Asymmetric Shapley values to quantify the importance of genes in clinico-genomic applications

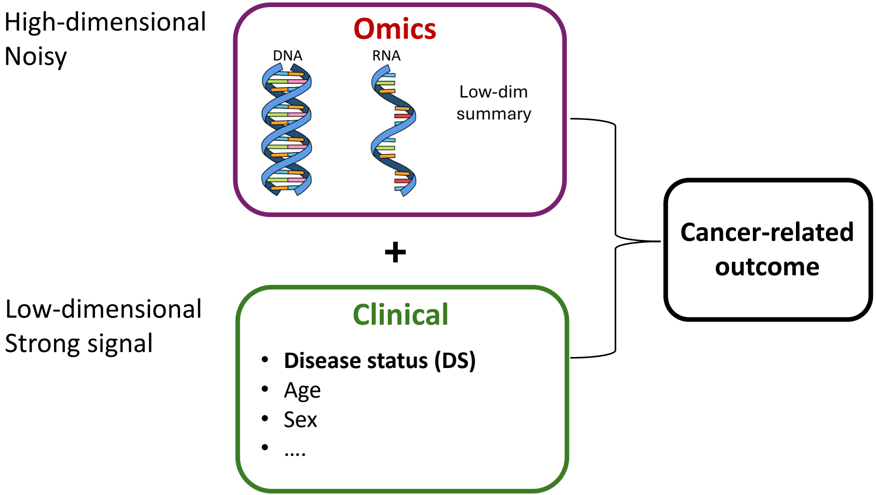
Jeroen Goedhart*, Mark van de Wiel, Martin Jullum, and Kjersti Aas

Workshop: Methods for Explainable Machine Learning in Health Care

4th Feb, 2026

Amsterdam UMC
University Medical Centers

How important are omics variables for predicting cancer?



Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).

Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*

Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype (CMS), clustering based on G_1; four categories

Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype (CMS), clustering based on G_1; four categories
- *Clinical covariates:*

Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype (CMS), clustering based on G_1; four categories
- *Clinical covariates:*
  - **DS: Tumor stage**; four categories (I, II, III, IV)

Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype (CMS), clustering based on G_1; four categories
- *Clinical covariates:*
  - **DS: Tumor stage**; four categories (I, II, III, IV)
  - age, gender, tumor site (left/right) $\rightarrow$ **confounders**

Leave covariates out to quantify the importance of the genes

- **Motivational case:** Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype (CMS), clustering based on G_1; four categories
- *Clinical covariates:*
  - **DS: Tumor stage**; four categories (I, II, III, IV)
  - age, gender, tumor site (left/right) $\rightarrow$ **confounders**

Table 1: C-index estimated on test data

Model	**C-index**
Clinical only (Cox PH)	0.72
Clinical + Omics (Ridge)	0.73

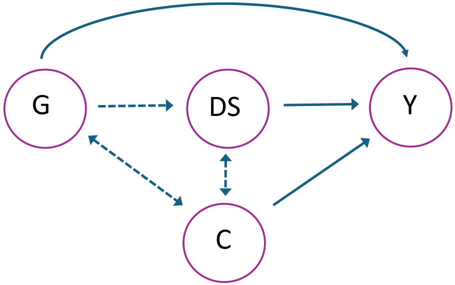
We assume that the genes (G) drive disease status (DS)

G: Gene expression, Clustering

DS: Disease state = I, II, III, IV

C: age, gender, tumor site

Y: Relapse-free survival; N = 845



Shapley values to quantify the importance of the genes

- Shapley values [1] are useful because:
  - Allows to incorporate (partial) causal knowledge through asymmetric Shapley values [2]

Shapley values to quantify the importance of the genes

- Shapley values [1] are useful because:
  - Allows to incorporate (partial) causal knowledge through asymmetric Shapley values [2]
  - Applies to any model; e.g., `blockForest` [3]

Shapley values to quantify the importance of the genes

- Shapley values [1] are useful because:
  - Allows to incorporate (partial) causal knowledge through asymmetric Shapley values [2]
  - Applies to any model; e.g., `blockForest` [3]
  - Quantifies interactions, nonlinearities, and correlations between the features

