Shapley Value

计算机学院
浙江大学

2024 年 7 月 12 日

Contents

Contents

Shapley Value Computation
- Efficient Sampling Approaches to Shapley Value Approximation
- Dynamic Shapley Value Computation
- Computing Shapley Values in the Plane

Application across Various Domains
- A Unified Approach to Interpreting Model Predictions (SHAP)
- Improving KernelSHAP: Practical Shapley Value Estimation via Linear Regression
- Data Shapley: Equitable Valuation of Data for Machine Learning.
- Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms.
- 2D-Shapley: A Framework for Fragmented Data Valuation.
- 2D-Shapley: A Framework for Fragmented Data Valuation
- P-Shapley: Shapley Values on Probabilistic Classifiers

Shapley Value

- Marginal Contribution: $MC(S \cup \{i\}) = U(S \cup \{i\}) - U(S)$
- Shapley Value: $SV_i = \frac{1}{n!} \sum_{\pi} U(S_i^{\pi} \cup \{i\}) - U(S_i^{\pi}) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \frac{MC(S \cup \{i\})}{\binom{n-1}{|S|}}$
- Exact Computation: $O(2^n)$

$N = \{1, 2, 3\}$

$$SV_1 = \frac{1}{3}[U(\{1\}) - U(\emptyset) + \frac{U(\{1,2\}) - U(\{2\}) + U(\{1,3\}) - U(\{3\})}{2} + U(\{1,2,3\}) - U(\{2,3\})]$$

Sampling Marginal Contributions

- Permutation Sampling

Initialize $\widehat{SV_1} = 0, \widehat{SV_2} = 0, \widehat{SV_3} = 0$
Sample 1^{th} permutation: 1,2,3
$\widehat{SV_1} += U(\{1\}) - U(\emptyset), \widehat{SV_2} += U(\{1,2\}) - U(\{1\}), \widehat{SV_3} += U(\{1,2,3\}) - U(\{1,2\})$
Sample 2^{nd} permutation: 3,2,1
$\widehat{SV_3} += U(\{3\}) - U(\emptyset), \widehat{SV_2} += U(\{3,2\}) - U(\{3\}), \widehat{SV_1} += U(\{3,2,1\}) - U(\{2,1\})$
Average: $\widehat{SV_1} /= 2, \widehat{SV_2} /= 2, \widehat{SV_3} /= 2$

Sampling Marginal Contributions

- Stratified Sampling

Stratum 1	Stratum 2	Stratum 3
$U(\{1\}) - U(\emptyset)$	$U(\{1,2\}) - U(\{2\}), U(\{1,3\}) - U(\{3\})$	$U(\{1,2,3\}) - U(\{2,3\})$

表: $SV_1 = \frac{1}{3}(SV_{1,1} + SV_{1,2} + SV_{1,3})$

Initialize $\widehat{SV_{1,1}} = 0, \widehat{SV_{1,2}} = 0, \widehat{SV_{1,3}} = 0$
Take one sample in Stratum 1: $U(\{1\}) - U(\emptyset)$
$\widehat{SV_{1,1}} += U(\{1\}) - U(\emptyset)$
Take one sample in Stratum 2: $U(\{1,3\}) - U(\{3\})$
$\widehat{SV_{1,2}} += U(\{1,3\}) - U(\{3\})$
Take one sample in Stratum 3: $U(\{1,2,3\}) - U(\{1,2\})$
$\widehat{SV_{1,3}} += U(\{1,2,3\}) - U(\{1,2\})$
Average: $\widehat{SV_1} = \frac{\widehat{SV_{1,1}} + \widehat{SV_{1,2}} + \widehat{SV_{1,3}}}{3}$

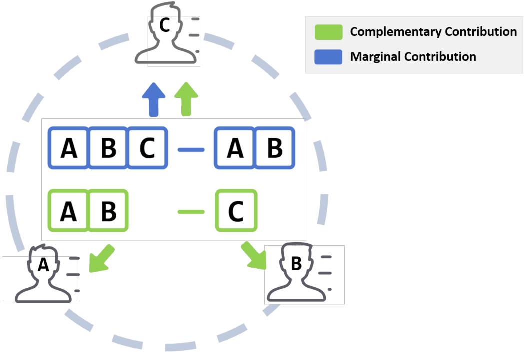
Complementary Contribution

- Complementary Contribution: $CC(S) = U(S) - U(N \setminus S)$
- Shapley Value: $SV_i = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \frac{CC(S \cup \{i\})}{\binom{n-1}{|S|}}$

$N = \{1, 2, 3\}$

$$SV_1 = \frac{1}{3}[U(\{1\}) - U(\{2,3\}) + \frac{U(\{1,2\}) - U(\{3\}) + U(\{1,3\}) - U(\{2\})}{2} + U(\{1,2,3\}) - U(\emptyset)]$$

Complementary Contribution



$$A: U(\{A,B\}) - U(\{C\})$$
$$B: U(\{A,B\}) - U(\{C\})$$
$$C: U(\{C\}) - U(\{A,B\})$$

浙江大學 ZHEJIANG UNIVERSITY

Stratification

Stratum 1	Stratum 2	Stratum 3
$U(\{1\}) - U(\{2,3\})$	$U(\{1,2\}) - U(\{3\})$	$U(\{1,2,3\}) - U(\emptyset)$
	$U(\{1,3\}) - U(\{2\})$	

表: $SV_1 = SV_{1,1} + SV_{1,2} + SV_{1,3}$

Stratum 1	Stratum 2	Stratum 3
$U(\{2\}) - U(\{1,3\})$	$U(\{1,2\}) - U(\{3\})$	$U(\{1,2,3\}) - U(\emptyset)$
	$U(\{2,3\}) - U(\{1\})$	

表: $SV_2 = SV_{2,1} + SV_{2,2} + SV_{2,3}$

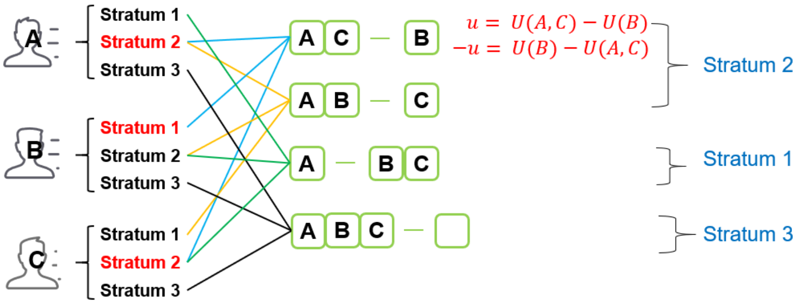
Stratum 1	Stratum 2	Stratum 3
$U(\{3\}) - U(\{1,2\})$	$U(\{1,3\}) - U(\{2\})$	$U(\{1,2,3\}) - U(\emptyset)$
	$U(\{2,3\}) - U(\{1\})$	

