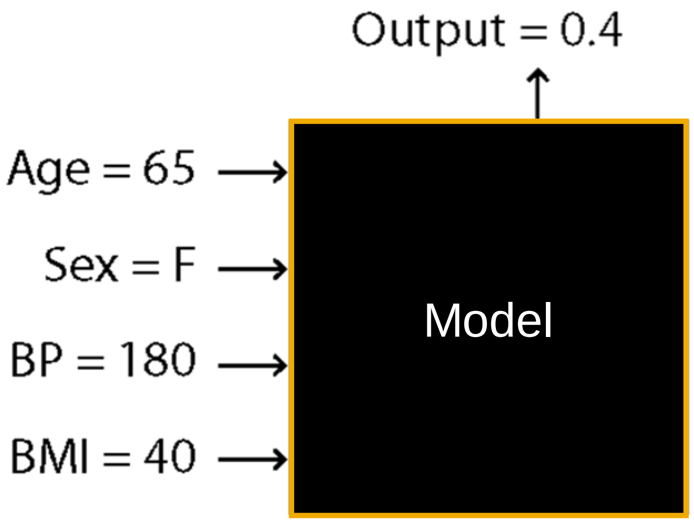
Shapley Values
and Bayesian Network

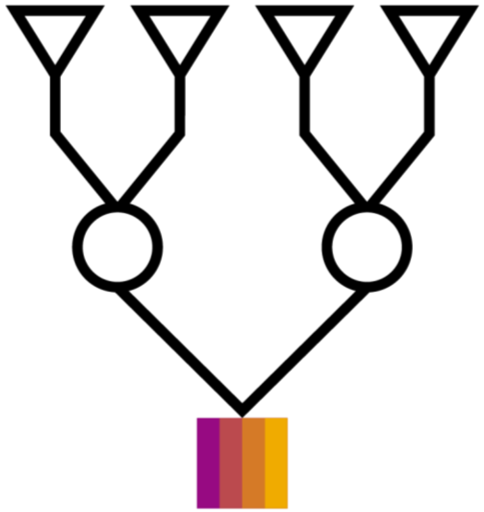
Mahdi HADJ ALI, SAP & LIP6
October 11, 2021

PUBLIC

THE BEST RUN SAP

Trust and explainability





- **ML Health:**the ML model and production deployment system must be healthy - ie behaving in production as expected and within norms specified by the data scientist.
- **ML Security:**the ML algorithm must be healthy and explainable in the face of malicious or non-malicious attacks - ie efforts to change or manipulate its behavior.
- **MLreproducibility :**All predictions must be reproducible.
- **ML Explainability:**It must be possible to determine why the ML algorithm behaved the way that it did for any particular prediction and what factors led to the prediction..

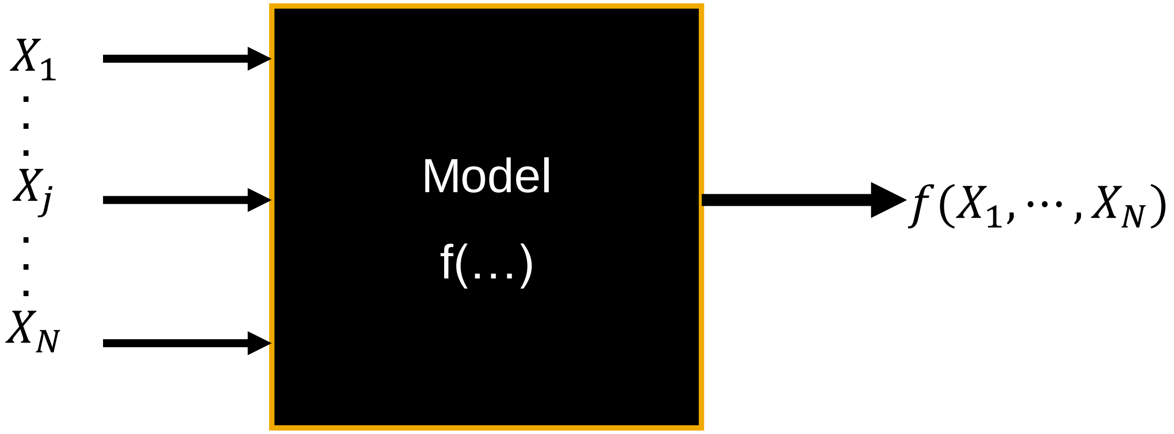
Diary

- Shapley Values
- Shapley Values in Bayesian Network
- Shapley Values in Causal Model
- Bayesian Networks ⇋ Predictive Models

