Chapter 1: The Science of Information

April 18, 2020

Two Basic Concepts

- Information is uncertainty: modeled by random variable
- Information is digital: transmission should be 0's and 1's (bits) w/o reference to what they really represent.

Two Fundamental Theorem

- Source Coding Theorem
  - No matter how smart the **compression system** designed, there is a fundamental limit the data compression.
- Channnel Coding Theorem
  - No matter how smart the **communication system** designed, for reliable communication, there exists a maximum rate that information can be transmitted through the channel.
  - This quality is called the **communication capacity**.
  - P.S. data storage can also be a form a data communication (from the past to the future).

Links

Wikipedia: [Information Theory]

> Information Theory

- **Bit**: the most fundamental unit of information.
- **Important quantities of information**:
  - **Entropy**: a measure of information in a single random variable.
    - Gives a limit on the rate at which data can be reliably compressed.
  - **Mutual Information**: a measure of information in common for two random variables.
    - Gives a limit on the rate of reliable communication across a noisy channel.
- **Terms**:
  - **Entropy**:
    - $H = \mathbb{E}_X[-\log p(x)] = -\sum_i p_i \log_2(p_i)$
      - This equation gives the entropy in the units of "bits" (per symbol) because it uses a logarithm of base 2.
    - $H(X) = \mathbb{E}_X[I(x)] = -\sum_{x\in\mathbb{X}} p(x) \log p(x)$

            - $I(x)$ is the self-information, which is the entropy contribution of an individual message.
    - **Joint Entropy**:
        - $H(X,Y) = \mathbb{E}_{X,Y}[-\log p(x,y)] = -\sum_{x,y} p(x,y) \log p(x,y)$
    - **Conditional Entropy**:
        - $H(X|Y) = \mathbb{E}_Y[H(X|y)] = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) = -\sum_{x,y} p(x,y) \log p(x|y)$
            - $H(X|Y) = H(X,Y) - H(Y)$
    - **Mutual Information**:
        - Measures the amount of information that can be obtained about one $r.v.$, by observing another (e.g. the mutual information of x relative to y).
        - It can be used to maximize the information shared between sent and received signals.
        - $I(X;Y) = \mathbb{E}_{X,Y}[SI(x,y)] = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$
            - $I(X;Y) = H(X) - H(X|Y)$
            - $I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y)$
            - $I(X;Y) = \mathbb{E}_{p(y)}\left[D_{\mathrm{KL}}(p(X|Y=y) \| p(X))\right]$
            - $I(X;Y) = D_{\mathrm{KL}}(p(X,Y) \| p(X)p(Y))$
    - **KL-Divergence / Information Gain**:
        - KL-Divergence is the number of average additional bits per datum necessary for compression.
        - "Unnecessary surprise" by the prior from the truth:
        - $D_{\mathrm{KL}}(p(X) \| q(X)) = \sum_{x \in X} -p(x) \log q(x) - \sum_{x \in X} -p(x) \log p(x)$
        - $D_{\mathrm{KL}}(p(X) \| q(X)) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$

> Coding Theory

- **Coding theory** is concerned with **finding explicit methods**, called *codes*, for *increasing the efficiency* and *reducing the error rate* of **data communication** over noisy channels to near the *channel capacity*.
    - Data compression: Source coding.
        - Any process that generates successive messages can be considered as a **source** of information.
    - Error correction: Channel coding.
- **Classification**
    - **Source Coding** (Data Compression)
        - **Lossless Data Compression**
        - **Lossy Data Compression / Rate-Distortion Theory**
    - **Channel Coding** (Error Correcting)
- Compression followed by transmission is optimal for only one transimitter and one receiver. When there are multiple, we will use network information theory.
- **Information rate** is the average entropy per symbol.
    - $r = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \ldots X_n)$

    - The rate of a source of information is related to its *redundancy* and how well it can be *compressed*.
- **Channel capacity**:
    - $C = \max_f I(X;Y)$

Wikipedia: Claude Shannon

- Shannon's mouse appears to have been the first artificial learning device of its kind.
- Shannon developed Alzheimer's disease and spent the last few years of his life in a nursing home.