表: $SV_3 = SV_{3,1} + SV_{3,2} + SV_{3,3}$

Stratification

Stratum 1	Stratum 2	Stratum 3
$U(\{1\}) - U(\{2,3\})$	$U(\{1,2\}) - U(\{3\})$	$U(\{1,2,3\}) - U(\emptyset)$
$U(\{2\}) - U(\{1,3\})$	$U(\{1,3\}) - U(\{2\})$	
$U(\{3\}) - U(\{1,2\})$	$U(\{2,3\}) - U(\{1\})$	

Example



$$SV_A = \frac{1}{3}[U(\{A\}) - U(\{B,C\}) + \frac{1}{2}(U(\{A,B\}) - U(\{C\})) + \frac{1}{2}(U(\{A,C\}) - U(\{B\})) + U(\{A,B,C\}) - U(\emptyset)]$$

$$SV_B = \frac{1}{3}[U(\{B\}) - U(\{A,C\}) + \frac{1}{2}(U(\{A,B\}) - U(\{C\})) + \frac{1}{2}(U(\{B,C\}) - U(\{A\})) + U(\{A,B,C\}) - U(\emptyset)]$$

$$SV_C = \frac{1}{3}[U(\{C\}) - U(\{A,B\}) + \frac{1}{2}(U(\{A,C\}) - U(\{B\})) + \frac{1}{2}(U(\{B,C\}) - U(\{A\})) + U(\{A,B,C\}) - U(\emptyset)]$$

Sample Allocation

Allocate sample number to minimize the sum of variances of estimated $\widehat{SV}_i$
Denote by $m_{i,j}$ the number of samples for estimating $SV_{i,j}$ and by $\sigma^2_{i,j}$ the variance of $CC_{i,j}$.

$$\textit{minimize} \sum_{i=1}^{n} Var[\widehat{SV}_i] = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\sigma^2_{i,j}}{m_{i,j}}$$

$$\textit{subject to} \sum_{j=1}^{n} m_{i,j} = m$$

- $\sigma^2_{i,j}$? sampling estimation
- $m_{i,j}$? $\mathbb{E}[m_{i,j}] = m_j \cdot \frac{j}{n}$

Stratum 1	Stratum 2	Stratum 3
$U(\{1\}) - U(\{2,3\})$	$U(\{1,2\}) - U(\{3\})$	$U(\{1,2,3\}) - U(\emptyset)$
$U(\{2\}) - U(\{1,3\})$	$U(\{1,3\}) - U(\{2\})$	
$U(\{3\}) - U(\{1,2\})$	$U(\{2,3\}) - U(\{1\})$	

Sample Allocation

Adaptively allocate more samples to the estimator $\widehat{SV}_{i,j}$ that is relatively farther from its expected value

Use Empirical Bound Inequality to estimate error bound $\epsilon_{i,j}$ of $\widehat{SV}_{i,j}$ in an online manner

Given the previous sampling results, the next sample is selected as

$$\arg \max_{C \subseteq N(S)} \left\{ \sum_{z_i \in S} \epsilon_{i,|S|} + \sum_{z_i \in N \setminus S} \epsilon_{i,|N \setminus S|} \right\} \quad (1)$$

Dynamic Shapley Value

- Addition: A set of data points $D_{add} = \{z_{n+1}, ..., z_m\}$ are added to D to form a new dataset $N^+ = D \cup D_{add}$. The Shapley value of $z_i \in N^+$ is

$$SV_i^+ = \frac{1}{n+m} \sum_{S \subseteq N^+ \setminus \{i\}} \frac{U(S \cup \{i\}) - U(S)}{\binom{n-1}{|S|}}$$

- Deletion: A subset of data points $D_{del} = \{z_p, ..., z_q\} \subseteq D$ are removed from D to form a new dataset $N^- = D \setminus D_{del}$. The Shapley value of $z_i \in N^-$ is

$$SV_i^- = \frac{1}{n+p-q-1} \sum_{S \subseteq N^- \setminus \{i\}} \frac{U(S \cup \{i\}) - U(S)}{\binom{n+p-q-2}{|S|}}$$

The problem of dynamic Shapley value computation is to compute SV_i^+ / SV_i^- for all the data points in N^+ / N^- efficiently

Motivation

- A dynamic patient dataset

id	age	sex	cp	rbps	chol	fbs	disease
z_1	61	Male	Level 4	138	204	<120	Severe
z_2	46	Female	Level 2	105	166	<120	None
z_3	59	Male	Level 3	150	212	>120	None

Add

- Marginal contributions

permutation	z_1	z_2
z_1 z_2	5	7
z_2 z_1	6	6
SV	11/2	13/2

$U(\{z_1\}) - \emptyset$

$U(\{z_2\}) - \emptyset$

$U(\{z_1, z_2\}) - U(\{z_1\})$

$U(\{z_1, z_2\}) - U(\{z_2\})$

permutation	z_1	z_2	z_3
z_1 z_2 z_3	5	7	8
z_1 z_3 z_2	5	10	5
z_2 z_1 z_3	6	6	8
z_2 z_3 z_1	7	6	7
z_3 z_1 z_2	7	10	3
z_3 z_2 z_1	7	10	3
SV	37/6	49/6	17/3

Algorithms

- For Addition: The Pivot Algorithm, The Delta Algorithm
- For Deletion: The YN Algorithm, The Delta Algorithm
- The Heuristic Algorithm

Pivot: Motivation

The marginal contribution of z_i in permutations of $D \cup \{z_i\}$ where z_i appears before z_{n+1} has already been precomputed in D and can be reused.

