# Explaining probabilistic predictions on the simplex with Shapley compositions

**Paul-Gauthier Noé** [1], Miquel Perelló-Nieto [2],
Peter Flach [2], Jean-François Bonastre [1]

[1]Avignon Université

[2]University of Bristol

LIA seminar, Avignon Université
January 18, 2024

# Introduction

**Local explanation in machine learning:**

$$\boldsymbol{x} \in \mathbb{R}^d \longrightarrow \boxed{\text{Model}} \longrightarrow \text{Prediction}$$

*Given one instance* $\boldsymbol{x}$ *with* $d$ *features, what is the contribution/effect of a feature's value on the prediction?*

$\neq$ Global explanation

# Introduction

Examples of local explanation methods:

- Local Interpretable Model-Agnostic Explanations (LIME)[1],
- Shapley values[2] (SHAP toolkit[3])

[1] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "" Why should i trust you?" Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1135–1144.

[2] Erik Štrumbelj and Igor Kononenko. "Explaining prediction models and individual predictions with feature contributions". In: *Knowledge and information systems* 41 (2014), pp. 647–665.

[3] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774.

# Introduction

**Shapley values in cooperative game theory**[4]

- Distributes the total payoff among the players.
- The unique quantity respecting a set of desired axiomatic properties:
  - Linearity:

$$\phi_{\alpha v+(1-\alpha)w}(i) = \alpha\phi_v(i) + (1-\alpha)\phi_w(i), \tag{1}$$

    for a player $i$ and two games $v$ and $w$, and for $\alpha \in [0,1]$
  - Efficiency,

$$\sum_{i\in\mathcal{C}} \phi_v(i) = v(\mathcal{C}), \tag{2}$$

    (the sum of the value is equal to the total payoff)
  - Symmetry

[4] Lloyd S Shapley et al. "A value for n-person games". In: (1953), pp. 307–317.

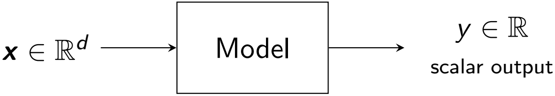
# Introduction

**Shapley values in machine learning**

# Introduction

**Shapley values in machine learning**

- Features are treated as players, and the scalar output of the model as the payoff,



# Introduction

### Shapley values in machine learning

- Features are treated as players, and the scalar output of the model as the payoff,

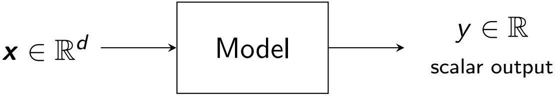
$$\boldsymbol{x} \in \mathbb{R}^d \longrightarrow \boxed{\text{Model}} \longrightarrow \underset{\text{scalar output}}{y \in \mathbb{R}}$$

- Binary classifier, regressor with one-dimensional output ✓

# Introduction

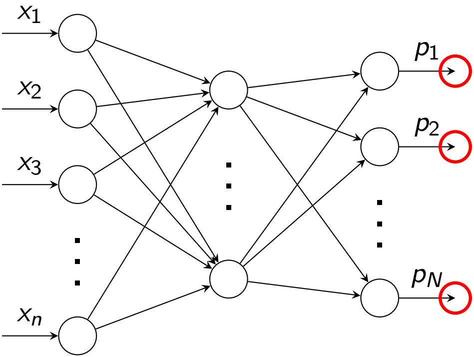
**Shapley values in machine learning**

- Features are treated as players, and the scalar output of the model as the payoff,



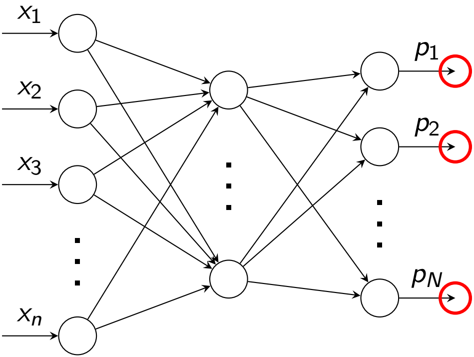

- Binary classifier, regressor with one-dimensional output ✓
- Multiclass classifier ✗
  ex: The output of a softmax lives on a multidimensional simplex!

