BULLS AND COWS

Mathematical Foundations for Data Science

Aslesha Sanjana Kodavali

AGENDA

- Introduction
- Overview of Bulls and Cows Game
- Entropy and Mutual Information
- Practical Implications
- Applications of Entropy
- Results

INTRODUCTION

Objective

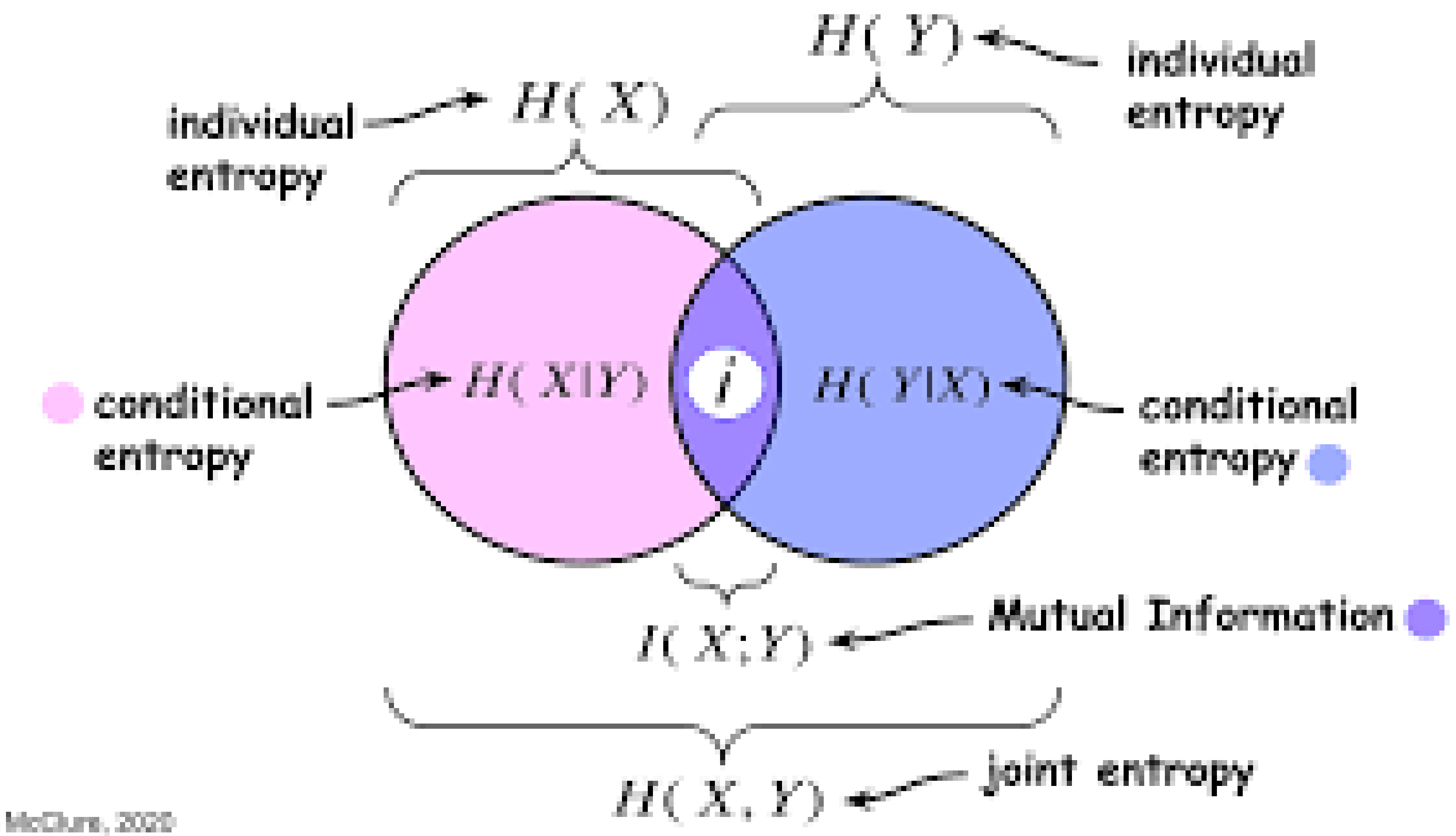
- Build an interactive Bulls and Cows game with entropy calculations.
- Explore mutual information for uncertainty reduction.

Why Entropy?

Entropy quantifies the uncertainty in data, making it crucial for decision-making systems, especially in data science applications like classification and prediction.

Importance of Mutual Information

- Measures the effectiveness of each guess in reducing uncertainty.
- Provides insights into how guesses refine possible outcomes.