Shapley Values

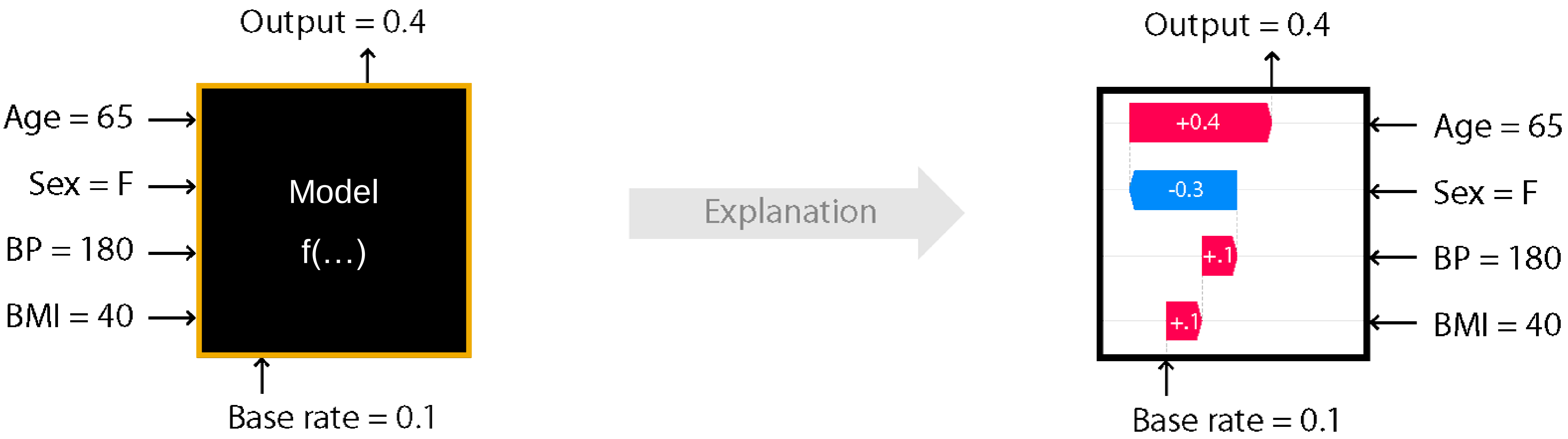
PredicitveModel: task

- Binary class prediction problem$\boldsymbol{Y}$,
- Database composed of $\boldsymbol{N}$ variables: $\boldsymbol{X} = \{ X_1, X_2, \cdots, X_j, \cdots, X_N \}$ and $\boldsymbol{D}$ rows.
- $f(X_1, \cdots, X_n)$ prediction function that takes those variables as inputs.