# Introduction




Some explain the output one-by-one,

# Introduction

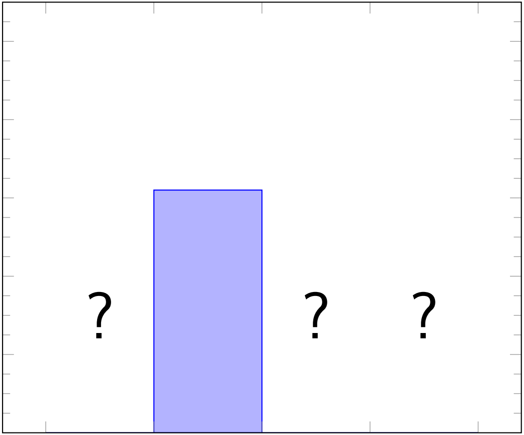



Some explain the output one-by-one,

But a probability distribution lives on a simplex,
**The relative information matter!!**

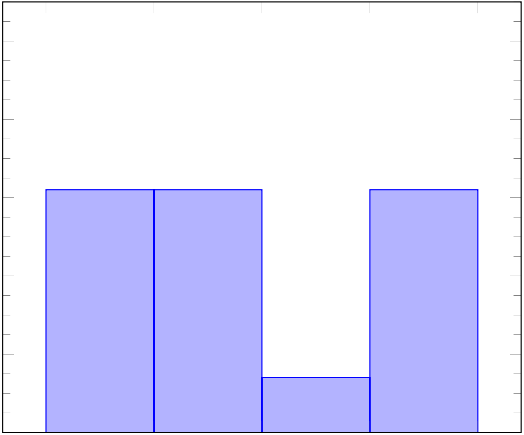
# Introduction

The relative information between parts matter!



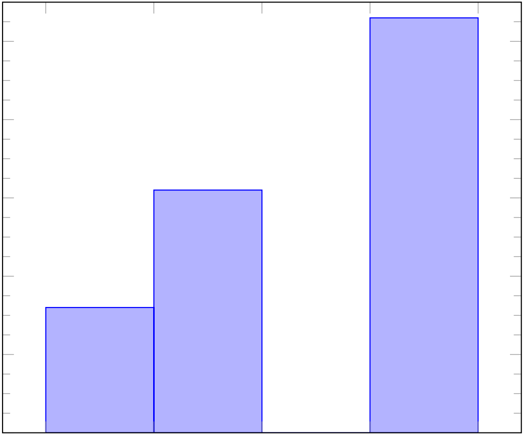
# Introduction

The relative information between parts matter!

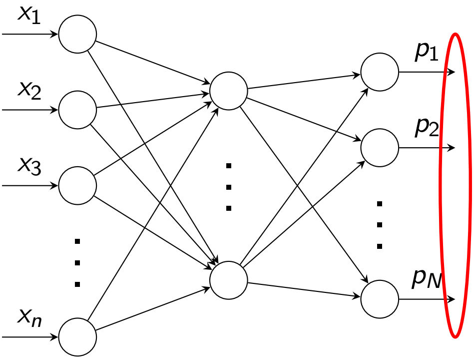


# Introduction

The relative information between parts matter!



# Introduction




We will explain the probablities all together using the *Aitchison geometry of the simplex*[5].

[5] John Aitchison. "The Statistical Analysis of Compositional Data". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 44.2 (1982), pp. 139–177, Vera Pawlowsky-Glahn, Juan José Egozcue, and Raimon Tolosana-Delgado. *Modeling and Analysis of Compositional Data*. John Wiley & Sons, 2015.

# Outline

# The Shapley values in machine learning

We want to explain a prediction $f(\boldsymbol{x})$ on the instance $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$, where $f : \mathcal{X} \to \mathbb{R}$ is the learned model.

$$\boldsymbol{x} \longrightarrow \boxed{f} \longrightarrow \underset{\text{prediction}}{f(\boldsymbol{x}) \in \mathbb{R}}$$

# The Shapley values in machine learning

We want to explain a prediction $f(\boldsymbol{x})$ on the instance $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$, where $f : \mathcal{X} \to \mathbb{R}$ is the learned model.

$$\boldsymbol{x} \longrightarrow \boxed{f} \longrightarrow \underset{\text{prediction}}{f(\boldsymbol{x}) \in \mathbb{R}}$$

Some notation. Let:

- Pr be the probability distribution over $\mathcal{X}$ of the data.