Sean McClure, 2020

Overview of Bulls and Cows Game

Game Rules

- The system generates a random 4-digit number with unique digits.
- Players guess a number.
- Feedback:
- Bulls: Correct digits in correct positions.
- Cows: Correct digits in wrong positions.

Winning Condition

The player guesses the secret number correctly (4 Bulls).

Entropy Integration

With each guess, entropy is recalculated to show the remaining uncertainty.

Visual: Example

Secret: 1234, Guess: 1256, Feedback: 2 Bulls, 0 Cow.

WHAT IS ENTROPY?

Definition

- Entropy is a measure of uncertainty or randomness in a system.
- It quantifies how unpredictable a variable's outcome is.

Mathematical Formula

$$H = -\sum p(x) \log p(x)$$

H: Entropy of the random variable

P(x): probability of a specific outcome x

Intuition

- Higher entropy: More uncertainty (e.g., choosing randomly among 10,000 numbers).
- Lower entropy: Less uncertainty (e.g., choosing among 2 remaining numbers).

Application in Game

- At the start, entropy is high since there are 10,000 possible secret numbers.
- After each guess, entropy reduces as possibilities narrow down.

Example in Game

- Initial state: 10,000 possible numbers.
- $H=\log2(10{,}000)\approx 13.29H=\log2(10{,}000)\approx 13.29$.
- After a guess that eliminates 75% of possibilities:
- Remaining: 2,500 numbers.
- Updated $H=\log2(2{,}500)\approx 11.29H=\log2(2{,}500)\approx 11.29$.

HOW TO CALCULATE ENTROPY?

1. **Initial Entropy:**
   - Calculate the total number of possible 4-digit numbers with unique digits.
   - Formula: H(X)=log2(total possibilities)H(X)=log2(total possibilities).
   - Example:
     - Total possibilities: 10,000.
     - Entropy: log2(10,000)≈13.29log2(10,000)≈13.29.
2. **Filtering Possibilities:**
   - After a guess, use the Bulls and Cows feedback to filter numbers.
   - Example:
     - Guess: 1256, Feedback: 2 Bulls, 0 Cow.
     - Filter the set of numbers that could generate the same feedback.
3. **Updated Entropy:**
   - Calculate entropy for the reduced set of possibilities.
   - Formula remains H(X)=log2(remaining possibilities)
   - Example:
     - Remaining possibilities: 2,500.
     - Entropy: log2(2,500)≈11.29log2(2,500)≈11.29.

Guess	Remaining Numbers	Entropy H(X)	Mutual Info I(X;Y)
Initial	10,000	13.29	N/A
Guess 1	2,500	11.29	2.00
Guess 2	1,024	10.00	1.29
Guess 3	512	9.00	1.00

4. **Mutual Information:**
   - Subtract the updated entropy from the initial entropy.
   - Example:I(X;Y)=H(X)−H(X ∣ Y)=13.29−11.29=2.00I(X;Y)=H(X)

WHAT IS MUTUAL INFORMATION?

Definition

- Mutual information measures the reduction in uncertainty about one variable given information about another.
- It quantifies the "information gained" about the secret number after a guess.

Mathematical Formula

I(X;Y)=H(X)−H(X ∣ Y)

I(X;Y)I(X;Y): Mutual information between variables XX and YY.

H(X)H(X): Entropy of XX before observing YY (initial state).

H(X ∣ Y)H(X ∣ Y): Conditional entropy of XX given YY (after observing the guess).

Intuition

- High mutual information: A guess significantly reduces the number of possibilities.
- Low mutual information: A guess has little impact on reducing possibilities

