Dr. Denton Bobeldyk

CIS 365 Artificial Intelligence

Information Theory

Week in Review

Blackboard Check-in

Delivery Methods

Lecture

Videos

Lab Time

Small Groups

Information Theory

- Information theory is a branch of applied mathematics that revolves around quantifying how much information is present in a signal

Information Theory

How much information is contained in the following?

- The sun rose this morning
- There was a solar eclipse this morning

Information Theory

- How much information is contained in the following?
  - The sun rose this morning — "Tell me something I don't know!"
  - There was a solar eclipse this morning — "Oh wow that's cool!"

Information Theory Class Exercise

What types of communications have (or could you have had) this morning that contained little to no information?

Information Theory Class Exercise

What types of communications have (or could you have had) this morning that contained a large amount of information?

Information Theory - Quantification

- Quantify information in a way that formalizes this intuition:
  - Likely events should have low information content
  - Events that are guaranteed to happen should have no information content
  - Less likely events should have higher information content
  - Independent events should have additive information.
    - For example: Finding out that a tossed coin has come up as heads twice should convey twice as much information as finding out that a tossed coin has come up heads once.

Father of Information Theory

Self-Information

Define the self-information of an event $\mathrm{x} = x$

$$I(x) = -\, log P(x)$$

Self-Information (Units)

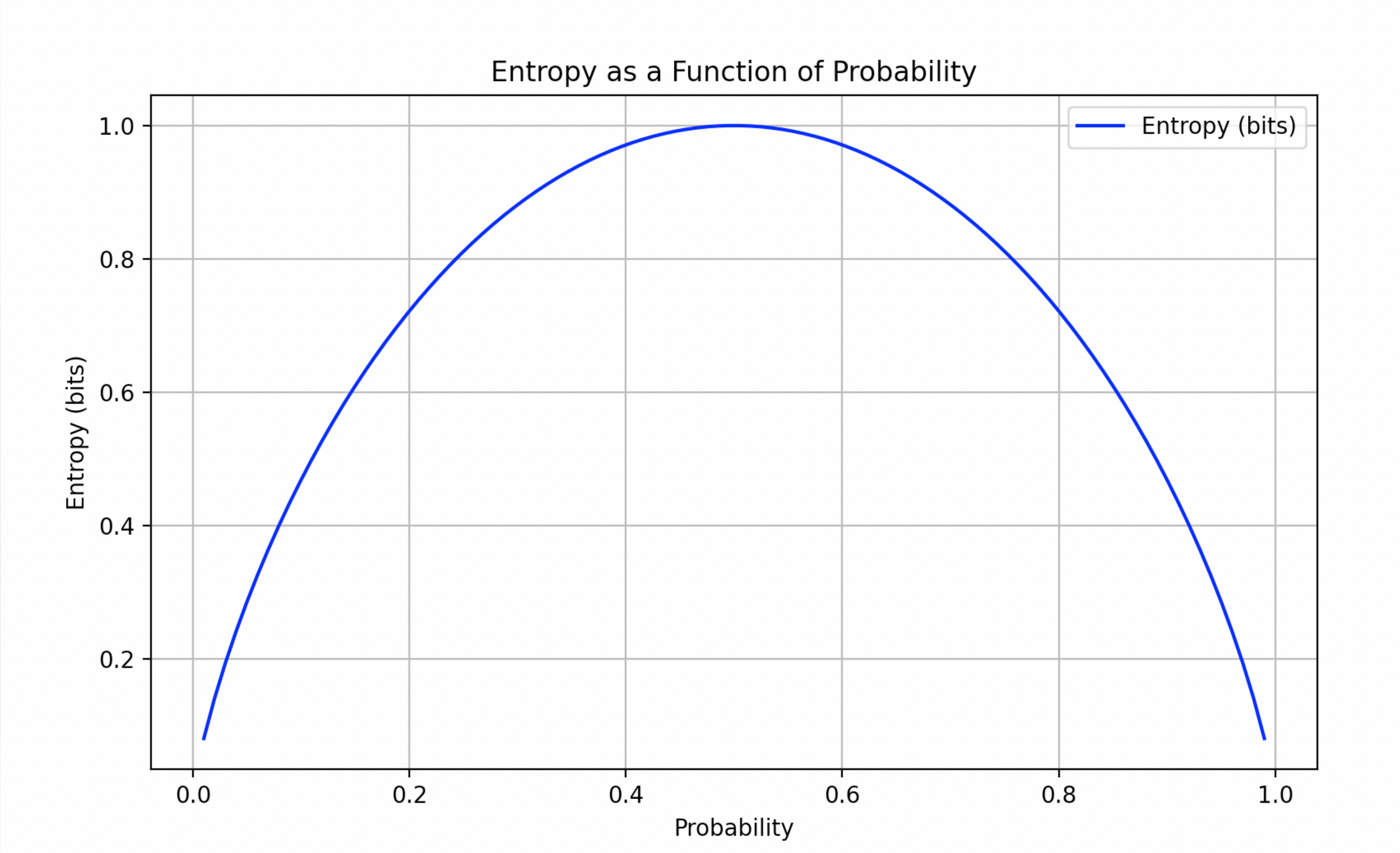
Define the self-information of an event $\mathrm{x} = x$

