Shapley Values: Mathematical Definition

$$\varphi_i = \Sigma_{S\subseteq N\setminus\{i\}} [|S|!(n-|S|-1)! / n!] \times [v(S\cup\{i\}) - v(S)]$$

Shapley value for feature i

φ_i

Shapley Value

Contribution of feature i to the prediction

$\Sigma_{S\subseteq N\setminus\{i\}}$

Sum Over Coalitions

All possible subsets S without feature i

$|S|!(n-|S|-1)! / n!$

Weight Factor

Probability of each ordering (combinatorial weight)

$v(S\cup\{i\}) - v(S)$

Marginal Contribution

Value added when i joins coalition S

Fairness Axioms (Uniqueness Theorem)

✓ Efficiency | ✓ Symmetry | ✓ Dummy | ✓ Additivity

⚠ Complexity: $O(2^n)$ for n features

🔍 Practical Calculation Example: House Price Prediction

📊 Scenario

A house has a predicted price of **$4,000,000**. Three features contributed to this prediction:

x_1: Area (85m²) | x_2: Number of Rooms (3) | x_3: Near Subway (5 min walk)

Goal: Calculate how much each feature contributed to the prediction using Shapley Values

1 Enumerate All Coalitions for Feature x_1 (Area)

Consider all possible subsets S that exclude feature x_1. With n=3, there are $2^{3-1} = 4$ coalitions.

Coalition S	|S|	v(S)	v(S∪{x_1})	Marginal v(S∪{x_1}) - v(S)
∅ (empty set)	0	$3,000,000	$3,500,000	+$500,000
{x_2} (rooms only)	1	$3,200,000	$3,800,000	+$600,000
{x_3} (subway only)	1	$3,400,000	$3,900,000	+$500,000
{x_2, x_3} (both)	2	$3,700,000	$4,000,000	+$300,000

2 Calculate Weight for Each Coalition

Weight formula: **|S|!(n-|S|-1)! / n!** where n=3

$S = \emptyset$:

Weight = 0! × (3-0-1)! / 3! = 1 × 2! / 6 = 2/6 = **1/3**

$S = \{x_2\}$:

Weight = 1! × (3-1-1)! / 3! = 1 × 1! / 6 = 1/6 = **1/6**

$S = \{x_3\}$:

Weight = 1! × (3-1-1)! / 3! = 1 × 1! / 6 = 1/6 = **1/6**

$S = \{x_2, x_3\}$:

Weight = 2! × (3-2-1)! / 3! = 2 × 0! / 6 = 2/6 = **1/3**

3 Calculate Shapley Value for Feature x_1

$\varphi_1 = \Sigma$ [Weight × Marginal Contribution]

φ_1 = (1/3 × 500,000) + (1/6 × 600,000) + (1/6 × 500,000) + (1/3 × 300,000)

φ_1 = 166,667 + 100,000 + 83,333 + 100,000

φ_1 = \$450,000

4 Shapley Values for Remaining Features (Same Method)

Repeating the same process for x_2 and x_3:

Shapley Value for x_2 (Number of Rooms): $\varphi_2 = \$350,000$

Shapley Value for x_3 (Near Subway): $\varphi_3 = \$200,000$

✨ Final Results: Shapley Values

Area (x_1)	Rooms (x_2)	Subway (x_3)
+$450K	**+$350K**	**+$200K**

Verification: 450K + 350K + 200K = 1,000K (difference between baseline $3M and prediction $4M) ✓

💡 Interpretation

- **Area ($450K)** made the largest contribution. This means that across all possible feature combinations, area added the most value on average.
- **Number of Rooms ($350K)** is the second most important feature.
- **Near Subway ($200K)** has the smallest contribution, but still makes a positive impact.
- The sum of all Shapley Values exactly equals the difference between prediction and baseline ($1M). This satisfies the **Efficiency axiom**.

- Each feature's contribution is fairly distributed by considering all possible combinations with other features.