# The Shapley values in machine learning

We want to explain a prediction $f(\boldsymbol{x})$ on the instance $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$, where $f : \mathcal{X} \to \mathbb{R}$ is the learned model.

$$\boldsymbol{x} \longrightarrow \boxed{f} \longrightarrow \underset{\text{prediction}}{f(\boldsymbol{x}) \in \mathbb{R}}$$

Some notation. Let:

- Pr be the probability distribution over $\mathcal{X}$ of the data.
- $S \subseteq \mathcal{I} = \{1, 2, \dots d\}$ be a subset of indices,

# The Shapley values in machine learning

We want to explain a prediction $f(\boldsymbol{x})$ on the instance $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$, where $f : \mathcal{X} \to \mathbb{R}$ is the learned model.

$$\boldsymbol{x} \longrightarrow \boxed{f} \longrightarrow \underbrace{f(\boldsymbol{x}) \in \mathbb{R}}_{\text{prediction}}$$

Some notation. Let:

- Pr be the probability distribution over $\mathcal{X}$ of the data.
- $S \subseteq \mathcal{I} = \{1, 2, \dots d\}$ be a subset of indices,
- $\boldsymbol{x}_S$ refers to an instance $\boldsymbol{x}$ restricted to the features indicated by the indices in $S$.

# The Shapley values in machine learning

The **prediction difference** (knowing only features indexed by $S$):

$$\begin{aligned} \delta_{f,\boldsymbol{x},\mathrm{Pr}} : 2^{\mathcal{I}} &\to \mathbb{R}, \\ S &\mapsto \mathbb{E}_{\mathrm{Pr}}[f(\boldsymbol{x}) \mid \boldsymbol{x}_S] - \mathbb{E}_{\mathrm{Pr}}[f(\boldsymbol{x})], \end{aligned} \tag{3}$$

where $\mathbb{E}_{\mathrm{Pr}}[f(\boldsymbol{x}) \mid \boldsymbol{x}_S] = \int_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})\mathrm{Pr}(\boldsymbol{x} \mid \boldsymbol{x}_S)d\boldsymbol{x}$.

When an instance $\boldsymbol{x}$ is observed, the expected value of the prediction is simply $\mathbb{E}[f(\boldsymbol{x}) \mid \boldsymbol{x}] = f(\boldsymbol{x})$. However, when only $\boldsymbol{x}_S$ is given with $\mathcal{S} \neq \mathcal{I}$, there is uncertainty about the non-observed features and we therefore compute the expected prediction given $\boldsymbol{x}_S$.

# The Shapley values in machine learning

The **contribution** of the feature indexed by $i \notin S$ in the prediction $f(\boldsymbol{x})$ given the known features indexed by $S$ is given by:

$$c_{f,\boldsymbol{x},\mathrm{Pr}}(i, \boldsymbol{X}_S) = \delta_{f,\boldsymbol{x},\mathrm{Pr}}(\boldsymbol{X}_{S\cup\{i\}}) - \delta_{f,\boldsymbol{x},\mathrm{Pr}}(\boldsymbol{X}_S), \tag{4}$$

This measures the contribution of the $i$th features with a particular *coalition* of features indexed by $S$.

# The Shapley values in machine learning

The whole contribution of the $i$th feature is computed by averaging this quantity over all possible coalitions of features as follows:

$$\phi_{f,\boldsymbol{x},\mathrm{Pr}}(i) = \frac{1}{d!} \sum_{\pi} c_{f,\boldsymbol{x},\mathrm{Pr}}(i, \pi_{\boldsymbol{X}}^{<i}), \tag{5}$$

where $\pi$ is a permutation of the set $\mathcal{I}$ of indexes and $\pi_{\boldsymbol{X}}^{<i}$ is the features of $\boldsymbol{X}$ coming before the $i$th feature in the ordering given by $\pi$.