$$I(x) = -\,log P(x)$$

Log base e the units are written in terms of **nats**

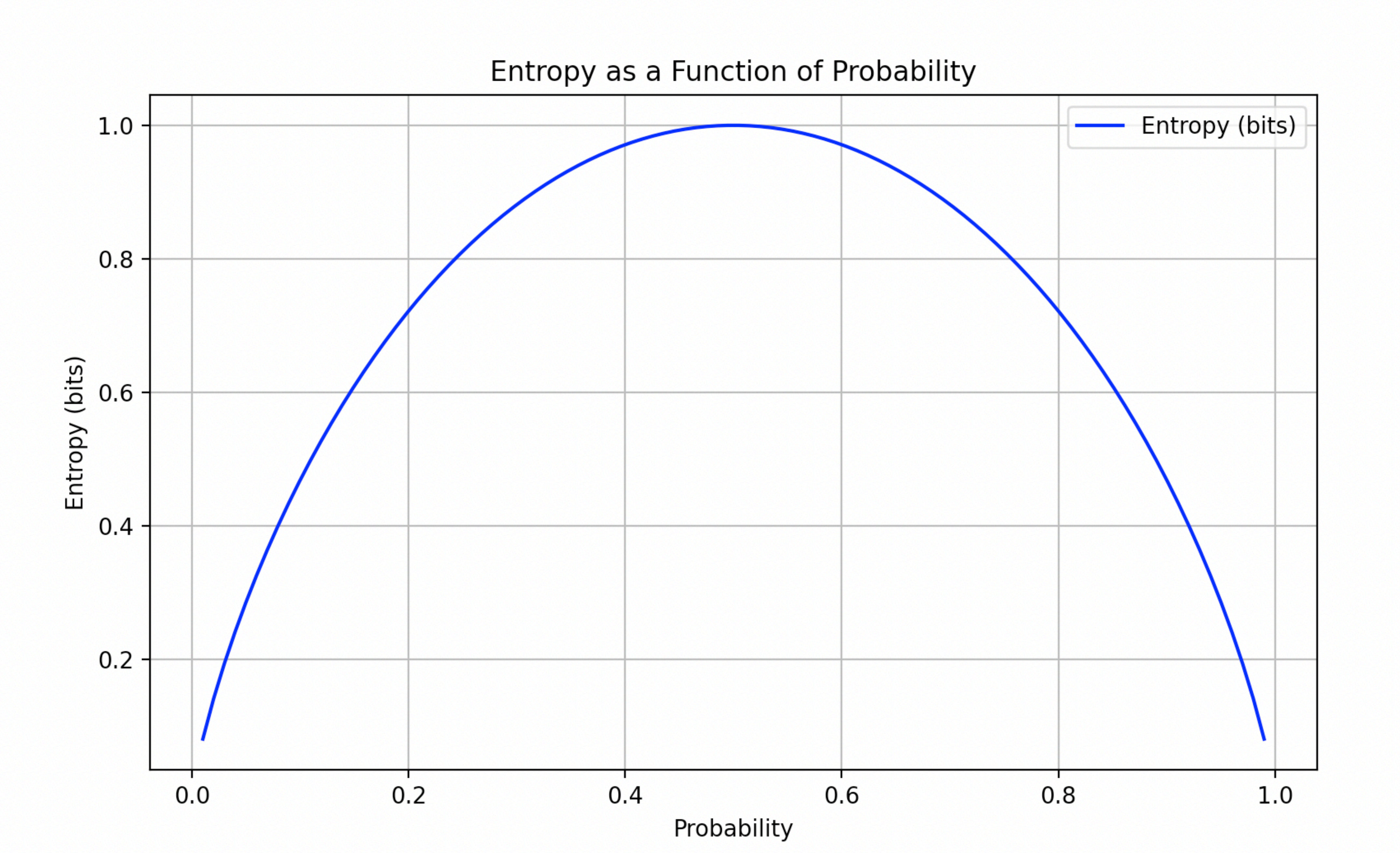
Log base 2 the units are written in terms of **bits** or **shannons**