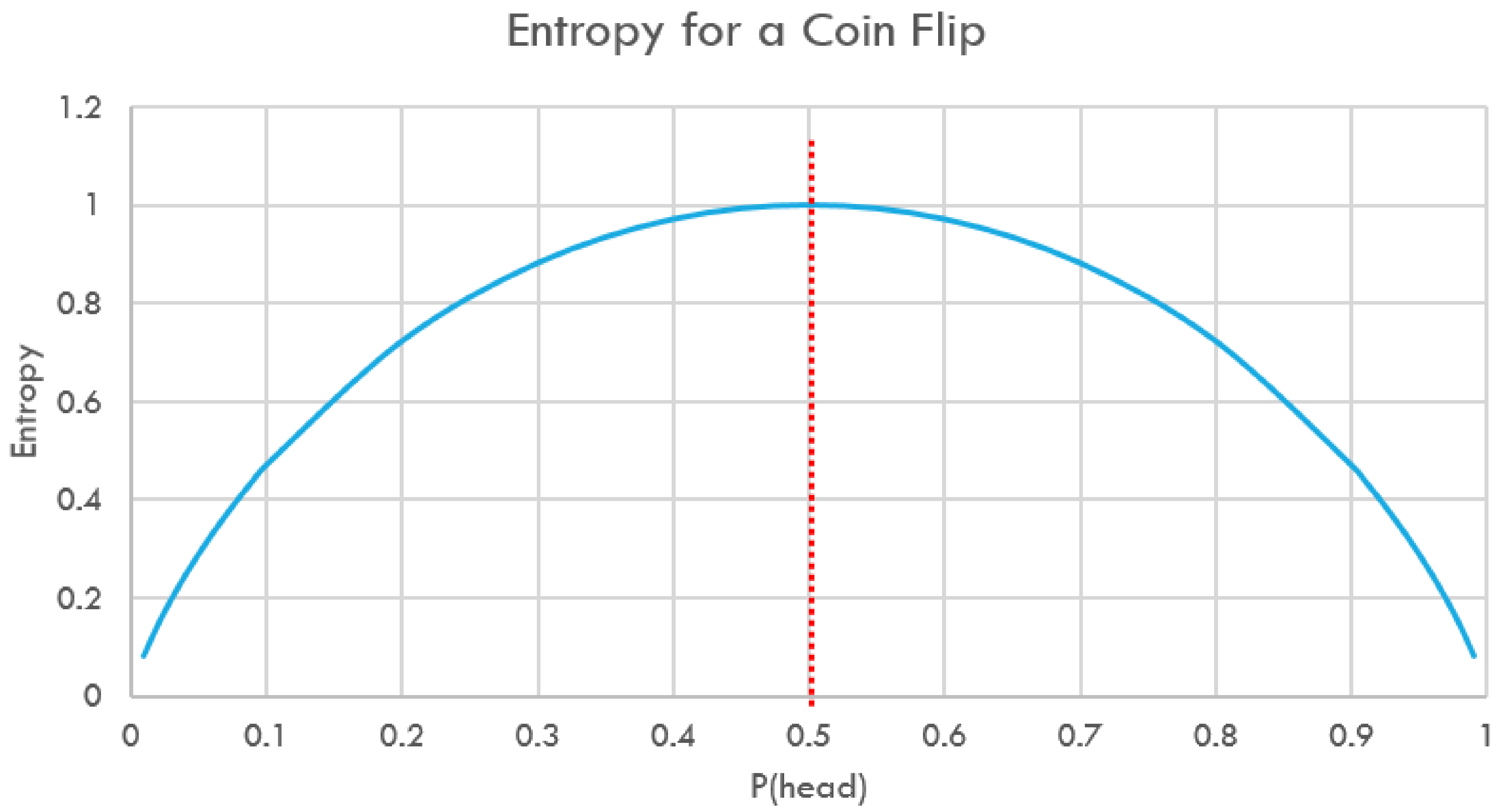
Application in Game

- With each guess, mutual information indicates how effectively uncertainty is reduced.

PRACTICAL IMPLICATIONS :

- Applications of Entropy:
  - Decision Trees: Splitting nodes based on information gain.
  - Machine Learning: Feature selection using mutual information.
  - Communication Systems: Measuring data encoding efficiency.

- Insights from the Game:
  - Entropy helps quantify progress in narrowing down possibilities.
  - Mutual information highlights the effectiveness of guesses.
  - Game mechanics are analogous to real-world data science problems, where reducing uncertainty is key



Applications of Entropy

Understanding Uncertainty:

- Entropy provides a numerical measure of uncertainty in a system.
- In the game:
  - At the start, with 10,000 possible secret numbers, uncertainty is high (entropy ≈ 13.29).
  - Each guess reduces the possible numbers, lowering uncertainty.

Practical Insight:

- Entropy helps quantify progress in problem-solving.
- It shows how much uncertainty remains, enabling players to gauge the effectiveness of their strategy.

Quantifying the Value of Information:

- Mutual information measures how much a guess reduces uncertainty.
- Effective guesses yield high mutual information, eliminating many possibilities.

Practical Insight:

- Mutual information provides feedback on the quality of guesses.
- Players can adjust their strategy to maximize uncertainty reduction.

Feature Selection:

- Mutual information is used to identify the most informative features in datasets.
- This reduces the dimensionality of data while retaining predictive power.

RESULTS

Entropy Reduction:

Example:

- Initial entropy: 13.2913.29 (10,000 possibilities).
- After Guess 1: 11.2911.29 (2,500 possibilities).
- After Guess 2: 10.0010.00 (1,024 possibilities).

Mutual Information:

- Highlight which guesses were most effective in reducing uncertainty.
- Example:
  - A guess with mutual information of 2.002.00 was twice as effective as one with 1.001.00

THANK YOU!