# The Shapley values in machine learning

The whole contribution of the $i$th feature is computed by averaging this quantity over all possible coalitions of features as follows:

$$\phi_{f,\boldsymbol{x},\mathrm{Pr}}(i) = \frac{1}{d!} \sum_{\pi} c_{f,\boldsymbol{x},\mathrm{Pr}}(i, \pi_{\boldsymbol{X}}^{<i}), \tag{5}$$

where $\pi$ is a permutation of the set $\mathcal{I}$ of indexes and $\pi_{\boldsymbol{X}}^{<i}$ is the features of $\boldsymbol{X}$ coming before the $i$th feature in the ordering given by $\pi$.

This quantity is known as the **Shapley value** for the $i$th feature.

# The Shapley values in machine learning

It comes from cooperative game theory and is known to be the only quantity respecting a set of desired axiomatic properties[6].

[6] Lloyd S Shapley et al. "A value for n-person games". In: (1953), pp. 307–317.

# The Shapley values in machine learning

It comes from cooperative game theory and is known to be the only quantity respecting a set of desired axiomatic properties[6].

- Linearity with respect to the model ($\alpha, \beta \in \mathbb{R}$): $\phi_{\alpha f+\beta g}(i) = \alpha\phi_f(i) + \beta\phi_g(i)$,

[6] Lloyd S Shapley et al. "A value for n-person games". In: (1953), pp. 307–317.

# The Shapley values in machine learning

It comes from cooperative game theory and is known to be the only quantity respecting a set of desired axiomatic properties[6].

- Linearity with respect to the model ($\alpha, \beta \in \mathbb{R}$): $\phi_{\alpha f+\beta g}(i) = \alpha\phi_f(i) + \beta\phi_g(i)$,
- The "centered" learned model is additively separable with respect to the Shapley values:

$$f(\boldsymbol{x}) - \mathbb{E}_{\mathrm{Pr}}[f(\boldsymbol{X})] = \sum_{i=1}^{d} \phi_f(i), \tag{6}$$

which is known as the *efficiency* property.

[6] Lloyd S Shapley et al. "A value for n-person games". In: (1953), pp. 307–317.

# The Shapley values in machine learning

It comes from cooperative game theory and is known to be the only quantity respecting a set of desired axiomatic properties[6].

- Linearity with respect to the model ($\alpha, \beta \in \mathbb{R}$): $\phi_{\alpha f + \beta g}(i) = \alpha \phi_f(i) + \beta \phi_g(i)$,
- The "centered" learned model is additively separable with respect to the Shapley values:

$$f(\boldsymbol{x}) - \mathbb{E}_{\mathrm{Pr}}[f(\boldsymbol{X})] = \sum_{i=1}^{d} \phi_f(i), \tag{6}$$

  which is known as the *efficiency* property.
- Symmetry

[6] Lloyd S Shapley et al. "A value for n-person games". In: (1953), pp. 307–317.

# The Shapley values in machine learning

Example of explanation:

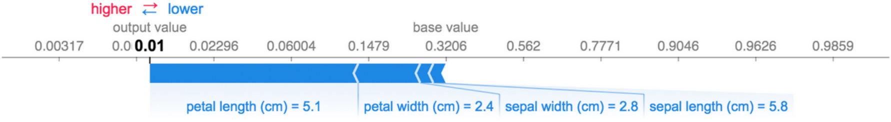
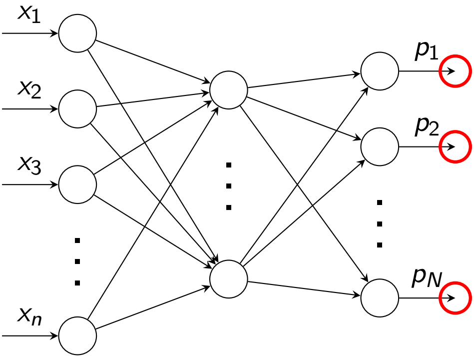



Figure: Explanation of the probability for the class Setosa for a flower from the Iris dataset. The classifier is an SVM with radial basis function and pairwise coupling. Image from `https://github.com/shap/shap/tree/master`.