Shannon Entropy of a binary random variable



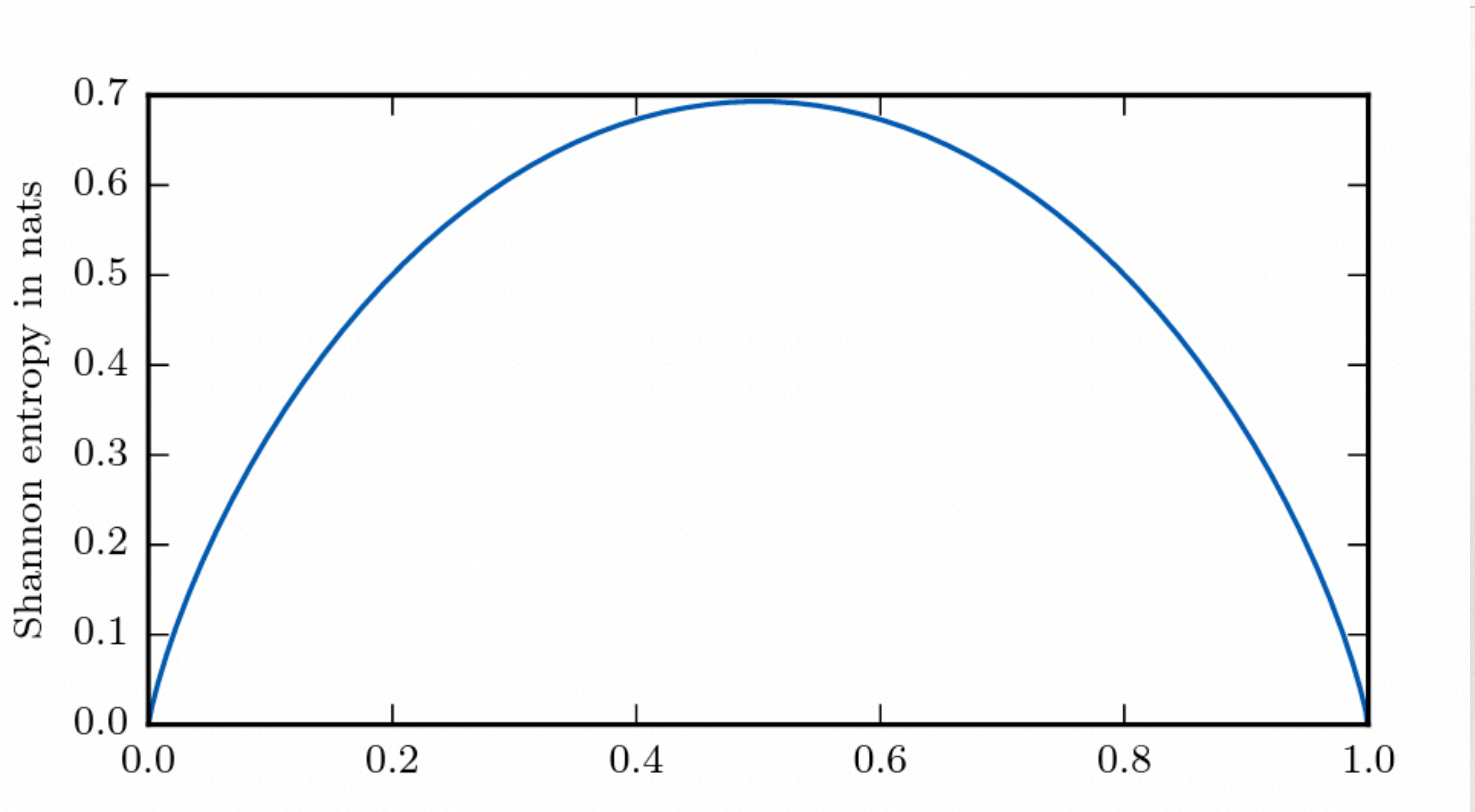
Horizontal Axis shows p, the probability of a binary random variable being equal to 1
The entropy increases while close to uniform and decreases on the tails

Entropy of a binary random variable



Read this as the entropy is largest from an event that has a 50% probability of occurring. For example flipping a fair coin gives more information than flipping a bias coin.

