Unit 12: Channel coding-1

(Channel Capacity)

$\boxed{A}$ Channel coding and repetition code

Perhaps information theory has been most successful in providing a framework for designing error correcting codes. (We will highlight the application in communication systems in Unit 14 addendum)

In this problem, we encode a message $m \in \{1, \dots, M\}$ as a codeword $\underline{x} \in \mathcal{X}^n$, which is transmitted over n uses of the channel. The message is decoded as $\hat{U}$ using the channel output. We want to ensure that the avege probability of error to be small, below ε, i.e.,

$$\frac{1}{M}\sum_{m=1}^{M} \mathbb{P}(\hat{U} \neq m) \leq \varepsilon.$$

How many messages can be sent per channel use?

Throughout, we focus on memoryless channels $W: \underset{\text{input alphabet}}{\mathcal{X}} \to \underset{\text{output alphabet}}{\mathcal{Y}}$, that is

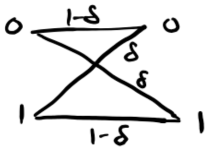
$$W^n(\underline{y}|\underline{x}) = \prod_{t=1}^{n} W(y_t|x_t),$$

where W^n denotes n uses of the channel.

Three channels have received special attention:

1) Binary symmetric channel

BSC(δ)

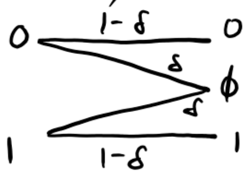


$\mathcal{X} = \mathcal{Y} = \{0, 1\}$

2) Binary erasure channel BEC(δ)

$\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1,\phi\}$

erasure



3) Additive Gaussian Noise channel

$$\mathcal{X} = \mathcal{Y} = \mathbb{R}, \quad W(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right).$$

Let us try to send a bit reliably over BSC(δ) and BEC(δ). In both cases, we simply repeatedly send the bit $b \in \{0,1\}$ over the channel n times. This code is called the repetition code.

$*$ For BEC(δ), the bit will be received correctly in n uses if there is no erasure. Thus, the prob. of error $= \delta^n \le \varepsilon$ when $n \ge \left(\log 1/\varepsilon\right) / \log \frac{1}{\delta}$. Therefore, the number of bits per channel use

$$= \frac{\log 1/\delta}{\log 1/\varepsilon} \to 0 \quad \text{as } \varepsilon \to 0.$$

$*$ For BSC(δ), we first design the decoder. When b is sent n times, the output sequence $y^n \in \{0,1\}^n$ has probability $(1-\delta)^{\sum_{i=1}^n \mathbb{1}_{\{y_i = b\}}} \delta^{\sum_{i=1}^n \mathbb{1}_{\{y_i = \bar{b}\}}}$. We use the maximum likelihood decoding rule: $P(\underline{y} | B=1) \underset{0}{\overset{1}{\gtrless}} P(\underline{y} | B=0)$, which is the same as

$$(1-\delta)^k \delta^{n-k} \underset{0}{\overset{1}{\gtrless}} (1-\delta)^{n-k} \delta^k,$$

where k = no. of 1's in $\underline{y}$. Thus, we declare 1 if

$$\left(\frac{1-\delta}{\delta}\right)^k \geq \left(\frac{1-\delta}{\delta}\right)^{n-k} \iff k \log \frac{1-\delta}{\delta} \geq (n-k) \log \frac{1-\delta}{\delta}.$$

$\boxed{\text{For } \delta < \frac{1}{2}}$ Declare 1 if $k \geq n-k$ (majority rule).