- $D = \{z_1, z_2\}, N^+ = \{z_1, z_2, z_3\}$
- Choose z_3 as a pivot
- For z_1 ,$[z_1, z_2, z_3], [z_2, z_1, z_3], [z_1, z_3, z_2]$;
  $[z_3, z_2, z_1], [z_2, z_3, z_1], [z_3, z_1, z_2]$

permutation	z_1	z_2
z_1 z_2	5	7
z_2 z_1	6	6
SV	11/2	13/2

permutation	z_1	z_2	z_3
z_1 z_2 z_3	5	7	8
z_1 z_3 z_2	5	10	5
z_2 z_1 z_3	6	6	8
z_2 z_3 z_1	7	6	7
z_3 z_1 z_2	7	10	3
z_3 z_2 z_1	7	10	3
SV	37/6	49/6	17/3

$$SV_1^+ = LSV_1^+ + RSV_1^+$$

LSV_1^+ Precomputed

RSV_1^+ To be computed

Pivot: Definition

- Given $D = \{z_1, \ldots, z_n\}, N^+ = D \cup \{z_{n+1}\}$
- Choose z_{n+1} as a pivot
  - G_L^i consists of permutations where z_i is located in front of z_{n+1}
    - Marginal contributions of z_i in G_L^i:

$$LSV_i^+ = \frac{1}{(n+1)!} \sum_{\pi^k \in G_L^i} [U(\{z_{\pi^k(1)}, \ldots, z_{\pi^k(j)}\}) - U(\{z_{\pi^k(1)}, \ldots, z_{\pi^k(j-1)}\})]$$

  - G_R^i consists of permutations where z_i is located behind z_{n+1}
    - Marginal contributions of z_i in G_R^i:

$$RSV_i^+ = \frac{1}{(n+1)!} \sum_{\pi^k \in G_R^i} [U(\{z_{\pi^k(1)}, \ldots, z_{\pi^k(j)}\}) - U(\{z_{\pi^k(1)}, \ldots, z_{\pi^k(j-1)}\})]$$

- $SV_i^+ = LSV_i^+ + RSV_i^+$

Pivot: Algorithm

- LSV_i^+ is precomputed on D
  - When applying Monte Carlo algorithm to compute SV on D, store marginal contributions in front of a uniformly sampled position as LSV_i^+
- RSV_i^+ needs to be computed on N^+

① Use sampled permutations or uniformly sample a permutation over $[n+1]$

② Compute marginal contributions behind the new data point in permutation as RSV_i^+

③ Repeat k times and output $SV_i^+ = LSV_i^+ + RSV_i^+$

Pivot: Example

①Precompute LSV_i^+ with 2 sampled permutations

permutation	LSV_1^+	LSV_2^+
$z_1 \backslash z_2$	5	0
$z_2 z_1 \backslash$	6	6
LSV_i^+	11/2	3

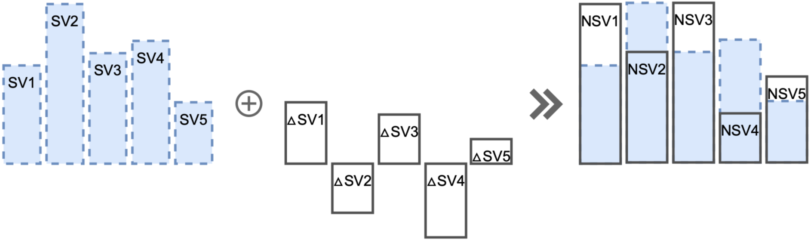
②Compute RSV_i^+ with 2 sampled permutations

permutation	RSV_1^+	RSV_2^+	RSV_3^+
$z_1 z_3 z_2$	0	10	5
$z_3 z_2 z_1$	7	10	3
RSV_i^+	7/2	10	4

③$SV_1^+ = LSV_1^+ + RSV_1^+ = 9, SV_2^+ = LSV_2^+ + RSV_2^+ = 13, SV_3^+ = RSV_3^+ = 4$

Delta: Motivation

Compute ΔSV_i rather than NSV_i ,requiring fewer sampled permutations for the same accuracy



Delta: Algorithm

- Given $D = \{z_1, \dots, z_n\}, N^+ = \{z_1, \dots, z_n, z_{n+1}\}$,for $z_i \in D$

$$\Delta MC_i = [U(S \cup \{z_{n+1}\} \cup \{z_i\}) - U(S \cup \{z_i\})] - [U(S \cup \{z_{n+1}\}) - U(S)]$$

$$\Delta SV_i = \frac{1}{n+1} \sum_{S \subseteq D \setminus z_i} \frac{|S| + 1}{n \binom{n-1}{|S|}} \Delta MC_i$$

$$SV_i^+ = SV_i + \Delta SV_i$$

① Uniformly sample a permutation over [n]

② Scan permutation and compute ΔMC_i of each data point for ΔSV_i

③ Compute marginal contributions of z_{n+1} for SV_{n+1}

④ Repeat k times and update Shapley value by $SV_i^+ = SV_i + \Delta SV_i$

YN: Motivation

Store precomputed utilities in multiple-dimensional arrays and reuse

- $D = \{z_1, z_2, z_3\}, N^- = \{z_1, z_2\}$

permutation	z_1	z_2	z_3
$z_1\ z_2\ z_3$	5	7	8
$z_1\ z_3\ z_2$	5	10	5
$z_2\ z_1\ z_3$	6	6	8
$z_2\ z_3\ z_1$	7	6	7
$z_3\ z_1\ z_2$	7	10	3
$z_3\ z_2\ z_1$	7	10	3
SV	37/6	49/6	17/3

permutation	z_1	z_2
$z_1\ z_2$	5	7
$z_2\ z_1$	6	6
SV	11/2	13/2

YN: Algorithm

- Given $D = \{z_1, \ldots, z_n\}$

$$YN[z_i][z_j][k] = \sum_{S \subseteq D, |S|=k, z_i \in S, z_j \notin S} U(S)$$

$$NN[z_i][z_j][k] = \sum_{S \subseteq D, |S|=k, z_i \notin S, z_j \notin S} U(S)$$

- Given $N^- = D \setminus z_{del}$, for $z_i \in N^-$

$$SV_i^- = \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{(YN[z_i][z_{del}][k] - NN[z_i][z_{del}][k-1])}{\binom{n-2}{k-1}}$$

YN: Example

- $D = \{z_1, z_2, z_3\}$

$YN[z_1][z_3][0] = U(\emptyset), NN[z_1][z_3][0] = U(\emptyset)$

$YN[z_1][z_3][1] = U(\{z_1\}), NN[z_1][z_3][1] = U(\{z_2\})$

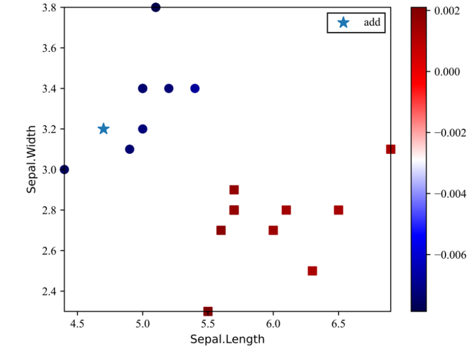
$YN[z_1][z_3][2] = U(\{z_1, z_2\}), NN[z_1][z_3][2] = U(\emptyset)$

- $N^- = \{z_1, z_2\}$

$$SV_1^- = \frac{1}{2}\sum_{k=1}^{2}\frac{YN[z_1][z_3][k] - NN[z_1][z_3][k-1]}{\binom{1}{k-1}} = \frac{1}{2}U(\{z_1, z_2\}) - \frac{1}{2}U(\{z_2\}) + \frac{1}{2}U(\{z_1\}) - \frac{1}{2}U(\emptyset)$$

Heuristi

- Data points with similar features tend to have similar performance on models, resulting in similar utility functions and Shapley values
- Find adjacent data points of added data points using k-NN, and assign the averages over their Shapley values to new data points



Changes in Shapley Values with the Addition of a New Data Point

Sergio Cabello, Timothy M. Chan: Computing Shapley Values in the Plane. SoCG 2019: 20:1-20:19

Convex Hull

Convex Hull

The convex hull of a set of points S in $\mathbb{R}^n$ is the set of all convex combinations of points in S.