Shapley values to quantify the importance of the genes

- Shapley values [1] are useful because:
  - Allows to incorporate (partial) causal knowledge through asymmetric Shapley values [2]
  - Applies to any model; e.g., `blockForest` [3]
  - Quantifies interactions, nonlinearities, and correlations between the features
  - Both global (SAGE [4]) and local (inference)

Shapley values to quantify the importance of the genes

- Shapley values [1] are useful because:
  - Allows to incorporate (partial) causal knowledge through asymmetric Shapley values [2]
  - Applies to any model; e.g., `blockForest` [3]
  - Quantifies interactions, nonlinearities, and correlations between the features
  - Both global (SAGE [4]) and local (inference)

Table 2: Global Shapley (SAGE): average performance

	G	DS	C	**Total**
C-index	0.22	0.3	0.2	0.72

Shapley values to quantify the importance of the genes

- Shapley values [1] are useful because:
  - Allows to incorporate (partial) causal knowledge through asymmetric Shapley values [2]
  - Applies to any model; e.g., `blockForest` [3]
  - Quantifies interactions, nonlinearities, and correlations between the features
  - Both global (SAGE [4]) and local (inference)

Table 2: Global Shapley (SAGE): average performance

	G	DS	C	**Total**
C-index	0.22	0.3	0.2	0.72

Table 3: Local Shapley: average prediction

	ϕ_{G}	ϕ_{DS}	ϕ_{C}	$\widehat{\mathbf{Y}}_{\mathrm{pred}}$
Patient 1	1.1	2.3	0.3	3.7
Patient 2	-2.8	0.3	0.5	-2.0
⋮	⋮	⋮	⋮	⋮

Shapley values as a generalization of partial dependence

- **Partial dependence (PD):** effect of a feature x_j, averaged over all others

$$\mathrm{PD}(x_j) = \mathbb{E}_{X_{-j}}[\hat{f}(x_j, X_{-j})]$$

Shapley values as a generalization of partial dependence

- **Partial dependence (PD):** effect of a feature x_j, averaged over all others

$$\mathrm{PD}(x_j) = \mathbb{E}_{X_{-j}}[\hat{f}(x_j, X_{-j})]$$

- **Limitation:** PD ignores interactions and correlations

Shapley values as a generalization of partial dependence

- **Partial dependence (PD):** effect of a feature x_j, averaged over all others

$$\mathrm{PD}(x_j) = \mathbb{E}_{X_{-j}}[\hat{f}(x_j, X_{-j})]$$

- **Limitation:** PD ignores interactions and correlations

- **Shapley values:** average *PD contrast* of a feature across *all possible subsets of other features* [5]

Shapley values as a generalization of partial dependence

- **Partial dependence (PD):** effect of a feature x_j, averaged over all others

$$\mathrm{PD}(x_j) = \mathbb{E}_{X_{-j}}[\hat{f}(x_j, X_{-j})]$$

- **Limitation:** PD ignores interactions and correlations

- **Shapley values:** average *PD contrast* of a feature across *all possible subsets of other features* [5]

$$\phi_j = \sum_{\mathcal{S} \subseteq \{1,\dots,p\} \setminus \{j\}} w_{\mathcal{S}} \Big(\mathrm{PD}(\mathcal{S} \cup \{j\}) - \mathrm{PD}(\mathcal{S}) \Big)$$

$$\mathrm{PD}(\mathcal{S}) = \mathbb{E}_{X_{-\mathcal{S}} | X_{\mathcal{S}}}[\hat{f}(X_{\mathcal{S}}, X_{-\mathcal{S}})] \quad \text{(conditional PD)}$$