Shannon Entropy of a binary random variable



Probability(X)

Horizontal Axis shows p, the probability of a binary random variable being equal to 1

The entropy increases while close to uniform and decreases on the tails

Maximum Entropy

Maximum entropy occurs when all outcomes are equally probable and there is the most uncertainty about the result

Entropy

Entropy of a system $= -\sum_{i-1}^{n} p(x_i) log_2 p(x_i)$

Where $p(x_i)$ is the probability of each outcome

Entropy

Entropy of a system $= -\sum_{i-1}^{n} p(x_i) log_2 p(x_i)$

Where $p(x_i)$ is the probability of each outcome

Given the following set, calculate the entropy:

[12(+), 3(-)] =

Entropy Calculation Examples

Recall the entropy of a system $= -\sum_{i-1}^{n} p_i log_2 p_i$

Given the following set, calculate the entropy:

$[12(+), 3(-)] = -\frac{12}{15} log_2(\frac{12}{15}) + (-\frac{3}{15} log_2(\frac{3}{15}))$

Entropy Calculation Class Exercise

Recall the entropy of a system $= -\sum_{i-1}^{n} p_i log_2 p_i$

Given the following sets, calculate their entropy:

[6(+), 3(-)]=

[4(+), 8(-)] =

[7(+), 7(-)] =

[9(+), 2(-)] =

Entropy Calculation First Example

Recall the entropy of a system $= -\sum_{i-1}^{n} p_i log_2 p_i$

- Given the following sets, calculate their entropy:

$$[6(+), 3(-)] = -\frac{6}{9}log_2(\frac{6}{9}) + (-\frac{3}{9}log_2(\frac{3}{9}))$$

$$p(+) = 6/9$$

$$p(-) = 3/9$$

Python Code

```
mport numpy as np
import matplotlib.pyplot as plt

# Function to calculate entropy in bits
def entropy_bits(p):
    if p == 0 or p == 1:
        return 0
    return -p * np.log2(p) - (1 - p) * np.log2(1 - p)

# Generate probabilities from 0 to 1 (excluding 0 and 1)
probabilities = np.linspace(0.01, 0.99, 100)

# Calculate entropy for each probability
entropies = [entropy_bits(p) for p in probabilities]

# Plot the entropy
plt.plot(probabilities, entropies, label='Entropy (bits)', color='blue')
plt.xlabel('Probability')
plt.ylabel('Entropy (bits)')
plt.title('Entropy as a Function of Probability')
plt.grid(True)
plt.legend()
plt.show()
```

Information Gain

Information Gain Example

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

Target: Predict if we'll play outside
Step 1: Calculate the entropy of the entire dataset
Step 2: Calculate the entropy after splitting by "Weather"
Step 3: Calculate the weighted average entropy after the split
Step 4: Calculate the information gain

Information Gain Example

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

Target: Predict if we'll play outside
Step 1: Calculate the entropy of the entire dataset
Step 2: Calculate the entropy after splitting by "Weather"
Step 3: Calculate the weighted average entropy after the split
Step 4: Calculate the information gain

Will be useful for decision trees later

Step 1: Calculate the entropy of the entire dataset

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

10 examples:
6 Yes
4 No

Formula for Entropy:

$$H(S) = -\sum_{i-1}^{n} p_i log_2 p_i$$

$$H(S) = -(p_{yes} log_2(p_{yes}) + p_{no} log_2(p_{no}))$$

Notation Note: H(S) is common way to represent entropy of a system

Step 1: Calculate the entropy of the entire dataset

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

$$H(S) = -(p_{yes}log_2(p_{yes}) + p_{no}log_2(p_{no}))$$

$$P_{yes} = \frac{6}{10} \qquad P_{no} = \frac{4}{10}$$

$$H(S) = -(\frac{4}{10}log_2(\frac{4}{10}) + \frac{6}{10}log_2(\frac{6}{10}))$$

$$H(S) = -(.4log_2(.4) + .6log_2(.6))$$

$$H(S) = -(.4 \times -1.32 + .6 \times -0.737)$$

$$H(S) \approx 0.971$$

Step 2: Calculate entropy after splitting by “Weather”