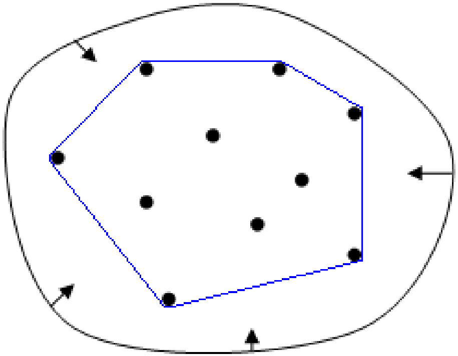
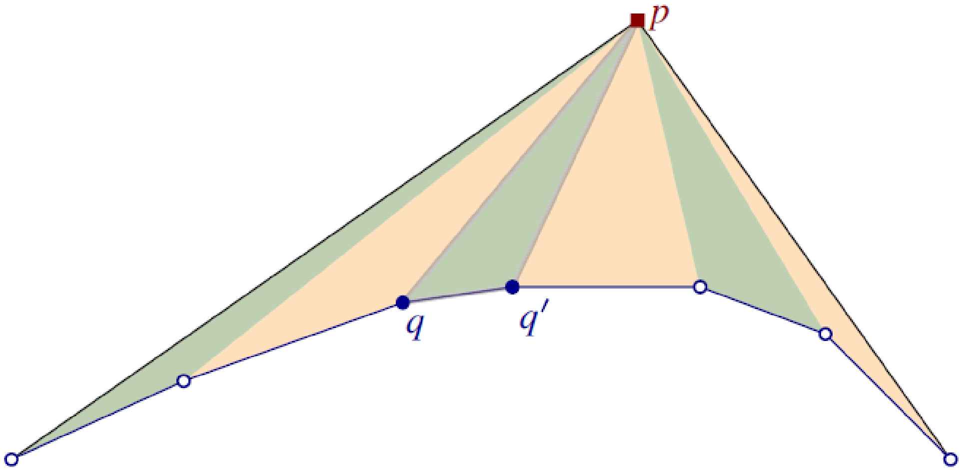


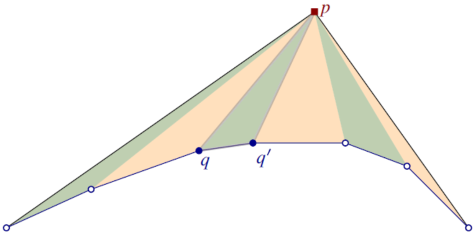
图: Convex Hull

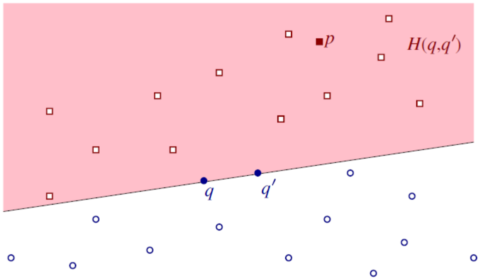
Utility Definition: Area

Given a fixed set P of points in the plane, consider the area of the convex hull formed by these points as a utility function. Compute the Shapley value for each point.

Marginal Contributions







In order for Δpqq to be a part of the marginal contribution of p in the permutation, q and q' must appear before p, □ must appear after p, and ○ can be positioned anywhere.

□: α, ○: $n-\alpha-3$

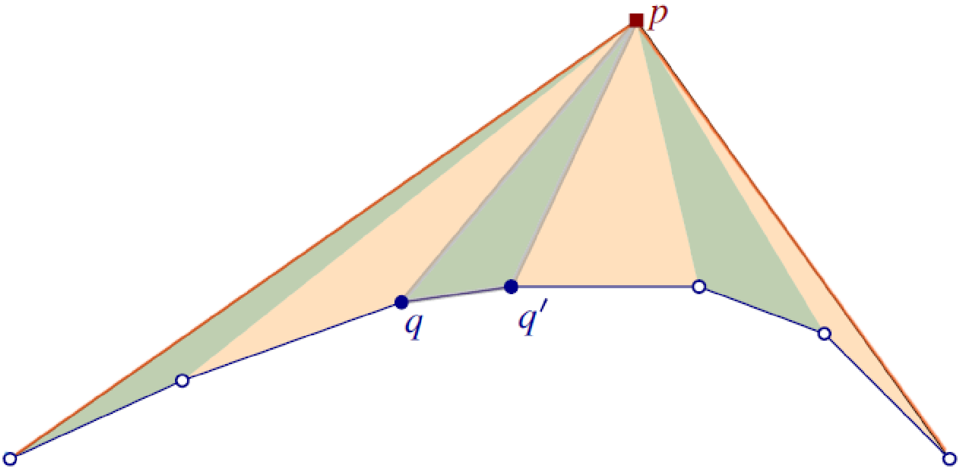
How many such permutations are there? $2!\alpha!((\alpha+4)\cdots n) = \frac{2n!}{(\alpha+1)(\alpha+2)(\alpha+3)}$, denoted by $\rho(q,q')$

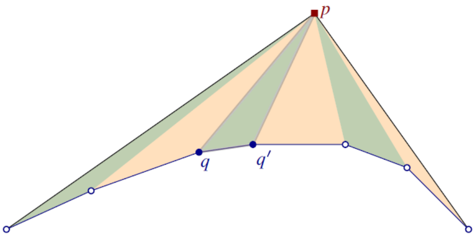
$$SV_p = \frac{1}{n!} \sum_{q,q' \in P(q \neq q') \text{ with } p \in H(q,q')} area(\Delta pqq') \cdot \rho(q,q')$$

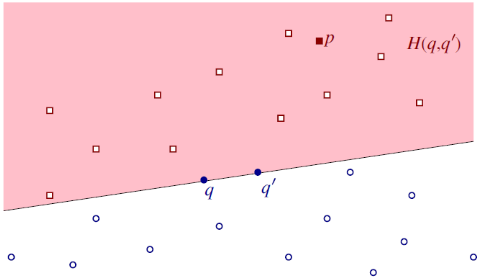
Utility Definition: Perimeter

Given a fixed set P of points in the plane, consider the perimeter of the convex hull formed by these points as a utility function. Compute the Shapley value for each point.