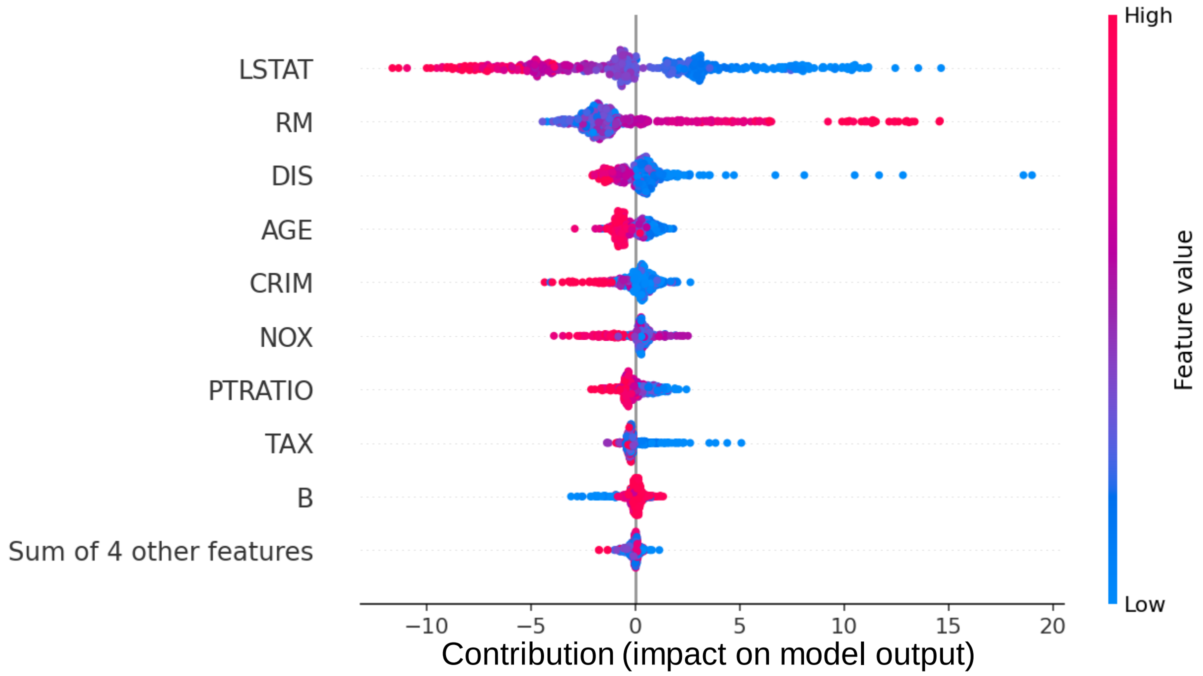


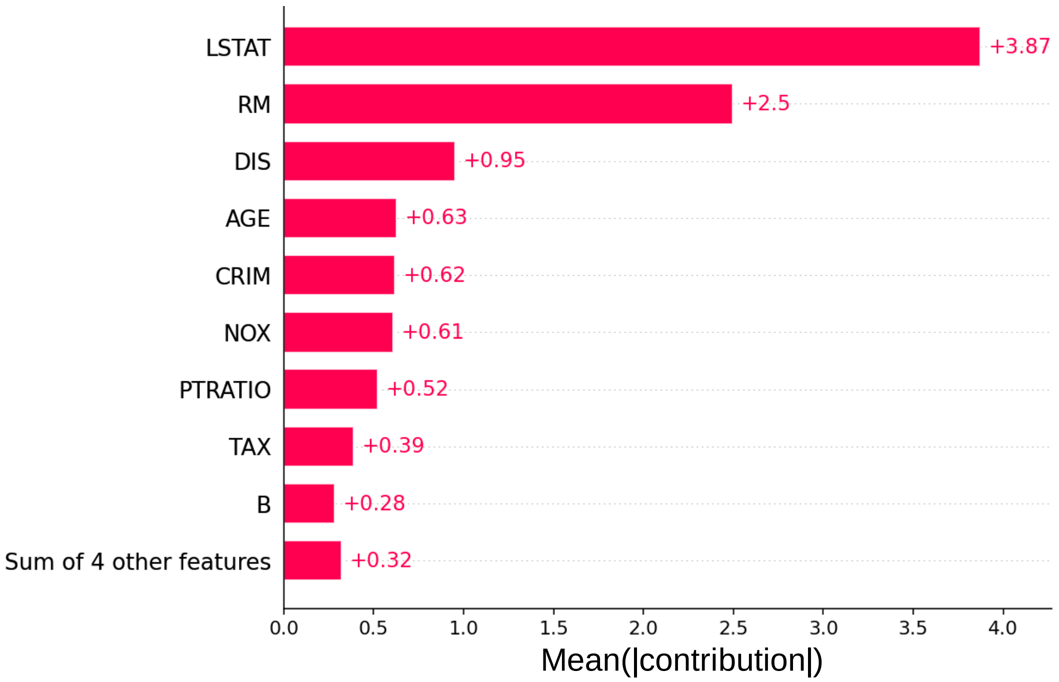
Contribution analysis: each line

Index	Age	Sex	BP	BMI
…	…	…	…	…
1953	65	F	180	40
…	…	…	…	…



Contribution analysis: all database





Graphics fromofSM Lundberg et al., "Explainable machine-learning predictions for the prevention ofhypoxaemiaduring surgery", Nat BiomedEng, flight. 2, no. 10, p. 749-760, Oct. 2018.

Shapley Values

- **Lloyd Shapley 1953**
- **Cooperative game theory**
- **Fair distribution**

Shapley Value formula for the player $\mathbf{X}_i$:

$$\phi_{X_i} = \sum_{S \subseteq \boldsymbol{X}/\{X_i\}} \frac{|S|!\,(N-|S|-1)!}{N!}\left(v(S \cup \{X_i\}) - v(S)\right)$$