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

Weather has 3 possible values: Sunny, Overcast, Rainy
Let’s calculate the entropy for each subset after splitting

Step 2: Calculate Entropy for *Sunny* Subset

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

$$H(S_{sunny}) = -(p_{yes} log_2(p_{yes}) + p_{no} log_2(p_{no}))$$

$$P_{Sunny_{yes}} = \frac{3}{4} \qquad P_{Sunny_{no}} = \frac{1}{4}$$

$$H(S_{sunny}) = -(\frac{3}{4} log_2(\frac{3}{4}) + \frac{1}{4} log_2(\frac{1}{4}))$$

$$H(S_{sunny}) = -(.75 log_2(.75) + .25 log_2(.25))$$

$$H(S_{sunny}) \approx 0.811$$

Step 2: Calculate Entropy for *Overcast* Subset

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

$$H(S_{overcast}) = -(p_{yes}log_2(p_{yes}) + p_{no}log_2(p_{no}))$$

$$P_{overcast_{yes}} = \frac{2}{2} \qquad P_{overcast_{no}} = \frac{0}{2}$$

$$H(S_{overcast}) = -(\frac{2}{2}log_2(\frac{2}{2}) + \frac{0}{2}log_2(\frac{0}{2}))$$

$$H(S_{overcast}) = -(1 \times log_2(1))$$

$$H(S_{overcast}) = 0$$

Entropy is 0 since all samples are of the same class

Step 2: Calculate Entropy for *Rainy* Subset

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

$$H(S_{Rainy}) = -(p_{yes}log_2(p_{yes}) + p_{no}log_2(p_{no}))$$

$$P_{Rainy_{yes}} = \frac{1}{4} \qquad P_{Rainy_{no}} = \frac{3}{4}$$

$$H(S_{Rainy}) = -(\frac{1}{4}log_2(\frac{1}{4}) + \frac{3}{4}log_2(\frac{3}{4}))$$

$$H(S_{Rainy}) = -(.25log_2(.25) + .75log_2(.75))$$

$$H(S_{Rainy}) \approx 0.811$$

Step 3: Calculate the weighted average entropy after the split

Weather	Play Outside?
Sunny	Yes
Sunny	No
Overcast	Yes
Rainy	No
Rainy	No
Rainy	Yes
Overcast	Yes
Sunny	Yes
Sunny	Yes
Rainy	No

$H(S_{Rainy}) \approx 0.811$ $H(S_{Sunny}) \approx 0.811$ $H(S_{Overcast}) = 0$

$$H(S_{AfterSplit}) = \frac{4}{10}H(S_{Sunny}) + \frac{2}{10}H(S_{Overcast}) + \frac{4}{10}H(S_{Rainy})$$

$$H(S_{AfterSplit}) = 0.3244 + 0 + 0.3244$$

$$H(S_{AfterSplit}) = 0.6488$$

Step 4: Calculate the Information Gain

Information Gain = H(S) - $H(S_{afterSplit})$

Information Gain = 0.971 - 0.6488

Information Gain = 0.3222

Step 4: Calculate the Information Gain

The information gain for splitting the dataset based on the “Weather” attribute is 0.3222. This means by using “Weather” as a splitting criterion, we reduce the uncertainty (entropy) in the dataset by 0.3222

Decision Trees

- What is a decision tree?
- Why do we want the item with the most information gain at the top of the decision tree?

Decision Trees & Entropy

Why do we want the item with the most information gain at the top of the decision tree?

1. Maximizing uncertainty reduction. This reduces the number of splits (decisions) that will occur

2. Efficient decision making. Each decision leads to the most effective separation of data

3. Minimizing tree depth

4. Better generalization. If suboptimal splits were made early, the model might need more splits to classify the data and lead to poor generalization on test data

Entropy Dataset Calculation

Weather	Temperature	Humidity	Wind	Play Outside?
Sunny	Hot	High	Weak	Yes
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rainy	Mild	High	Weak	No
Rainy	Cool	Normal	Weak	No
Rainy	Cool	Normal	Strong	Yes
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	Yes
Sunny	Cool	Normal	Weak	Yes
Rainy	Mild	Normal	Weak	No

END