Information Theory Basics

Xiangbo Li

February 5, 2020

1 Information Defintion

$$I = \log(1/p) = -\log(p) \tag{1}$$

Example:

- **One flip of a fair coin**: Before the flip, there are two equally probable choices: heads or tails. After the flip, we've narrowed it down to one choice. Amount of $information = log2(2/1) = 1$ bit.
- **Roll of two dice**: Each die has six faces, so in the roll of two dice there are 36 possible combinations for the outcome. Amount of $information = log2(36/1) = 5.2$ bits.

2 Entropy

Now that we know how to measure the information contained in a given event, we can quantify the expected information in a set of possible outcomes. Specifically, if an event i occurs with probability p_i, $1 \leq n \leq N$out of a set of N events, then the average or expected information is given by

$$H(p_1, p_2, ..., p_N) = \sum_{i=1}^{n} p_i \log(1/p_i) \tag{2}$$

H is also called the entropy (or Shannon entropy) of the probability distribution. It is often useful to think of the entropy as the average or expected uncertainty associated with this set of events. Shannon proved that the expected code length of any decodable code cannot be smaller than the entropy, H, of the underlying probability distribution over the symbols.