Marginal Contributions







Let $E(\pi, p)$ be the set of directed edges of $CH(P(\pi, p))$ that are not in $CH(P(\pi, p) \setminus \{p\})$, where edges are oriented in clockwise order.
Then $(q, p) \in E(\pi, p)$ if and only if q appears before p in π and no points before p in π lie in $H(q, p)$.
$\square$: α
How many such permutations are there? $\alpha!((\alpha+3)\cdots n) = \frac{n!}{(\alpha+1)(\alpha+2)}$, denoted by $\rho'(q, p)$

Let $E(\pi, p)$ be the set of directed edges of $CH(P(\pi, p))$ that are not in $CH(P(\pi, p) \setminus \{p\})$, where edges are oriented in clockwise order.
Then $(q, p) \in E(\pi, p)$ if and only if q appears before p in π and no points before p in π lie in $H(q, p)$.
□: α
How many such permutations are there? $\alpha!((\alpha+2)\cdots n) = \frac{n!}{(\alpha+1)(\alpha+2)}$, denoted by $\rho'(q, p)$

$$SV_p^+ = \frac{1}{n!} \sum_{q \in P \setminus \{p\}} \|p - q\| \cdot (\rho'(p, q) + \rho'(q, p))$$

$$SV_p^- = \frac{1}{n!} \sum_{q, q' \in P (q \neq q') \text{ with } p \in H(q, q')} \|q - q'\| \cdot \rho(q, q')$$

$$SV_p = SV_p^+ - SV_p^-$$

Shapley value

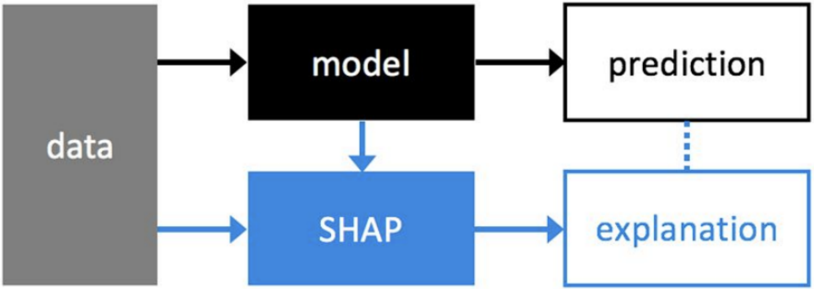
In a cooperative game, the Shapley value $\mathcal{SV}_i$ of a player z_i represents the payoff they deserve based on the following principles:

1. **Efficiency:** The total payoff received by all participants equals the total payoff generated by the entire cooperation.
2. **Symmetry:** If two participants contribute equally to any coalition, then these two participants should receive the same payoff.
3. **Dummy Player:** If a participant does not contribute to any possible coalition, then this participant should not receive any payoff.
4. **Additivity:** If you merge two cooperative games into one, then the payoff for any participant in the merged game equals the sum of their payoffs in the original two games.

$$\phi_i = \frac{1}{n} \sum_{S \subseteq N \setminus \{z_i\}} \frac{U(S \cup \{z_i\}) - U(S)}{\binom{n-1}{|S|}}.$$

SHAP:SHapley Additive exPlanations

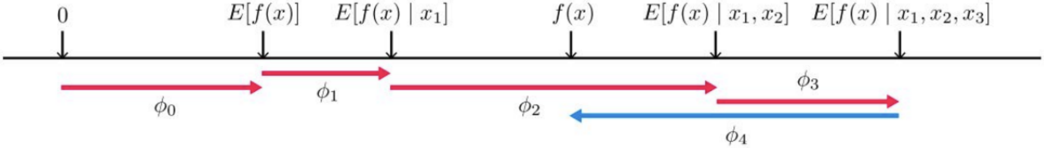
Additive feature attribution method to explain the output of any ML model (or any black-box models).



It assigns each feature a value to measure its importance for a particular prediction.

SHAP:SHapley Additive exPlanations

Additive feature attribution method:



Property 1 (Local accuracy) $f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i$

Property 2 (Missingness) $x'_i = 0 \Rightarrow \phi_i = 0$

Property 3 (Consistency) $f'_x(z') - f'_x(z' \setminus i) \geq f_x(z') - f_x(z' \setminus i) \Rightarrow \phi_i(f', x) \geq \phi_i(f, x)$

Only **Shapley values** satisfy all the properties.

SHAP:SHapley Additive exPlanations

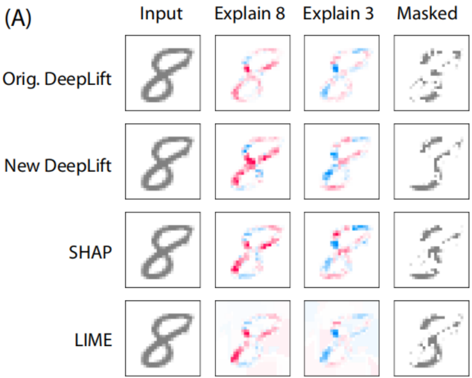


图: Red areas increase the probability of that class, and blue areas decrease the probability. Masked removes pixels in order to go from 8 to 3.

SHAP

Scott M. Lundberg, Su-In Lee: A Unified Approach to Interpreting Model Predictions. NIPS 2017: 4765-4774

KernelSHAP

Shapley values are the solutions of the weighted least-squares problem:

$$\min_{\phi_1,\dots,\phi_n} \sum_{S \subseteq N} \mu(S) || \sum_{i \in S} \phi_i - U(S) + U(\varnothing) ||^2$$

The Shapley kernel is given by

$$\mu(S) = \frac{n-1}{\binom{n}{S}} |S|(n - |S|)$$

KernelSHAP

Solving the problem requires evaluating the utilities of 2^n coalitions. KernelSHAP manages this challenge by subsampling a dataset and optimizing an approximate objective.