The probability of error for this rule is less than

$$\mathbb{P}\left(\sum_{t=1}^{n} Z_t \geq \frac{n}{2}\right) \qquad \text{where } Z_1, \ldots, Z_n \sim \text{iid Ber}(\delta).$$

$$\leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^{n} Z_t - \delta\right| > \frac{1}{2} - \delta\right)$$

$$\leq e^{-2n\left(\frac{1}{2}-\delta\right)^2}$$

(this is a Chernoff bound; this specific form is called Hoeffding's inequality)

$$\leq \varepsilon$$

if $n \geq \dfrac{1}{2\left(\frac{1}{2}-\delta\right)^2} \ln \dfrac{1}{\varepsilon}$. Thus, number of bits sent per channel use with probability of error less than ε is

$$\frac{2\left(\frac{1}{2}-\delta\right)^2}{\ln 1/\varepsilon} \longrightarrow 0 \quad \text{as } \varepsilon \to 0.$$

Thus, for both cases above, for vanishing probability of error, we cannot send positive number of bits per channel use.

B Channel capacity and Shannon's channel coding theorem

Can we find a scheme that sends messages at positive rate with vanishing probability of error?

Let us formulate this question more carefully.

* Channel code A channel code of length n and size M is given by a pair of mappings (e, d) with the encoder $e : \{1, \dots, M\} \to \mathcal{X}^n$ and the decoder $d : \mathcal{Y}^n \to \{1, \dots, M\}$. The prob. of error for this code is given by

$$\frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\left(d(Y^n) \neq m \,\middle|\, m \text{ sent} \right)$$

↳ output of the channel when the input is $e(m)$

$$= \frac{1}{M} \sum_{m=1}^{M} \sum_{\underline{y} : d(\underline{y}) \neq m} W^n(\underline{y} \mid e(m)).$$

(This notion of prob. of error is the average prob. of error. We may also consider the maximum prob. of error, but roughly the same answer will be attained)

The rate of this code is given by $\frac{\log M}{n}$.

Denote by $M(n, \varepsilon)$ the maximum size of a code of length n and prob. of error less than ε.

* Channel Capacity

A rate $R>0$ is an ε-achievable rate if $M(n,\varepsilon) \geq 2^{nR}$ for all n suff. large. Denote by $C_\varepsilon(W)$ the supremum (maximum) over all achievable rates. The capacity of a channel W is given by $C(W) = \lim_{\varepsilon \to 0} C_\varepsilon(W)$.

Roughly speaking, $C(W) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log M(n,\varepsilon)$.

* Shannon's Channel Capacity Theorem

For a channel $W: \mathcal{X} \to \mathcal{Y}$ with discrete $\mathcal{X}$ and $\mathcal{Y}$,

$$C(W) = \max_{P_X} I(P_X; W)$$

mutual info. b/w input and output of a channel

We will prove this result in the next unit.

C Duality bounds and examples

Earlier we saw the following "information radius" formula:

$$\underbrace{\max_{P_X} I(P_X;W)}_{\text{capacity formula}} = \underbrace{\min_{Q} \max_{x} D(W_x \| Q)}_{\text{information radius}}$$

Thus, for every input distribution P on $\mathcal{X}$ and output distribution Q on $\mathcal{Y}$, $I(P;W) \leq C(W) \leq \max_x D(W_x \| Q)$

We use these bounds to evaluate capacity for some channels.

1) Binary symmetric channel: $W \equiv BSC(\delta)$

Using $P_X = \text{unif}\{0,1\}$, $I(X \wedge Y) = H(Y) - H(Y|X) = 1 - h(\delta)$.

Using $Q = \text{unif}\{0,1\}$, $\max\{D(W_0 \| Q), D(W_1 \| Q)\} = 1 - h(\delta)$.

Thus, $C(W) = 1 - h(\delta)$.

2) Binary erasure channel: $W \equiv BEC(\delta)$

Using $P_X = \text{unif}\{0,1\}$,

$$I(X \wedge Y) = H(Y) - H(Y|X) = 1 - \delta$$

Using $Q = \left(\frac{1-\delta}{2}, \delta, \frac{1-\delta}{2}\right)$,

$$\max\{D(W_0 \| Q), D(W_1 \| Q)\} = 1 - \delta$$

Thus, $C(W) = 1 - \delta$.