Shapley values as a generalization of partial dependence

- **Partial dependence (PD):** effect of a feature x_j, averaged over all others

$$\mathrm{PD}(x_j) = \mathbb{E}_{X_{-j}}[\hat{f}(x_j, X_{-j})]$$

- **Limitation:** PD ignores interactions and correlations

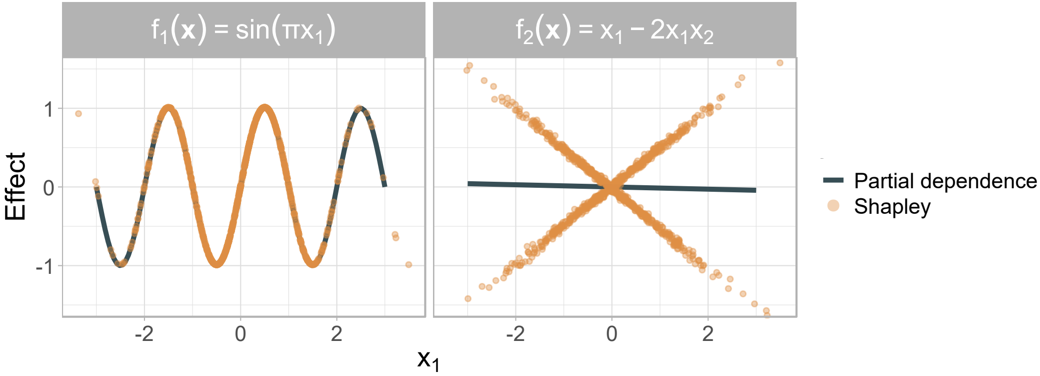
- **Shapley values:** average *PD contrast* of a feature across *all possible subsets of other features* [5]

$$\phi_j = \sum_{\mathcal{S} \subseteq \{1,\dots,p\} \setminus \{j\}} w_{\mathcal{S}} \Big(\mathrm{PD}(\mathcal{S} \cup \{j\}) - \mathrm{PD}(\mathcal{S}) \Big)$$

$$\mathrm{PD}(\mathcal{S}) = \mathbb{E}_{X_{-\mathcal{S}} | X_{\mathcal{S}}}[\hat{f}(X_{\mathcal{S}}, X_{-\mathcal{S}})] \quad \text{(conditional PD)}$$

$$\textbf{Example (genes G):} \quad \mathcal{S} \in \Big\{ \{\emptyset\}, \{\mathrm{DS}\}, \{\mathrm{C}\}, \{\mathrm{DS}, \mathrm{C}\} \Big\}$$

Shapley values find the interaction whereas PD does not



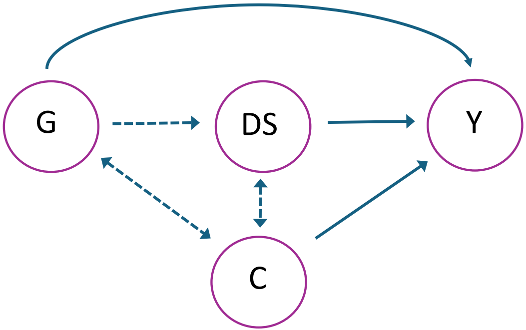
Asymmetric Shapley values ignore subsets that do not respect the (causal) ordering

For variable G with variables $\{\mathrm{G}, \mathrm{DS}, \mathrm{C}\}$:

- $\{\emptyset\}$
- $\{\mathrm{DS}\}$
- $\{\mathrm{C}\}$
- $\{\mathrm{DS}, \mathrm{C}\}$

For variable C with variables $\{\mathrm{G}, \mathrm{DS}, \mathrm{C}\}$:

- $\{\emptyset\}$
- $\{\mathrm{DS}\}$
- $\{\mathrm{G}\}$
- $\{\mathrm{G}, \mathrm{DS}\}$



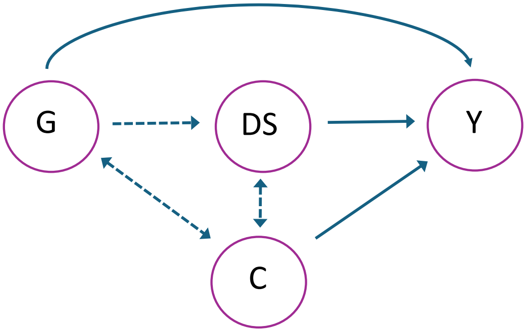
Asymmetric Shapley values ignore subsets that do not respect the (causal) ordering

