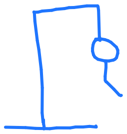
# Learning information theory with hangman



$$P(x,y) = \begin{cases} 1/N & \text{if } "xy_1y_2" \text{ is a word} \\ 0 & \text{else} \end{cases}$$

N chosen to satisfy $\sum_{x \in X} \sum_{y \in Y} P(x,y) = 1$

$\Rightarrow N =$ # of 3-letter words

$\underbrace{\_}_{X \in \mathbb{Z}^{26}} \ \underbrace{\_ \ \_}_{Y \in \mathbb{Z}^{26} \times \mathbb{Z}^{26}}$

The joint entropy is defined $H(X,Y) = -\sum_{\substack{x \in X \\ y \in Y}} p(x,y) \lg p(x,y)$.

It turns out, there are 759 three-letter words, in a state space of $26^3 = 17576$ combinations.

$\Rightarrow \frac{1}{N} = \frac{1}{759} = .0013$. $\quad \frac{759}{17576} = .043$

$\Rightarrow H(X,Y) = -\sum_{x,y} \frac{1}{N} \lg \frac{1}{N} = -\lg \frac{1}{N} = \lg N = 9.56$ bits

$\Rightarrow$ Roughly, if done properly it will take 9-10 yes/no questions to determine a given 3-letter word

The conditional entropy is defined

$$H(Y|X) = \sum_{x \in X} p(x) \, H(Y|X=x)$$
$$= -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \lg p(y|x)$$
$$= -\sum_{x,y} p(x,y) \lg p(y|x) = -\sum_{x,y} p(x,y) \lg \left[\frac{p(x,y)}{p(x)}\right]$$

$p(x) = \sum_{y \in Y} p(x,y)$

Lastly, the mutual information is given by

$$I(X;Y) = H(X) - H(X|Y)$$
$$= H(Y) - H(Y|X)$$

All together:

$$H(X,Y) = 9.56 \text{ bits}$$
$$H(X) = 4.44 \text{ bits}$$
$$H(Y) = 7.19 \text{ bits}$$
$$I(X;Y) = 2.11 \text{ bits}$$

$H(X|Y) = 2.29$ bits $\Rightarrow$ once you tell me the last two letters, I will determine the first letter after 2-3 yes/no questions

C A R
T
B
F

$H(Y|X) = 5.08$ bits $\Rightarrow$ once you tell me the first letter, I need ~5 guesses to determine the word

$$H(X,Y) = \underline{H(X)} + \underline{H(Y|X)}$$
$$= \underline{H(Y)} + \underline{H(X|Y)}$$

T H E

S H E

Lastly, the largest values of $p(X)$ are

S _ _ ~2x as common as average

A _ _ ~1.9x

T _ _ , ~1.7x

and the largest values of $p(Y)$ are

_ AN ~14x

_ AY ~12x

_ OW ~12x

_ AT ~11x

_ AM ~10x