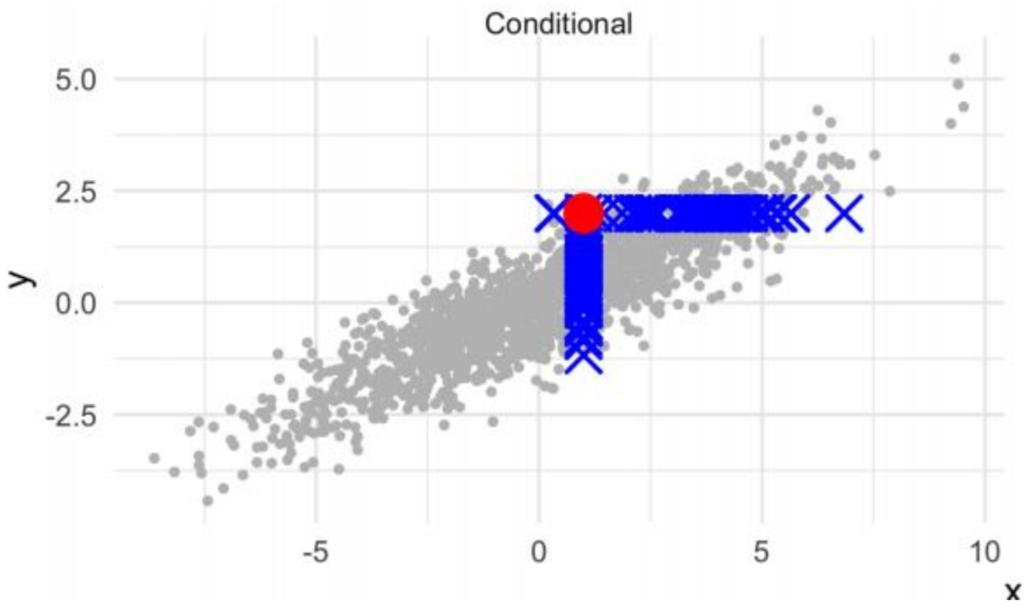
With $\boldsymbol{N}$**:** Number of players, $\boldsymbol{S}$: Coalition of players, $\boldsymbol{X_i}$: ith player and $\boldsymbol{v}(\boldsymbol{S})$: worth of coalition $\boldsymbol{S}$.

Definition of function *v (Conditional)*

Shapley Values Conditionals

$$v(S) = \mathbb{E}[f(x_S, X_{\bar{S}}) | X_S = x_S]$$

$$= \int P(\boldsymbol{X}_{\bar{S}} | \boldsymbol{x}_S)\, f(\boldsymbol{X}_{\bar{S}}, \boldsymbol{x}_S)\, d\boldsymbol{X}_{\bar{S}}$$



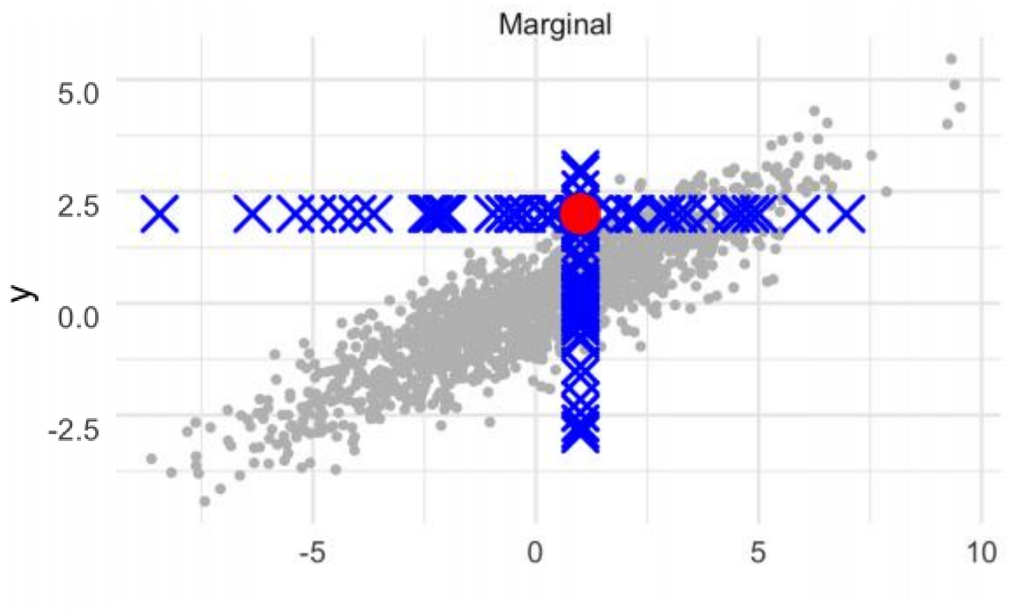
- Best estimate of $\boldsymbol{f}$ given $\boldsymbol{S}$.
- Analysis on the distribution of the data, at X fixed we are on the manifold.
- Possibly a non-zero value for a variable not used by the model.