For variable G with variables $\{G, DS, C\}$:

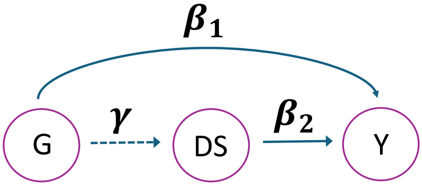
- $\{\emptyset\}$
- ~~$\{DS\}$~~
- $\{C\}$
- ~~$\{DS, C\}$~~

For variable C with variables $\{G, DS, C\}$:

- $\{\emptyset\}$
- ~~$\{DS\}$~~
- $\{G\}$
- $\{G, DS\}$

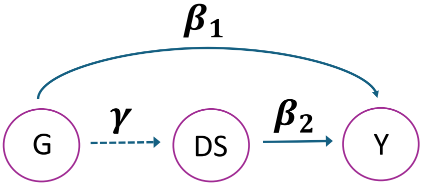


Intuition 1: Analytical expressions for a 2D toy example



$$\hat{f}\,(\mathrm{G}, \mathrm{DS}) = \beta_1 \mathrm{G} + \beta_2 \mathrm{DS}$$

Intuition 1: Analytical expressions for a 2D toy example



$$\widehat{f}\,(\mathrm{G}, \mathrm{DS}) = \beta_1 \mathrm{G} + \beta_2 \mathrm{DS}$$

Table 4: Asymmetric and symmetric Shapley values (and independent Shapley values).

Variable	Asymmetric	Symmetric	Independent
G	$\mathrm{G}\,(\beta_1 + \beta_2 \gamma)$	$\beta_1 \mathrm{G} + \frac{\gamma}{2}(\beta_2 \mathrm{G} - \beta_1 \mathrm{DS})$	$\beta_1 \mathrm{G}$

Intuition 2: Including a confounder and a nonlinearity

$$C_0,\ S_0,\ U_0,\ \sim \mathcal{N}(0,1)$$

$$\mathrm{G} \sim \mathcal{N}(0,1)$$

$$\mathrm{DS} = S_0 + \beta_1 G + U_0,$$

$$\mathrm{C}_1 = C_0 + U_0$$

$$\hat{f} = \beta_2 \mathrm{C}_1 + \beta_3 \mathrm{G} + \beta_4 \mathrm{DS}^2$$

Intuition 2: Including a confounder and a nonlinearity

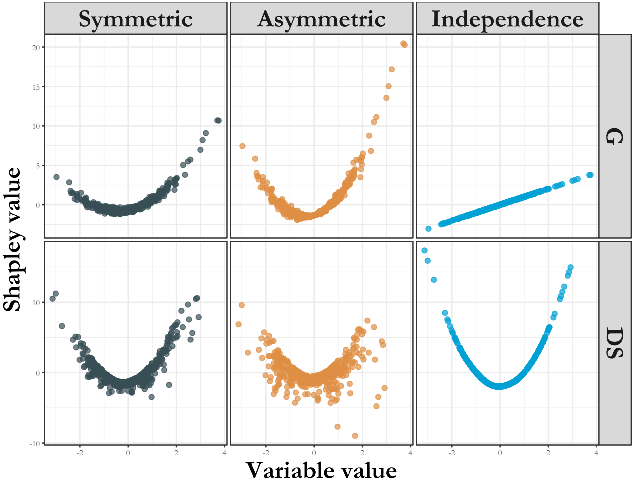
$$C_0,\ S_0,\ U_0,\ \sim \mathcal{N}(0,1)$$

$$\mathrm{G} \sim \mathcal{N}(0,1)$$

$$\mathrm{DS} = S_0 + \beta_1 G + U_0,$$

$$\mathrm{C}_1 = C_0 + U_0$$

$$\hat{f} = \beta_2 \mathrm{C}_1 + \beta_3 \mathrm{G} + \beta_4 \mathrm{DS}^2$$



Computing Shapley values for the data set described earlier

Recalling the set-up

- Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype: clustering based on G_1; four categories
- *Clinical covariates:*
  - DS: Tumor stage; four categories (I, II, III, IV)
  - age, gender, tumor site (left/right)