$$\min_{\phi_1,\ldots,\phi_n} \frac{1}{m}\sum_{i=1}^{m}(s_i^T\phi - U(s_i) + U(\mathbf{0}))$$

$$\text{s.t.} \quad \mathbf{1}^T\phi = U(\mathbf{1}) - \mathbf{0}.$$

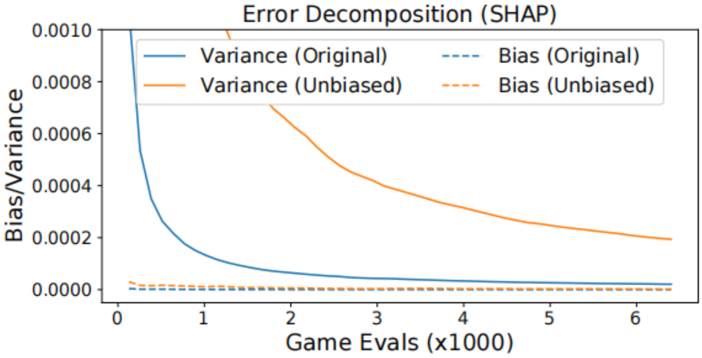
Given a set of samples $\{s_1, \ldots, s_m\}$, the solution is

$$\hat{\phi} = \hat{A}^{-1}(\hat{b} - \mathbf{1}\frac{\mathbf{1}\hat{A}^{-1}\hat{b} - U(\mathbf{1}) + U(\mathbf{0})}{\mathbf{1}^T\hat{A}\mathbf{1}}),$$

where $\hat{A} = \frac{1}{m}\sum_{i=1}^{m} s_i s_i^T$ and $\hat{b} = \frac{1}{m}\sum_{i=1}^{m} s_i(U(s_i) - U(\mathbf{0}))$.

KernelSHAP: an unbiased estimator

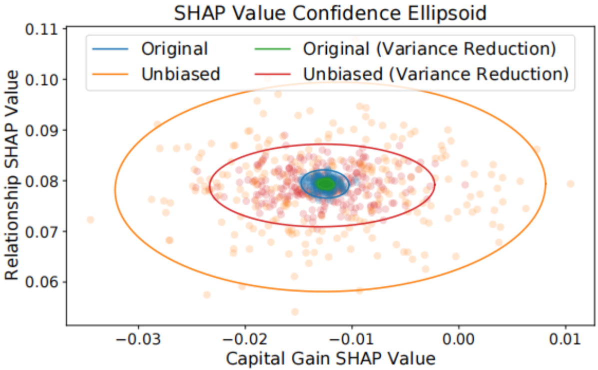
$$A = \mathbb{E}[ss^T] \quad b = \mathbb{E}[s(U(s) - U(\mathbf{0}))] \Rightarrow \bar{b_n} = \frac{1}{m}\sum_{i=1}^{m} s_i U(s_i) - \mathbb{E}[s]\, U(\mathbf{0}).$$



Improving KernelSHAP: paired Sampling

$$\check{b} = \frac{1}{2*m}\sum_{i=1}^{m} s_i U(s_i) + (\mathbf{1} - s_i)U(\mathbf{1} - s_i) - U(\mathbf{0})$$

.



KernelSHAP

Ian Covert, Su-In Lee: Improving KernelSHAP: Practical Shapley Value Estimation Using Linear Regression. AISTATS 2021: 3457-3465

Data valuation

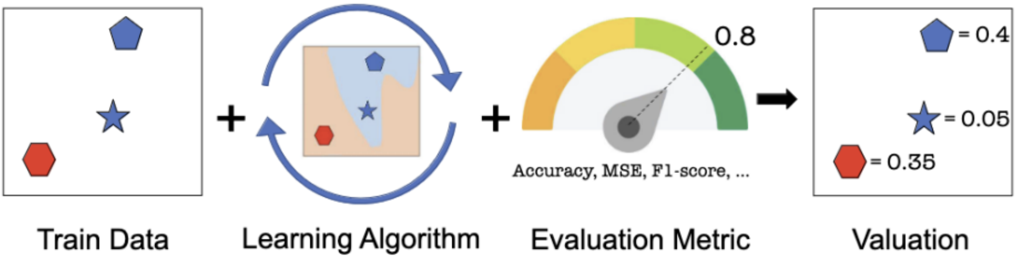


图: Data valuation.

Data pricing

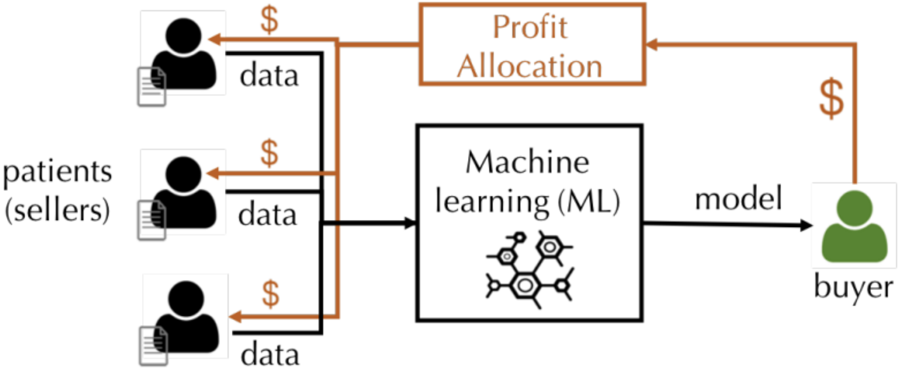
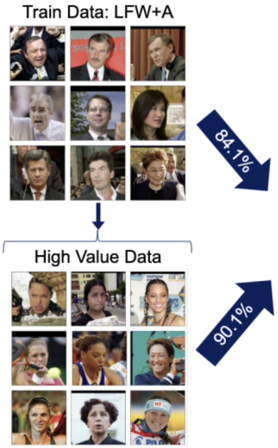


图: Data pricing.

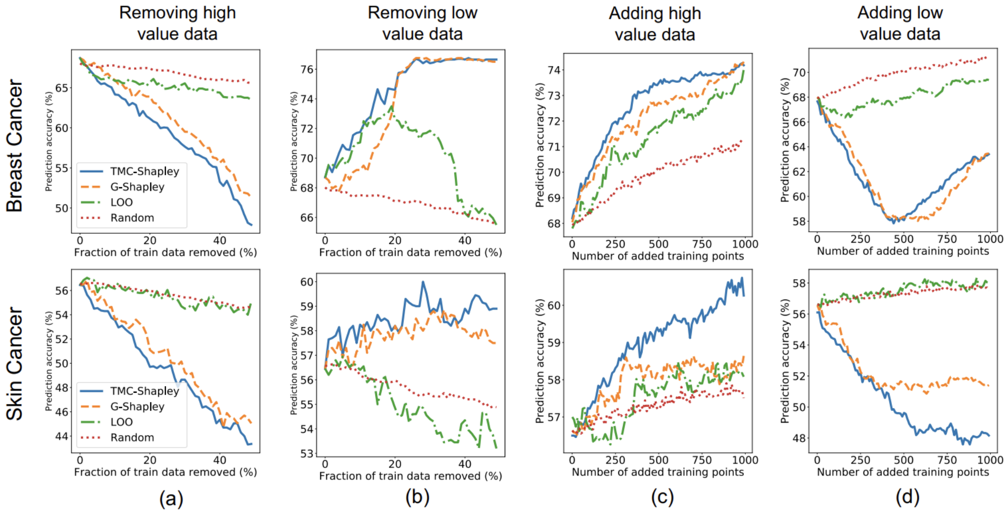
Data selection & Truncated Monte Carlo Shapley