Graphics from:Frye,rowat, FeigeATsymmetric Shapley values: incorporating causal knowledge into model-agnostic Explainability. Advances in Neural Information Processing Systems

Definitions of function *v (Marginal)*

Shapley Value Marginals

$$v(S) = \mathbb{E}[f(x_S, X_{\bar{S}})] = \int P(\boldsymbol{X}_{\bar{S}})\, f(\boldsymbol{X}_{\bar{S}}, \boldsymbol{x}_S)\, d\boldsymbol{X}_{\bar{S}}$$



- Marginal Expectation.
- Maycreate unrealistic data.
- Always a null value for a variable not used by the model.

Graphics from:Frye,rowat, FeigeATsymmetric Shapley values: incorporating causal knowledge into model-agnostic Explainability. Advances in Neural Information Processing Systems

TreeExplainer

Shap values are very expensive to calculate.

- The algorithm **TreeExplainer** is one of the fastest.

This approach uses the information computed during the training of a forest of decision trees.

- Optimized for decision trees, its complexity goes from $O\left(TLM2^{N}\right)$ à $O\left(TLP^{2}\right)$[1].

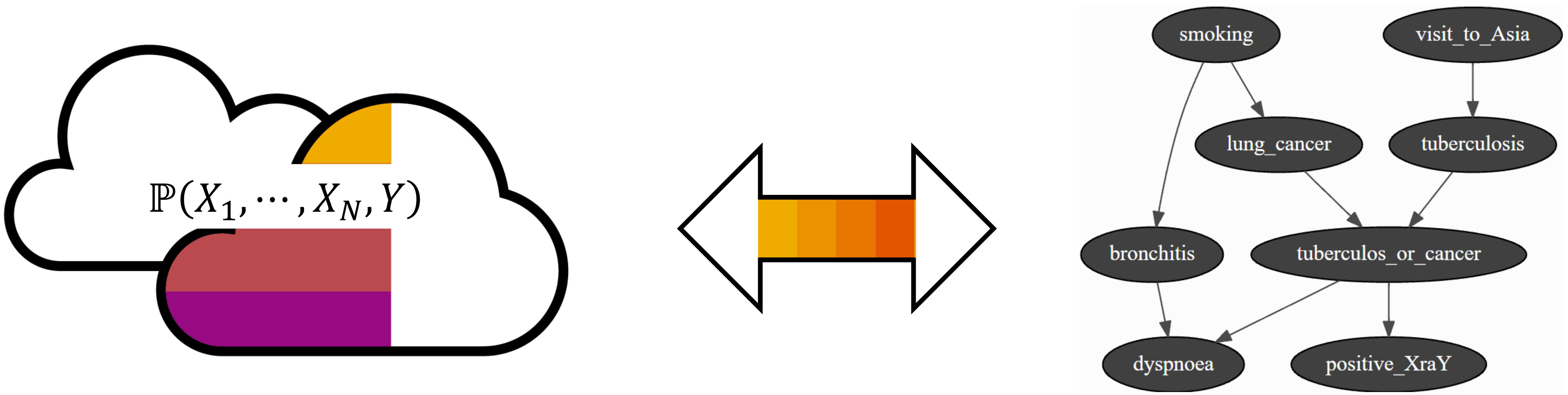
With : **T** the number of trees, **L** the maximum number of leaves in a tree, **N** the number of variables, **P** maximum tree depth

- Give **an approximate result**, they are neither marginals nor conditionals.

[1] Lundberg, SM,Erion, G., Chen, H. et al. From local explanations to global understanding with explainable AI for trees. Nat MachIntel ***2,****56–67 (2020).*

