Samples from (Fashion-)Mnist datasets with lowest (left) and highest (right) $p(x)$ in the test set.

Robustness measure in classification



(a) Robustness split by emotions.



(b) Examples of emotions.

TU/e

Ongoing Research

- Credal circuits for portfolio optimisation
- Credal clustering for learning more robust deep models
- Credal sets to combine probabilistic propositional logic with deep ML models

A difficulty with circuits (if not generated by compilation): structure learning!

Ongoing: Probabilistic propositional logic to Credal Bayesian nets to credal prob. circuits

- Unpublished work: an invitation to join the challenge?

- Build a credal Bayesian net with probabilistic propositional logic (PPL) assessments
  - Somehow force bounded treewidth induced by the assessments
  - Possibly run structure learning with bounded treewidth too

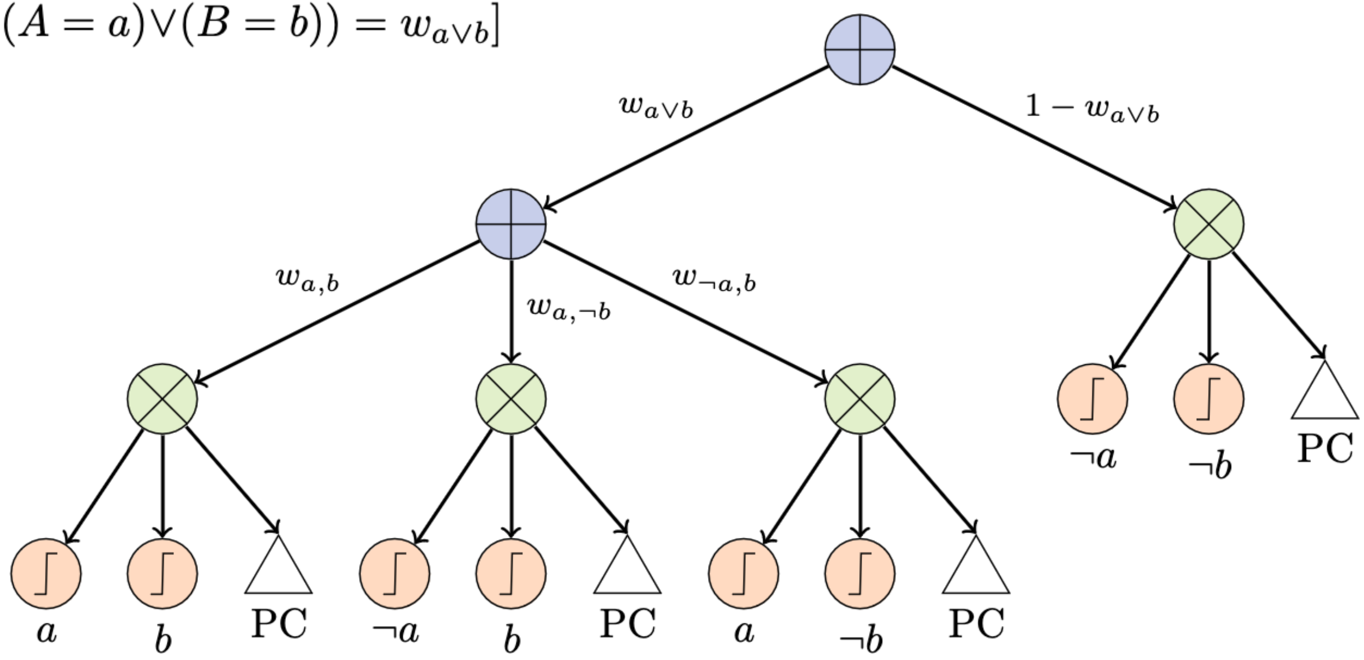
- Translate this network into a credal probabilistic circuit (akin to Darwiche's compilation)

- Train (some) model parameters of this circuit

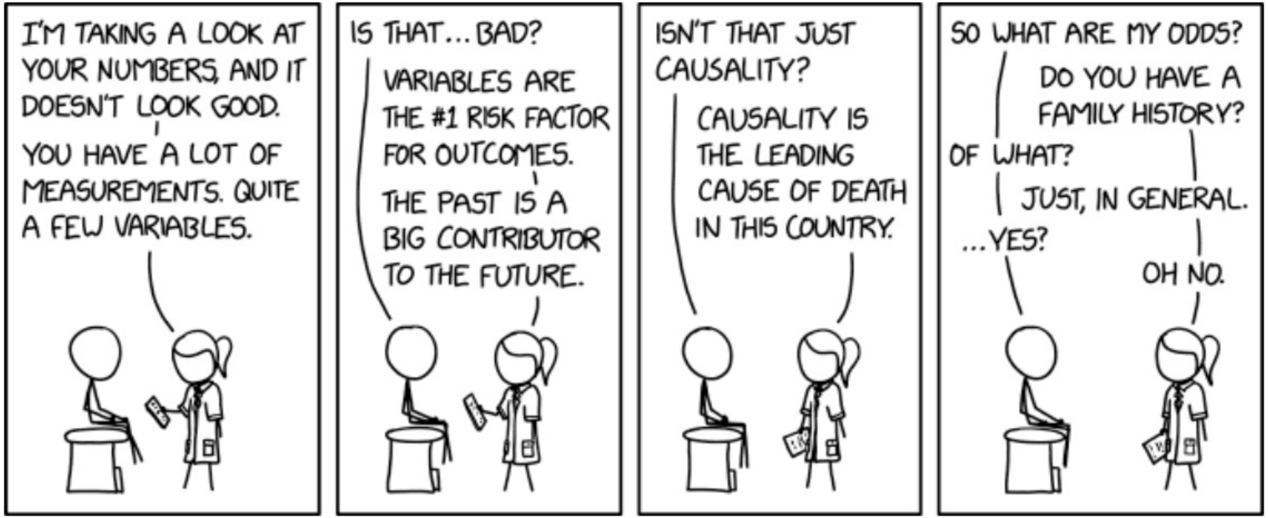
Result: a sort of neuro-symbolic AI?

Ongoing: Probabilistic propositional logic to Credal Bayesian nets to credal prob. circuits

$[P((A = a) \vee (B = b)) = w_{a \vee b}]$



Thank you for your attention



https://xkcd.com/2620/

Thanks for Alvaro Correia, Alessandro Antonucci, Soroush Ghandi for (parts of) slides and content