Notes for EECS 550: Information Theory

Yiwei Fu, Instructor: David Neuhoff

FA 2022

Contents

Office hours:

Chapter 1

Introduction

1.1 What is Information Theory

1.2 Lossless Coding

It is a type of data compression.

GOAL to encode data into bits so that

1. bits can be decoded perfectly or with very high accuracy back into original data;
2. we use as few bits as possible.

We need to model for data, a measure of decoding accuracy, a measure of compactness.

MODEL FOR DATA

Definition 1.2.1. A *source* is a sequence of i.i.d (discrete) random variables $U_1, U_2, \dots$

We would like to assume a known alphabet $A = \{a_1, a_2, \dots, a_Q\}$ and known probability distribution either through probability mass functions $p_U(u) = \Pr[U = u]$.

Definition 1.2.2. Source coding

PERFORMANCE MEASURES A measure of compactness (efficiency)

Definition 1.2.3. rate = encoding rate = average number of encoded bits per data symbol

Two versions: empricial avg rate $\langle r \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} L_k(U_1, \dots, U_k)$.

Statistical avg rate:

$$\overline{r} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[L_k(U_1, \dots, U_k)]$$

where L_K is the number of bits out of the encoder after U_k and before U_{k+1}.

Accuracy per-letter frequency of error

$$\langle F_{LE} \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} I(\hat{U}_k = U_k)$$

per-letter error probability

$$p_{LE} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[I(\hat{U}_k = U_k)] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \Pr(\hat{U}_k = U_k)$$

Fixed-length to fixed-length block codes (FFB)

characteristics

A code is perfectly lossless (PL) if the $\beta(\alpha(\underline{u})) = \underline{u}$ for all $\underline{u} \in A_U^k$ (the set of all sequences $u_1, \dots, u_k$).

In order to be perfectly loss, α must be one-to-one. Encode must assign a distinct codeword (L bits) to each data sequences. rate = L/K. We seek $R^*_{PL}(k)$ the smallest rate of any PL code.

Number of sequences of size $k = Q^k$, and number binary sequence of size $L = 2^L$. We need $2^L \gg Q^K$.

$$\overline{r} = \frac{L}{k} \geq \frac{k \log_2 Q}{k} = \log_2 Q$$

Choose $\lceil k \log_2 Q \rceil$, then we have

$$R^*_{PL}(k) = \frac{\lceil k \log_2 Q \rceil}{k} \leq \frac{k \log_2 Q + 1}{k} = \log_2 Q + \frac{1}{k}.$$

$$\log_2 Q \leq R^*_{PL}(k) \leq \log_2 Q + \frac{1}{k}$$

Let R^*_{PL} be the least rate of any PL FFB code with any k. $R^*_{PL}(k) \to \log_2 Q$ as $k \to \infty$.

$R^*_{PL} = \inf_k R^*_{PL}(k)$

Now we want rate less and $\log_2 Q$ almost lossless codes.

$R^*_{AL} = \inf\{r,$ there is an FFB code with $\bar{r} \leq n$ and arbitrarily small $P_{LE}\}$

$= \inf\{r,$ there is an FFB code with $\bar{r} \leq n$ and $P_{LE} < \delta$ for all $\delta > 0\}$

Instead of per-letter probability P_{LE}, we focus on block error probability $P_{BE} = \Pr(\underline{\hat{U}} \neq \underline{U})$

Lemma 1.2.1. $P_{BE} \geq P_{LE} \geq \frac{P_{BE}}{k}$

Proof. See homework 1. ■

To analyze, we focus on the set of correctly encoded sequences. $G = \{\underline{u} : \beta(\alpha(\underline{u})) = \underline{u}\}$

Then we have

$$P_{BE} = 1 - \Pr[U \in G], |G| \leq 2^k, L \geq \lceil \log_2 |G| \rceil .$$

<u>QUESTION</u> How large is the smallest set of sequences with length k form A_U with probability ≈ 1?

We need to use weak law of large numbers (WLLN).

Theorem 1.2.1. *Suppose $A_x = \{1, 2, \ldots, Q\}$ with probability $p_1, \ldots, p_Q$. Given $\underline{u} = (u_1, \ldots, u_k) \in A_U^k$.*

$$n_q(\underline{u}) := \#\textit{times } a_q \textit{ occurs in } \underline{u}, \quad f_q(\underline{u}) = \frac{n_q(\underline{u})}{k} = \textit{frequency}$$

Fix any $\varepsilon > 0$,

$$\Pr[f_q(\underline{u}) \doteq p_q \pm \varepsilon] \to 1 \textit{ as } k \to \infty.$$

Moreover,

$$\Pr[f_q(\underline{u}) \doteq p_q \pm \varepsilon, q = 1, \ldots, Q] \to q \textit{ as } k \to \infty.$$

<u>NOTATION</u> $a \doteq b \pm \varepsilon \iff |a - b| \leq \varepsilon$

Consider subset of A_U^k that corresponds to this event x.

$T_k = \{\underline{u} : f_q(\underline{u}) \doteq p_q \pm \varepsilon, q = 1, \ldots, Q\}$.

$\Pr[\underline{U} = \underline{u}] = p(u_1)p(u_2)\ldots p(u_k)$.

By WLLN, $\Pr(T_k) \to 1$ as $k \to \infty$.

<u>KEY FACT</u> all sequences in T_k have approximately the same probability.

For $\underline{u} \in T_k$,

$$\begin{aligned} p(\underline{u}) &= p(u_1)p(u_2)\dots p(u_k) \\ &= p_1^{n_1(u)} p_2^{n_2(u)} \dots p_k^{n_k(u)} \\ &= p_1^{kf_1(u)} p_2^{kf_2(u)} \dots p_k^{kf_k(u)} \\ &\approx \tilde{p}^k \text{ where } \tilde{p} = p_1^{p_1} p_2^{p_2} \dots p_Q^{p_Q}. \end{aligned}$$

So we have $|T_k| \approx \frac{1}{\tilde{p}^k}$.

Then we have
$$\overline{r} = \frac{\log_2 |T_k|}{k} = -\frac{k \log_2 \tilde{p}}{k} = -\log_2 \tilde{p}.$$

Is that rate good? Can we do better? Can we have a set S with probability ≈ 1 and significantly smaller?

Since $\Pr[\underline{U} \in A_U^k \setminus T_k] \approx 0 \implies \Pr[\underline{U} \in S] \approx \Pr[\underline{U} \in S \cap T_k] \approx \frac{|S|}{|T_k|}$. So when k is large, T_k is the smallest set with large probability. And $R_{AL}^* \approx -\log \tilde{p}$.

How to express $\tilde{p}$.

$$\begin{aligned} -\log \tilde{p} &= -\log \prod_{i=1}^{Q} p_i^{p_i} \\ &= -\sum_{i=1}^{Q} p_i \log p_i =: \text{entropy} = H. \end{aligned}$$

Some properties of H:

1. its unit is bits
2. $H \geq 0$.
3. $H = 0 \implies \iff p_q = 1$ for some q.
4. $H \leq \log_2 Q$.
5. $H = \log_2 Q \iff p_q = \frac{1}{Q}$ for all q.