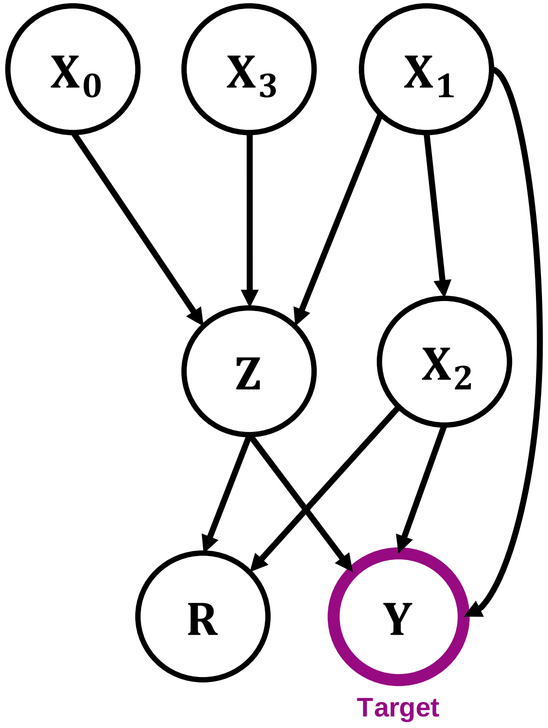
Shapley values and Bayesian Networks

Prediction and Bayesian Networks



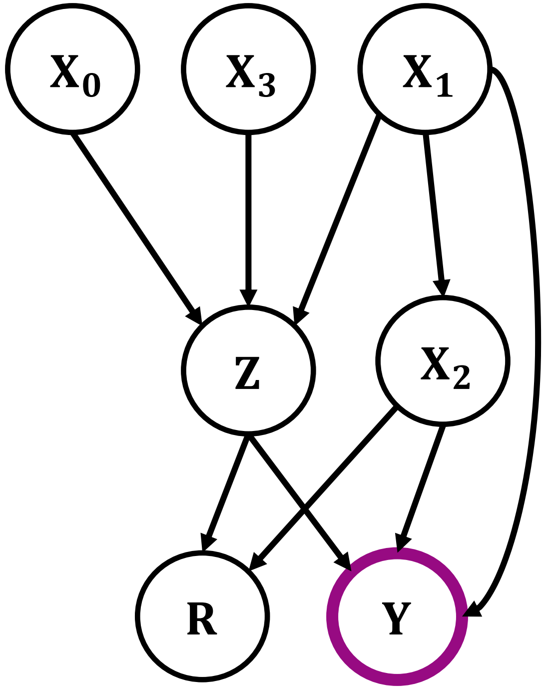
- The prediction of $\boldsymbol{Y}$ is given by $\mathrm{P}(\mathrm{Y}|\mathrm{X}_1 \ \cdots, \mathrm{X}_\mathrm{N})$ obtained from the joint distribution.
- We use the $logit(P(Y| \dots))$ in order to have an additive explanation.

Inference Exact



- Compute new probabilistic information from a Bayesian network and some observations.
- Exact inference calculates the posterior distribution for some variable in Bayesian networks given (partial) observations.
- $v(\{X_1, X_2\}) = logit(P(Y = 1|\ X_1 = x_1^d, X_2 = x_2^d))$