Efficiency: $\underbrace{f(\boldsymbol{x})}_{\text{prediction}} - \underbrace{\mathbb{E}_{\mathrm{Pr}}[f(\boldsymbol{X})]}_{\text{base value}} = \sum_{i=1}^{d} \phi_f(i),$

Note that the Shapley explanation is ran in the *logit* domain!

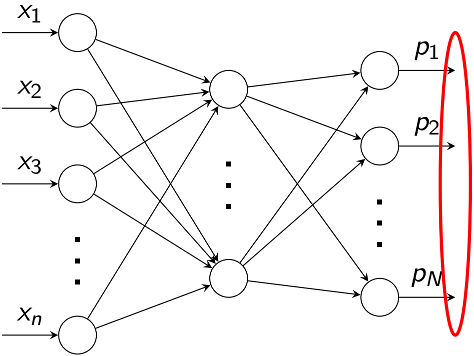
# The Shapley values in machine learning




Some explain the output one-by-one,

But a probability distribution lives on a simplex,
**The relative information matter!!**

# The Shapley values in machine learning




We will explain the probablities all together using the *Aitchison geometry of the simplex*.

# Compositional data analysis

Figure: A piece of basalt

Composition:

- 35% of pyroxene,
- 50% of plagioclase,
- 12% of olivine,
- 3% of magnetite.

$$\boldsymbol{x} = [35, 50, 12, 3]^T,$$
$$\sum_i x_i = 100\%.$$

# Compositional data analysis

A composition lives on a simplex:

$$\mathcal{S}^N = \left\{ \boldsymbol{x} = [x_1, x_2, \dots x_N]^T \in \mathbb{R}^N \mid \forall i \in [\![1, N]\!], x_i > 0 \text{ and } \sum_{i=1}^{N} x_i = k \right\}, \tag{7}$$

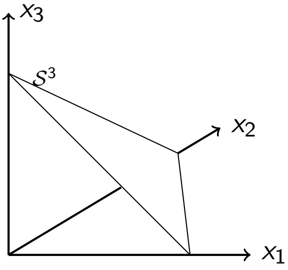



Figure: Two dimensional simplex as the sample space of three-parts compositional data.

Exemples of compositional data:

- vectors of percentages,
- vectors of concentrations,
- **discrete probability distributions** (probability simplex: $k = 1$).

# Compositional data analysis

**The Aitchison geometry of the simplex** gives to the simplex an **Euclidean vector space** structure.[7]

$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}^N$ and $\alpha \in \mathbb{R}$,

- perturbation: $\boldsymbol{x} \oplus \boldsymbol{y} = \mathcal{C}\left([x_1 y_1, \dots x_N y_N]\right)$,
- powering: $\alpha \odot \boldsymbol{x} = \mathcal{C}\left([x_1^\alpha, \dots x_N^\alpha]\right)$,
- inner product: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_a = \dfrac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{N} \log \dfrac{x_i}{x_j} \log \dfrac{y_i}{y_j}$

where, $\mathcal{C}(\boldsymbol{x}) = \left[\frac{x_1}{\|\boldsymbol{x}\|_1}, \frac{x_2}{\|\boldsymbol{x}\|_1}, \dots \frac{x_N}{\|\boldsymbol{x}\|_1}\right]^T$ is the *closure* operator.

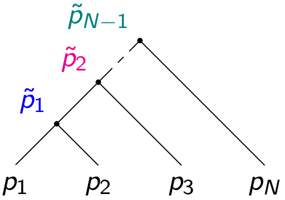
[7] John Aitchison. "The Statistical Analysis of Compositional Data". In: *Monographs on Statistics and Applied Probability* 25 (1986), John Aitchison. "Simplicial inference". In: *Algebraic Methods in Statistics and Probability*. Ed. by Donald St. P. Richards (eds.) Ams Special Session on Algebraic Methods in Statistics Marlos A. G. Viana. Contemporary Mathematics 287. American Mathematical Society, 2001.