Computing Shapley values for the data set described earlier

Recalling the set-up

- Relapse-free survival on $N = 845$ colorectal cancer patients (253 events).
- *Genes (G):*
  - G_1: gene expression ($p = 21,292$)
  - G_2: Consensus Molecular Subtype: clustering based on G_1; four categories
- *Clinical covariates:*
  - DS: Tumor stage; four categories (I, II, III, IV)
  - age, gender, tumor site (left/right)

Experiment

- Fit a `blockForest` model using a train test split and estimate the Shapley values (both asymmetric and symmetric)
- Consider global and local feature importance

Global importance: SAGE nicely decomposes the C-index

	Symmetric	Asymmetric
intercept	0.500	0.500
Genes (G)	0.130	0.173
Profile	*0.076*	*0.082*
CMS	*0.028*	*0.043*
Disease state (DS)	0.091	0.062
Gender (C_1)	0.010	0.008
Age (C_2)	0.021	0.010
Tumor site (C_3)	0.002	0.00
Total	**0.754**	**0.754**

Table 5: SAGE decomposition of the C-index of a `blockForest` model for the symmetric and asymmetric version.

Local importance: Interplay between different variables

	ϕ_{G}	ϕ_{DS}	ϕ_{Gender}	ϕ_{Age}	ϕ_{Site}
Patient 1	1.1	2.3	0.3	0.01	0.02
Patient 2	-2.8	0.3	0.5	-0.02	-0.3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Patient N	0.1	-1.0	0.08	0.0	-0.1

Local importance: Interplay between different variables

	ϕ_{G}	ϕ_{DS}	ϕ_{Gender}	ϕ_{Age}	ϕ_{Site}
Patient 1	1.1	2.3	0.3	0.01	0.02
Patient 2	-2.8	0.3	0.5	-0.02	-0.3
⋮	⋮	⋮	⋮	⋮	⋮
Patient N	0.1	-1.0	0.08	0.0	-0.1

Conditional on the model, we can ask many (inference) questions:

- Does the importance of DS differ between left and right tumor site
- Does the importance of Gender differ across DS categories
- *et cetera* ... (p-hacking)

Local importance: Interplay between different variables

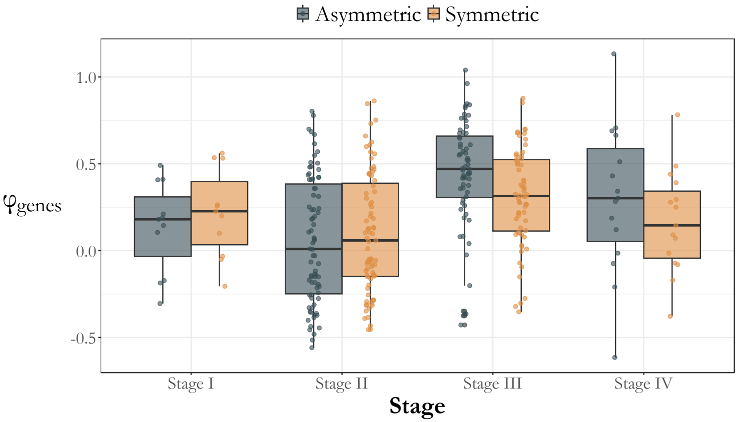
	ϕ_{G}	ϕ_{DS}	ϕ_{Gender}	ϕ_{Age}	ϕ_{Site}
Patient 1	1.1	2.3	0.3	0.01	0.02
Patient 2	-2.8	0.3	0.5	-0.02	-0.3
⋮	⋮	⋮	⋮	⋮	⋮
Patient N	0.1	-1.0	0.08	0.0	-0.1

Conditional on the model, we can ask many (inference) questions:

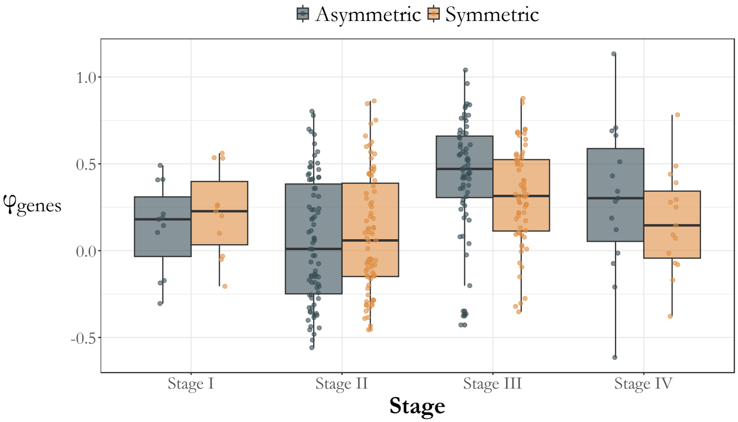
- Does the importance of DS differ between left and right tumor site
- Does the importance of Gender differ across DS categories
- *et cetera* . . . (p-hacking)

We are interested in the interplay between genes and disease status

The importance of genes differs more substantially across tumor stages in the asymmetric setting



The importance of genes differs more substantially across tumor stages in the asymmetric setting



Kruskal-Wallis test:

$$p_{\text{symmetric}} = 1.4 \times 10^{-3}$$
$$p_{\text{asymmetric}} = 4.3 \times 10^{-9}$$

A discussion and look ahead

- A (maybe not required) disclaimer: there is no general best way to quantify feature importance
- In this setting: asymmetric Shapley values are interesting as they put more weight on relevant aspects of the data generating mechanism
- Dependency modeling ($p\left(X_{-\mathcal{S}} \mid X_{\mathcal{S}}\right)$) is challenging and can be improved
- Asymmetric Shapley values are a good starting point to ask more meaningful biological and clinical questions
  - *Biological:* Grouping of genes
  - *Clinical:* For which patients omics variables are not relevant

Thank you

j.m.goedhart@amsterdamumc.nl

References

[1] Scott M. Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Curran Associates Inc., 2017, pp. 4768–4777. ISBN: 9781510860964.

[2] C. Frye, C. Rowat, and I. Feige. "Asymmetric Shapley values: incorporating causal knowledge into model-agnostic explainability". In: *arXiv* (2021). DOI: 10.48550/arXiv.1910.06358.

[3] R. Hornung and M. N. Wright. "Block Forests: random forests for blocks of clinical and omics covariate data". In: *BMC Bioinformatics* 20.1 (2019), p. 358. DOI: https://doi.org/10.1186/s12859-019-2942-y.

[4] Ian Covert, Scott M Lundberg, and Su-In Lee. "Understanding Global Feature Contributions With Additive Importance Measures". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 17212–17223.

[5] K. Aas, M. Jullum, and A. Løland. "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values". In: *Artificial Intelligence* 298 (2021), p. 103502. DOI: https://doi.org/10.1016/j.artint.2021.103502.

Technical details: Asymmetric Shapley value estimation

$$\phi_j = \sum_{\mathcal{S} \subseteq \{1,\dots,p\} \setminus \{j\}} w_{\mathcal{S}} \Big(\mathrm{PD}(\mathcal{S} \cup \{j\}) - \mathrm{PD}(\mathcal{S}) \Big)$$

$$\mathrm{PD}(\mathcal{S}) = \mathbb{E}_{X_{-\mathcal{S}} | X_{\mathcal{S}}} \left[\hat{f}(X_{\mathcal{S}}, X_{-\mathcal{S}}) \right]$$

- **Weights** $w_{\mathcal{S}}$
  - Combinatorial redefinition for omitted subsets
  - Importance sampling for large p
- **Conditional dependencies**
  - Dimension reduction for high-dimensional G
  - Dependency estimation in reduced space