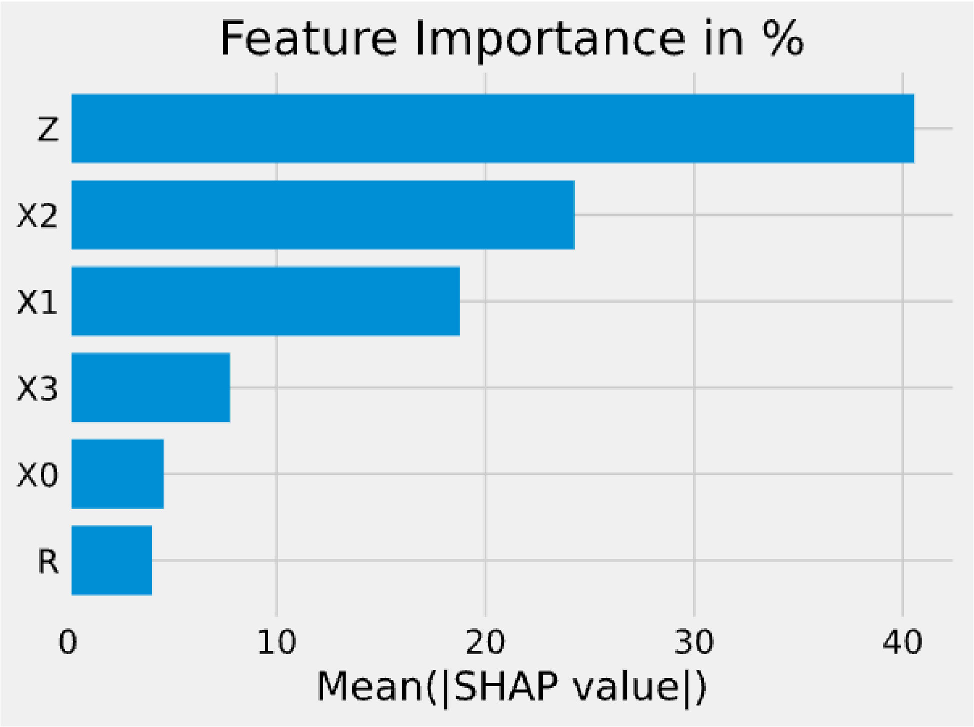
Simplification in Bayesian Networks

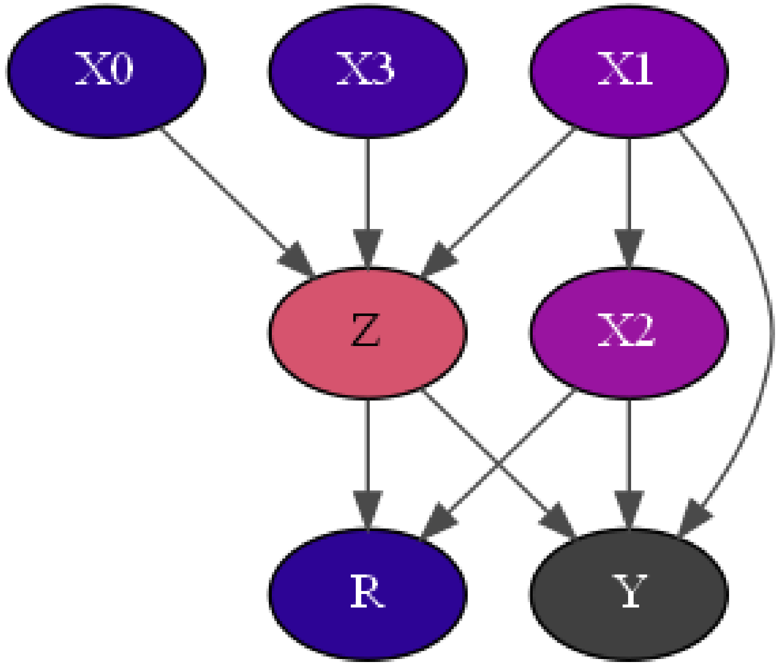


Possible combinations: $\mathbf{2^N}$

- V-structures and other graph specifications help us know which coalitions are interesting to compute.
- $v(\{Z, X_1, X_0\}) - v(\{Z, X_1\}) = 0$ because $Y \perp X_0 | Z$
- $v(\{Z, X_1, X_0\})$ *and* $v(\{Z, X_1\})$ are exchangeable.
- For marginal Shapley values: only the Markov Blanket matters.

Significance of variables





Shapley values and Causal Models

Shapley values causal

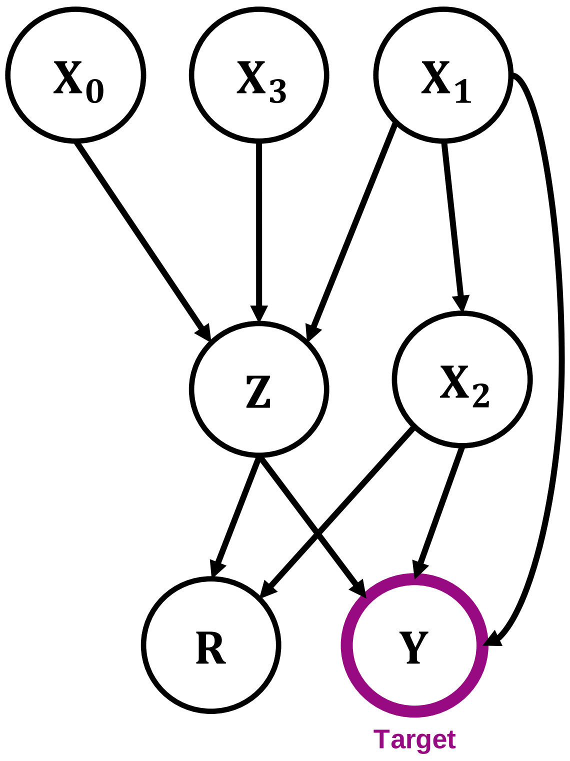
$$v(S) = \mathbb{E}[f(\boldsymbol{X})|do(\boldsymbol{X_S} = \boldsymbol{x_s})] = \int P(\boldsymbol{X}_{\bar{S}}|\, do(\boldsymbol{X}_S = \boldsymbol{x}_S))\, f(\boldsymbol{X}_{\bar{S}}, \boldsymbol{x}_S)\, d\boldsymbol{X}_{\bar{S}}$$ [2]

- To take into account the possible causal relationships between the 'in-coalition' characteristics and the 'out-of-coalition' characteristics, we condition 'by intervention' for which we use the do-calculus of Pearl.

