

Entropy and Information Gain*(find and work with a partner)*

1. Given the probabilities of each class below, what is the average number of bits needed to transmit one value?

Class	Probability	Encoding
Senior	0.5	
Junior	0.25	
Sophomore	0.125	
First-year	0.125	

Based on example from Ameet Soni

2. Given the movie data below (Liked (Li) is the response variable), select the feature that maximizes the information gain. This will be the root node in our decision tree. *Use a calculator for $H(Li)$.*

Movie	Type	Length	Director	Famous actors	Liked?
m1	Comedy	Short	Adamson	No	Yes
m2	Animated	Short	Lasseter	No	No
m3	Drama	Medium	Adamson	No	Yes
m4	Animated	Long	Lasseter	Yes	No
m5	Comedy	Long	Lasseter	Yes	No
m6	Drama	Medium	Singer	Yes	Yes
m7	Animated	Short	Singer	No	Yes
m8	Comedy	Long	Adamson	Yes	Yes
m9	Drama	Medium	Lasseter	No	Yes

$$P(Li = \text{yes}) =$$

$$H(Li) =$$

$$H(Li | T) = 0.61$$

$$H(Li | Le) = 0.61$$

$$H(Li | D) = 0.36$$

$$H(Li | F) = 0.85$$

$$\text{Gain}(Li, T) =$$

$$\text{Gain}(Li, Le) =$$

$$\text{Gain}(Li, D) =$$

$$\text{Gain}(Li, F) =$$

Based on materials by Jessica Wu and Ziv Bar-Joseph