# Compositional data analysis

## Cartesian coordinate system and the ILR transformation

$$\tilde{\boldsymbol{p}} = \mathrm{ilr}(\boldsymbol{p}) = \left[ \underbrace{\langle \boldsymbol{p}, \boldsymbol{e}^{(1)} \rangle_a}_{\tilde{\boldsymbol{p}}_1}, \underbrace{\langle \boldsymbol{p}, \boldsymbol{e}^{(2)} \rangle_a}_{\tilde{\boldsymbol{p}}_2}, \ldots \underbrace{\langle \boldsymbol{p}, \boldsymbol{e}^{(N-1)} \rangle_a}_{\tilde{\boldsymbol{p}}_{N-1}} \right]^T \in \mathbb{R}^{N-1}. \tag{8}$$

where $\{\boldsymbol{e}^{(i)}\}_{1 \leq i \leq N}$ forms an *Aitchison* orthonormal basis on the simplex[8].




$$\tilde{p}_i = \langle \boldsymbol{p}, \boldsymbol{e}^{(i)} \rangle_a = \frac{1}{\sqrt{i\,(i+1)}} \log \left( \frac{\prod_{j=1}^{i} p_j}{(p_{i+1})^i} \right). \tag{9}$$

Figure: Bifurcating tree corresponding to the orthonormal basis obtained with the Gram-Schmidt procedure.

[8] Juan José Egozcue et al. "Isometric logratio transformations for compositional data analysis". In: *Mathematical geology* 35.3 (2003), pp. 279–300.

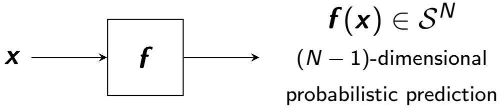
# Shapley composition on the simplex

Use the Aitchison geometry to extend the concept of Shapley value to the simplex for explaining multidimensional probabilistic predictions.

# Shapley composition on the simplex

Use the Aitchison geometry to extend the concept of Shapley value to the simplex for explaining multidimensional probabilistic predictions.

We want to explain a prediction $\boldsymbol{f}(\boldsymbol{x}) \in \mathcal{S}^N$, where $\boldsymbol{f} : \mathcal{X} \to \mathcal{S}^N$ is the learned model.



# Shapley composition on the simplex

We rewrite the **prediction difference** and the **contribution** as follows:

# Shapley composition on the simplex

We rewrite the **prediction difference** and the **contribution** as follows:

- $$\begin{aligned} \boldsymbol{\delta}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}} : 2^{\mathcal{I}} &\to \mathcal{S}^N, \\ S &\mapsto \mathbb{E}^{\mathcal{A}}_{\mathrm{Pr}}[\boldsymbol{f}(\boldsymbol{x}) \mid \boldsymbol{x}_S] \ominus \mathbb{E}^{\mathcal{A}}_{\mathrm{Pr}}[\boldsymbol{f}(\boldsymbol{x})], \end{aligned} \tag{10}$$

# Shapley composition on the simplex

We rewrite the **prediction difference** and the **contribution** as follows:

- $$\begin{aligned} \boldsymbol{\delta}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}} : 2^{\mathcal{I}} &\to \mathcal{S}^N, \\ S &\mapsto \mathbb{E}^{\mathcal{A}}_{\mathrm{Pr}}[\boldsymbol{f}(\boldsymbol{x}) \mid \boldsymbol{x}_S] \ominus \mathbb{E}^{\mathcal{A}}_{\mathrm{Pr}}[\boldsymbol{f}(\boldsymbol{x})], \end{aligned} \tag{10}$$

- $$\boldsymbol{c}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(i, \boldsymbol{X}_S) = \boldsymbol{\delta}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(\boldsymbol{X}_{S\cup\{i\}}) \ominus \boldsymbol{\delta}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(\boldsymbol{X}_S), \tag{11}$$

where $\boldsymbol{a} \ominus \boldsymbol{b}$ is the perturbation $\boldsymbol{a} \oplus ((-1) \odot \boldsymbol{b})$ which correspond to a substraction between compositions $\boldsymbol{a}$ and $\boldsymbol{b}$, and where:

# Shapley composition on the simplex

The Shapley quantity expressing the contribution of the $i$th feature on a prediction can simply be expressed on the simplex as the composition $\phi(i)$ given by:

$$\phi_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(i) = \frac{1}{d!} \odot \bigoplus_{\pi} \boldsymbol{c}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(i, \pi_{\boldsymbol{X}}^{<i}). \tag{12}$$

We call this quantity **Shapley composition**.
It lives on the probability simplex.