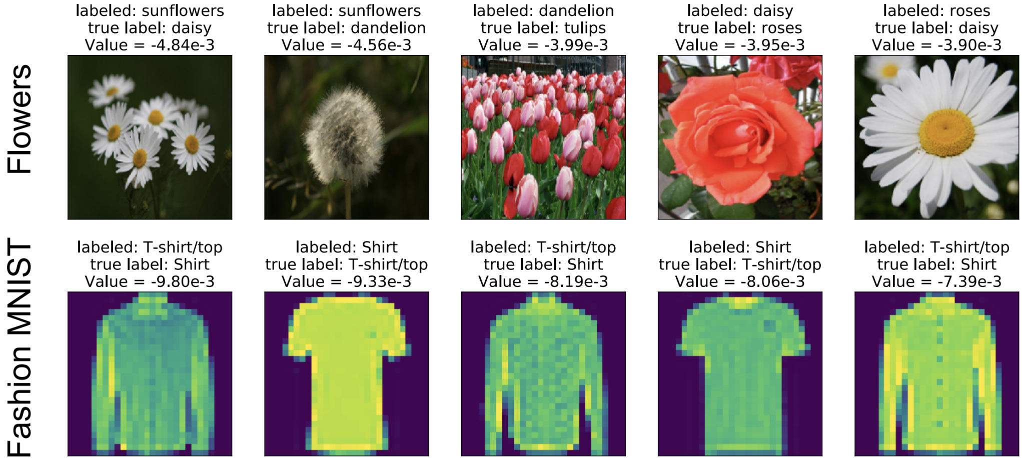
Algorithm 1 Truncated Monte Carlo Shapley

Input: Train data $D = \{1, \dots, n\}$, learning algorithm $\mathcal{A}$, performance score V
Output: Shapley value of training points: $\phi_1, \dots, \phi_n$
Initialize $\phi_i = 0$ for $i = 1, \dots, n$ and $t = 0$
while Convergence criteria not met **do**
  $t \leftarrow t + 1$
  π^t: Random permutation of train data points
  $v_0^t \leftarrow V(\emptyset, \mathcal{A})$
  for $j \in \{1, \dots, n\}$ **do**
    if $|V(D) - v_{j-1}^t| <$ Performance Tolerance **then**
      $v_j^t = v_{j-1}^t$
    else
      $v_j^t \leftarrow V(\{\pi^t[1], \dots, \pi^t[j]\}, \mathcal{A})$
    end if
    $\phi_{\pi^t[j]} \leftarrow \frac{t-1}{t}\phi_{\pi^{t-1}[j]} + \frac{1}{t}(v_j^t - v_{j-1}^t)$
  end for
end for

DataShapley:Experiments



DataShapley:Experiments



DataShapley

Amirata Ghorbani, James Y. Zou: Data Shapley: Equitable Valuation of Data for Machine Learning. ICML 2019: 2242-2251

KNNShapley: recursive calculation

Utility function:

$$U(S) = \frac{1}{K} \sum_{k=1}^{\min\{K,|S|\}} \mathbf{1}[y_{\alpha_k}(S) = y_{test}] \tag{2}$$

The Shapley value of each training point can be calculated recursively as follows:

$$\phi_{\alpha_N} = \frac{1[y_{\alpha_N} = y_{test}]}{N} \tag{3}$$

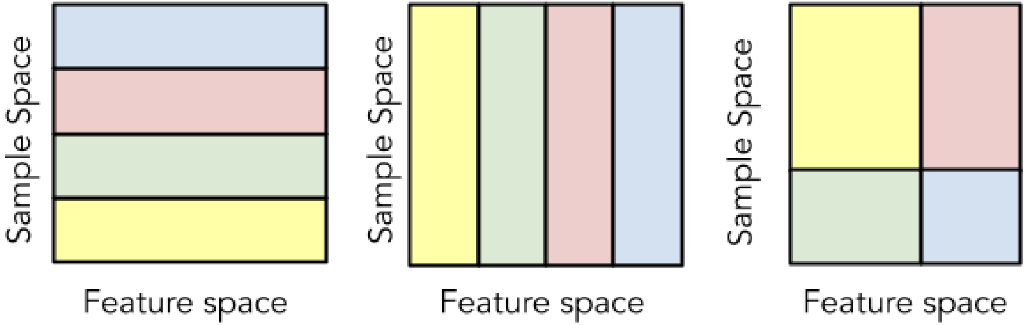
$$\phi_{\alpha_i} = \phi_{\alpha_{i+1}} + \frac{1[y_{\alpha_i} = y_{test}] - 1[y_{\alpha_{i+1}} = y_{test}]}{K} \frac{\min\{K, i\}}{i} \tag{4}$$

The complexity of the algorithm is $O(n \log n)$.

KNNShapley

Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gürel, Bo Li, Ce Zhang, Costas J. Spanos, Dawn Song: Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms. Proc. VLDB Endow. 12(11): 1610-1623 (2019)

2D-Shapley

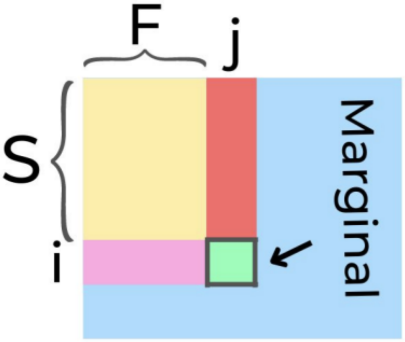

Sample Space
Feature space
Sample Space
Feature space
Sample Space
Feature space

2D-Shapley

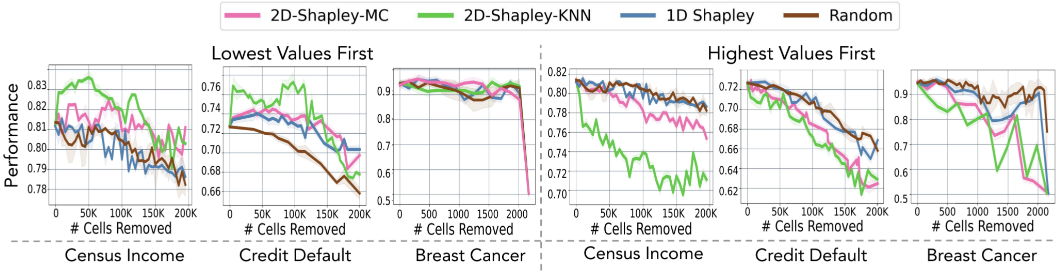
$$MC^{i,j}(S,F) := U(S\cup i, F\cup j) + U(S,F) - U(S\cup i, F) - U(S, F\cup j)$$

$$\psi_{ij}^{2d} = \frac{1}{nm}\sum_{s=1}^{n}\sum_{f=1}^{m}\Delta_{sf}$$

$$\Delta_{sf} = \frac{1}{\binom{n-1}{s-1}\binom{m-1}{f-1}}\sum_{(S,F)\in D_{sf}^{ij}} MC^{i,j}(S,F)$$