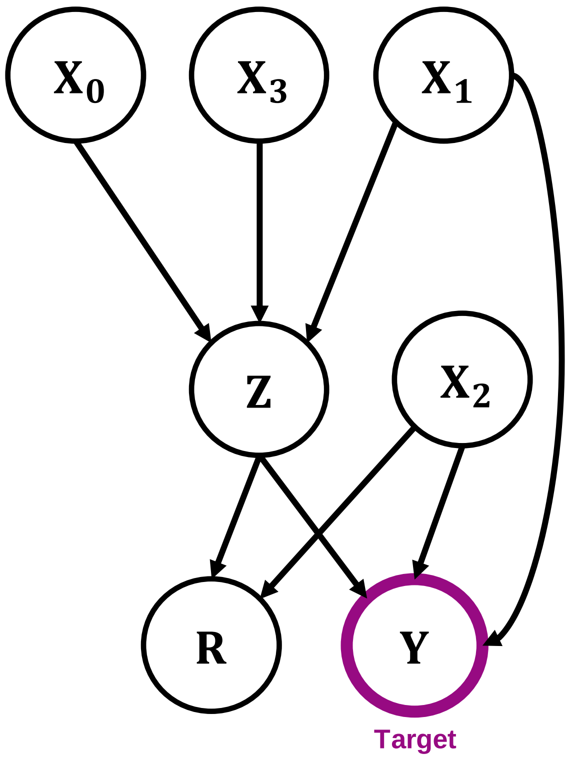
- The contribution ϕ_{X_i} measures the relevance of the variable X_i through the (average) prediction obtained if we intervene on the characteristic X_i at its value x_i with respect to (the counterfactual situation of) not knowing its value.

[2] TomHeskes, ,Evi Sijben,JohnGabriel Bucur, and Tom Claassen. "Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models." (2020).

Do-calculuswithout latent variable:Graph Mayhem

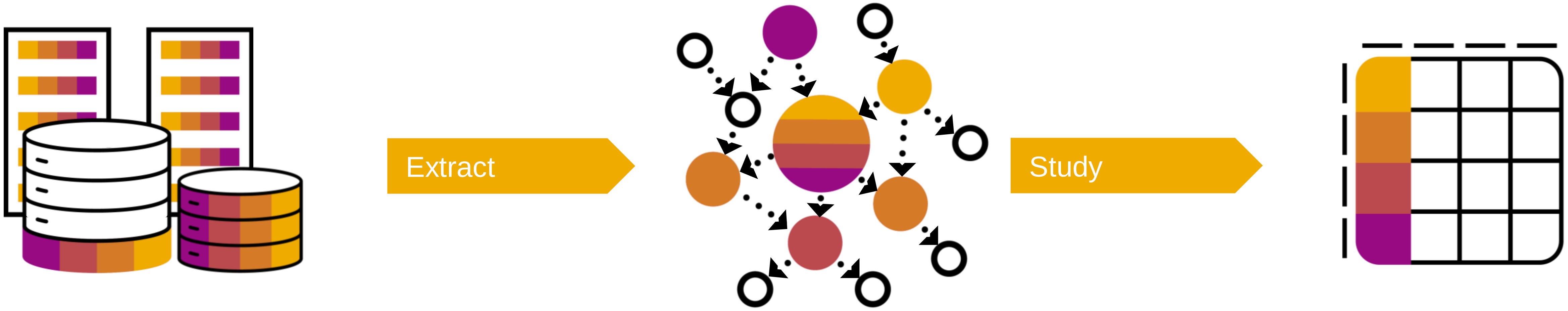


$$do(X_2 = x_2)$$



Bayesian networks ⇋ Predictive Models

Bayesian networks⟶Predictive analysis



Drive the predictive analysis:

- Do not take the consequences of the Target
- Markov blanket for variable selection

Bayesian networks⟵Predictive analysis

Graph Discovery:

- TreeShap and Marginal to find the Markov Blanket

References

- SMLundberget al., "Explainablemachine-learning predictionsfor thepreventionofhypoxaemia during surgery», NatBiomedEng, vol. 2, no. 10, p. 749-760, Oct. 2018.
- Frye,rowat, FeigeAsymmetricShapley values:incorporatedcausalknowledge intomodel-agnosticexplainability.Advancesin Neural InformationProcessing Systems
- Lundberg, SM,Erion, G., Chen, H. et al. From local explanations to global understanding with explainable AI for trees. Nat MachIntel2, 56–67 (2020).
- tomHeskes, ,Evi Sijben,JohnGabriel Bucur, and Tom Claassen. "Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models." (2020).