# Shapley composition on the simplex

The Shapley quantity expressing the contribution of the $i$th feature on a prediction can simply be expressed on the simplex as the composition $\phi(i)$ given by:

$$\phi_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(i) = \frac{1}{d!} \odot \bigoplus_{\pi} \boldsymbol{c}_{\boldsymbol{f},\boldsymbol{x},\mathrm{Pr}}(i, \pi^{<i}_{\boldsymbol{X}}). \tag{12}$$

We call this quantity **Shapley composition**.
It lives on the probability simplex.

It is:

- Linear
- Efficient: $\bigoplus_{i=1}^{d} \phi_{\boldsymbol{f}}(i) = \boldsymbol{f}(\boldsymbol{x}) \ominus \mathbb{E}^{\mathcal{A}}_{\mathrm{Pr}}[\boldsymbol{f}(\boldsymbol{X})]$
- Symmetric

# Explaining a prediction with Shapley compositions

## Visualizations in the isometric-log-ratio space

We can visualize the Shapley compositions in the isometric-log-ratio space.

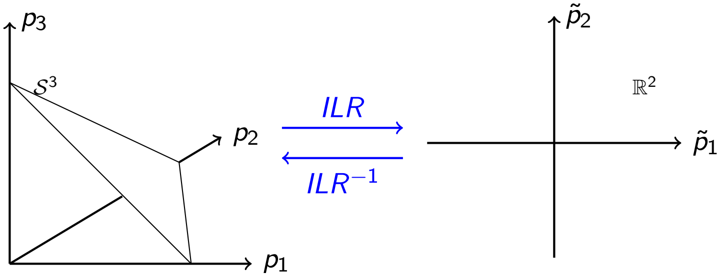



Figure: Isometric-log-ratio transformation of the 2-dimensional simplex (basis obtained with the Gram-Schmidt procedure)

where $\tilde{p}_1 = \frac{1}{\sqrt{2}} \log \frac{p_1}{p_2}$ and $\tilde{p}_2 = \frac{1}{\sqrt{6}} \log \frac{p_1 p_2}{p_3^2}$.

# Explaining a prediction with Shapley compositions

## Visualizations in the isometric-log-ratio space

And with more classes?

We can still visualize subspaces.

# Explaining a prediction with Shapley compositions

## Visualizations in the isometric-log-ratio space

And with more classes?

We can still visualize subspaces.

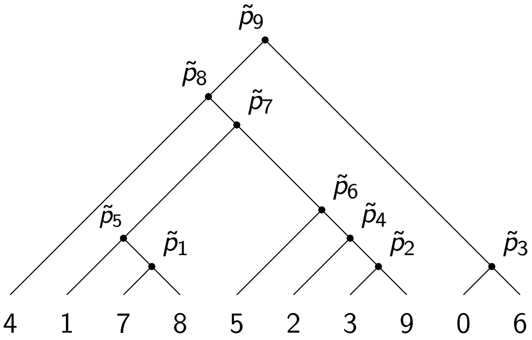



Figure: Bifurcation tree used in our 10-classes digit recognition task.

