Hamming’s Distiance

The Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different.
Put another way, it measures the minimum number of substitutions required to change one string into the other, or the number of errors that transformed one string into the other.

Hamming's Distiance

The Hamming distance is named after Richard Hamming, who introduced it in his fundamental paper on Hamming codes Error detecting and error correcting codes in 1950. It is used in telecommunication to count the number of flipped bits in a fixed-length binary word as an estimate of error, and therefore is sometimes called the signal distance. Hamming weight analysis of bits is used in several disciplines including information theory, coding theory, and cryptography.

Information

x_i	A	B	C	D
$p(x_i)$	0.5	0.25	0.125	0.125
$1/p(x_i)$	2	4	8	8
$\log(1/p(x_i)$	0.30103	0.60206	0.90309	0.90309
$p(x_i)[\log(1/p(x_i)]$	0.15051	0.15051	0.11288	0.11288

$$H(X) = \sum p(x_i) \left[\log(\frac{1}{p(x_i)}\right]$$

$$H(X) = 0.15051 + 0.15051 + 0.11288 + 0.11288$$

$$H(X) = 0.52680$$

Examples

The fair coin
H = -1/2 log2(1/2) - 1/2 log2(1/2)
= 1/2 + 1/2
= 1 bit
That biased coin, P(head)=0.75, P(tail)=0.25
$H = -3/4log2(3/4) - 1/4log2(1/4) = 3/4 \times 0.42 + 2/4 = 0.31 + 1/2$
$= 0.81$bits, approx. A biased four-sided dice, p(a)=1/2, p(c)=1/4, p(g)=p(t)=1/8

$$H = -1/2log2(1/2) - 1/4log2(1/4) - 1/8log2(1/8) - 1/8log2(1/8)$$

= 1 3/4 bits

bits = - $log_2 p$ where p is the probability with which a particular value occurs

- bits(A) = - log 2 1/2 = 1
- bits(B) = - log 2 1/4 = 2
- bits (C) = bits(D) = - log 2 1/8 = 3

Entropy

- High Entropy means that we are sampling from a uniform (boring) distribution. Would have a flat histogram, therefore we have an equal chance of obtaining any possible value.
- Low Entropy means that the distribution varies, it has peaks and valleys. The histogram of frequency distribution would have many lows and maybe one or two highs. Hence it is more predictable.
- Entropy is a measure of how pure or impure a variable is.

Information Gain

Information Gain is the number of bits saved, on average, if we transmit Y and both receiver and sender know X