2D-Shapley: Experiments

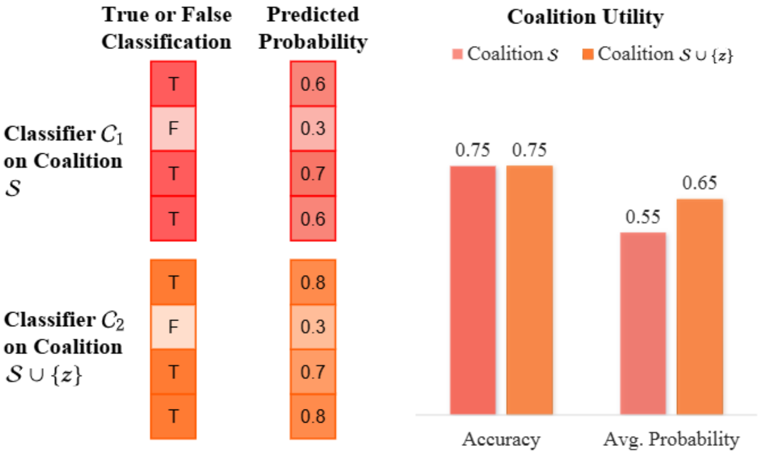


2D-Shapley

Zhihong Liu, Hoang Anh Just, Xiangyu Chang, Xi Chen, Ruoxi Jia: 2D-Shapley: A Framework for Fragmented Data Valuation. ICML 2023: 21730-21755

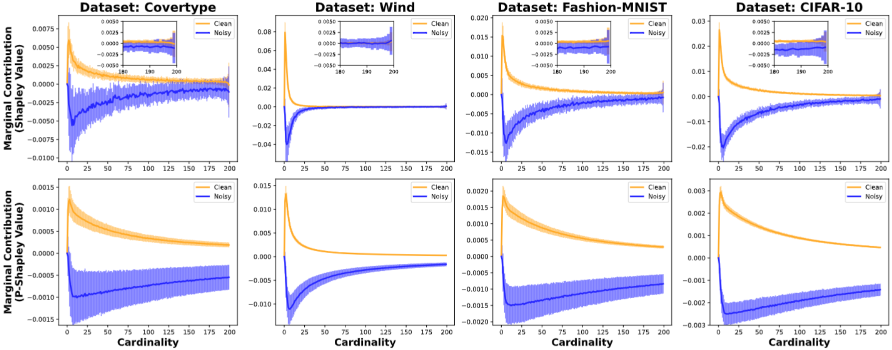
P-Shapley: Intuition

Models can present sample accuracy with different average probabilities.



P-Shapley: Intuition

$$U_p(\mathcal{S}) = \frac{1}{|V|} \sum_{z_k \in V} CF(Pr(\hat{y}_k = y_k)) \qquad PSV_i = \frac{1}{n} \sum_{S \subseteq N \setminus \{\mathbf{z}_i\}} \frac{(U_p(S \cup \{z_i\}) - U_p(S))}{\binom{n-1}{|S|}}$$

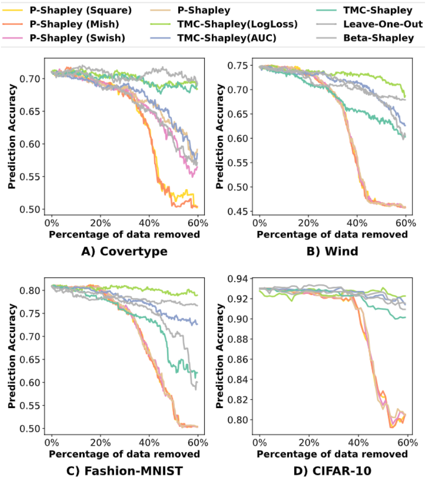


P-Shapley: Calibration function

$$U_p^+(S) = \frac{1}{|V|} \sum_{z_k \in \mathcal{V}} CF(Pr(\hat{y}_k = y_k)) \quad PSV_i^+ = \frac{1}{n} \sum_{S \subseteq N \setminus \{z_i\}} \frac{\left(U_p^+(S \cup \{z_i\}) - U_p^+(S)\right)}{\binom{n-1}{|S|}}$$

Calibration Function	Mathematical Expression
Square	$y = x^2$
Mish	$y = x \tanh\left[\ln(1 + \exp(x)\right]$
Swish*	$y = x\left[1 + \exp(-\beta x)\right]^{-1}$

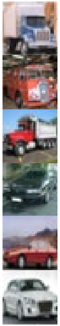
P-Shapley:experiment



P-Shapley:explanation

Base predicted probability for the correct label = 0.5

P-Shapley Value



Validation image	0	1	2	3	4	5	6	7	8	9	Output	Remove low-value data cluster
0	+0.440	-0.038	-0.058	+0.083	+0.063	-0.270	-0.055	+0.035	+0.045	+0.126	= 0.871	0.998
1	-0.023	+0.551	+0.122	-0.028	-0.049	+0.027	+0.003	-0.010	-0.165	+0.032	= 0.961	0.999
2	-0.035	+0.002	+0.490	+0.107	+0.045	-0.046	+0.025	+0.077	-0.161	-0.005	= 0.997	1.000
3	+0.063	-0.179	+0.051	+0.039	-0.005	+0.111	-0.001	+0.449	-0.021	-0.028	= 0.979	1.000
4	+0.010	+0.069	-0.267	-0.020	-0.266	-0.089	+0.065	-0.036	+0.413	-0.062	= 0.318	0.740
5	+0.044	+0.020	+0.032	+0.025	-0.252	+0.065	-0.066	+0.092	-0.072	+0.514	= 0.900	0.999

Data cluster

The closest image to the center of each data cluster (clustered by k-means)

P-Shapley

Haocheng Xia, Xiang Li, Junyuan Pang, Jinfei Liu, Kui Ren, Li Xiong: P-Shapley: Shapley Values on Probabilistic Classifiers. Proc. VLDB Endow. 17(7): 1737-1750 (2024)

Thank you for listening!