# Explaining a prediction with Shapley compositions

## Histograms

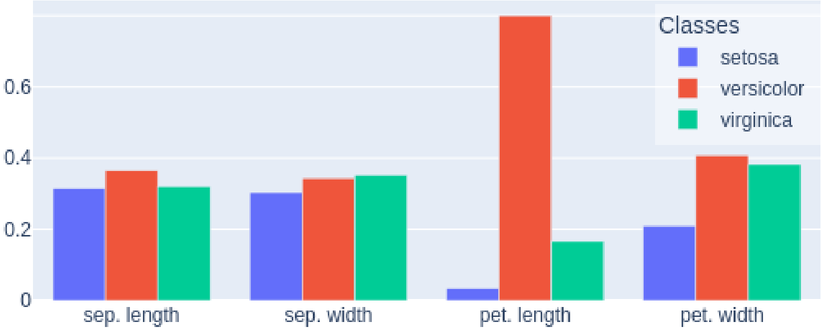



Figure: Shapley compositions visualized as histograms for the Iris classification example

# Explaining a prediction with Shapley compositions

## Histograms

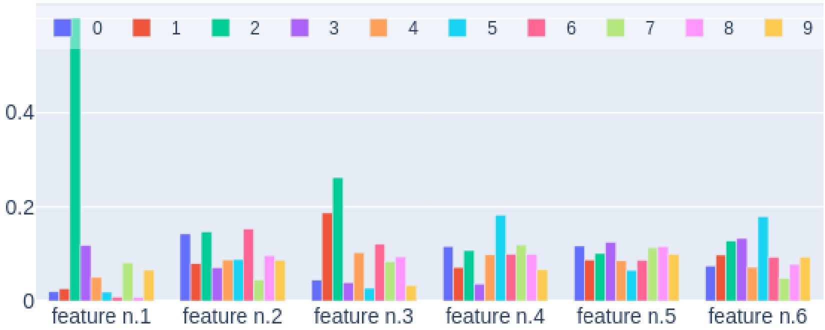



Figure: Shapley compositions visualized as histograms for the seven classes digit recognition example.

# Explaining a prediction with Shapley compositions

Use the geometrical tools!

We can summarize an explanation with:

- The norms of the Shapley compositions,
- Angles between them,
- And projection on the class-compositions.

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting
- Shapley composition,

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting
- Shapley composition,
- ~~scalar~~ **distribution/composition** living on the simplex

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting
- Shapley composition,
- ~~scalar~~ **distribution/composition** living on the simplex
- Visualisations in the ILR space, or as histograms,

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting
- Shapley composition,
- ~~scalar~~ **distribution/composition** living on the simplex
- Visualisations in the ILR space, or as histograms,
- Computation of norms, angles and projections.

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting
- Shapley composition,
- ~~scalar~~ **distribution/composition** living on the simplex
- Visualisations in the ILR space, or as histograms,
- Computation of norms, angles and projections.

Future work:

- The literature about Shapley value is prolific,

# Discussion and conclusion

Summarize:

- We proposed an extention of the concept of Shapley value for explaination in a multiclass setting
- Shapley composition,
- ~~scalar~~ **distribution/composition** living on the simplex
- Visualisations in the ILR space, or as histograms,
- Computation of norms, angles and projections.

Future work:

- The literature about Shapley value is prolific,
- Axiomatic formulation and uniqueness...

# References I


[1] John Aitchison. “Simplicial inference”. In: *Algebraic Methods in Statistics and Probability*. Ed. by Donald St. P. Richards (eds.) Ams Special Session on Algebraic Methods in Statistics Marlos A. G. Viana. Contemporary Mathematics 287. American Mathematical Society, 2001.

[2] John Aitchison. “The Statistical Analysis of Compositional Data”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 44.2 (1982), pp. 139–177.

[3] John Aitchison. “The Statistical Analysis of Compositional Data”. In: *Monographs on Statistics and Applied Probability* 25 (1986).

[4] Juan José Egozcue et al. “Isometric logratio transformations for compositional data analysis”. In: *Mathematical geology* 35.3 (2003), pp. 279–300.

# References II


[5] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774.

[6] Vera Pawlowsky-Glahn, Juan José Egozcue, and Raimon Tolosana-Delgado. *Modeling and Analysis of Compositional Data*. John Wiley & Sons, 2015.

[7] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. """ Why should i trust you?" Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1135–1144.

[8] Lloyd S Shapley et al. "A value for n-person games". In: (1953), pp. 307–317.

# References III

[9] Erik Štrumbelj and Igor Kononenko. "Explaining prediction models and individual predictions with feature contributions". In: *Knowledge and information systems* 41 (2014), pp. 647–665.

# Thank you!!