Lecture 16 Channel Coding theorem for BSC

Textbook 7.5

October 25th, 2022

Outline

Definition

A discrete channel, denoted by $(\mathcal{X}, p(y|x), \mathcal{Y})$, consists of two finite sets $\mathcal{X}$ and $\mathcal{Y}$ and a collection of probability mass functions $p(y|x)$, one for each $x \in \mathcal{X}$, such that for every x and y, $p(y|x) \geq 0$, and for every x, $\sum_x p(y|x) = 1$, with the interpretation that X is the input and Y is the output of the channel.

Definition

The nth extension of the discrete memoryless channel (DMC) is the channel $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$, where

$$p(y_k|x^k, y^{k-1}) = p(y_k|x_k), \ \ k = 1, 2, \ldots, n.$$

Remark

If the channel is used without feedback [i.e., if the input symbols do not depend on the past output symbols, namely, $p(x_k|x^{k-1}, y^{k-1}) = p(x_k|x^{k-1})$*], the channel transition function for the* n*th extension of the discrete memoryless channel reduces to*

$$p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i).$$

When we refer to the discrete memoryless channel, we mean the discrete memoryless channel without feedback unless we state explicitly otherwise.

Definition

An (M, n) code for the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of the following:

1. An index set $\{1, 2, \ldots, M\}$.
2. An encoding function $X^n : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$. The set of codewords is called the cordbook.
3. A decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, M\},$$

which is a deterministic rule that assigns a guess to each possible received vector.

Conditional probability of error

Let

$$\lambda_i = \Pr(g(Y^n) \neq i | X^n = x^n(i)) = \sum_{y^n} p(y^n | x^n(i)) I(g(y^n) \neq i)$$

be the conditional probability of error given that index i was sent, where $I(\cdot)$ is the indicator function.

Definition

The maximal probability of error $\lambda^{(n)}$ for an (M, n) code is defined as

$$\lambda^{(n)} = \max_{i \in \{1,2,\ldots,M\}} \lambda_i.$$

Definition

The (arithmetic) average probability of error $P_e^{(n)}$ for an (M, n) code is defined as

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i.$$

Note that if the index W is chosen according to a uniform distribution over the set $\{1, 2, \ldots, M\}$, and $X^n = x^n(W)$, then

$$P_e^{(n)} = \mathsf{Pr}(W \neq g(Y^n)).$$

Also, obviously,

$$P_e^{(n)} \leq \lambda^{(n)}.$$

Definition

The *rate* of an (M, n) code is

$$R = \frac{\log M}{n} \text{ bits per transmission.}$$

Definition

A rate R is said to be *achievable* if there exists a sequence of $(\lceil 2^{nR} \rceil,\ n)$ codes such that $\lambda^{(n)}$ tends to 0 as $n \to \infty$.

Definition

The *capacity* of a channel is the supremum of all achievable rates.

Outline

If **x** and **y** are two tuples of 0s and 1s, then we shall say that their *Hamming-distance* is

$$d(\mathbf{x}, \mathbf{y}) := |\{i | 1 \leq i \leq n, x_i \neq y_i\}|.$$

Maximum-likelihood-decoding

If $\mathbf{y}$ is received we try to find a codeword $\mathbf{x}$ such that $d(\mathbf{x}, \mathbf{y})$ is minimal.

Outline

Suppose that we use a code C consisting of M words of length n, each word occurring with equal probability. If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ are the codewords and we use maximum-likelihood-decoding, let P_i be the probability of making an incorrect decoding of a received word is:

$$p_C := M^{-1} \sum_{i=1}^{M} P_i.$$

Now consider all possible codes C with the given parameters and define:

$$P^*(M, n, p) := \text{minimal value of } P_C.$$

Theorem

If $0 < R < 1 + p\log p + q\log q$ *and* $M_n := 2^{\lfloor Rn \rfloor}$ *then* $P^*(M, n, p) \to 0$ *if* $n \to 0$.

The probability of an error pattern with w errors is $p^w q^{n-w}$, i.e., it depends on w only.

The number of errors in a received word is a random variable with expected value np and variance $np(1-p)$. If $b := (np(1-p)/(\epsilon/2))^{1/2}$, then by Chebyshev's inequality we have

$$P(w > np + b) \leq \frac{1}{2}\epsilon.$$

Since $p < \frac{1}{2}$, the number $\rho := \lfloor np + b \rfloor$ is less than $\frac{1}{2}n$ for sufficiently large n.

Let $B_\rho(\mathbf{x})$ be the set of words $\mathbf{y}$ with $d(\mathbf{x},\mathbf{y}) \leq \rho$. Then

$$|B_\rho(\mathbf{x})| = \sum_{i \leq \rho} \binom{n}{i} < \frac{1}{2} n \binom{n}{p} \leq \frac{1}{2} \cdot \frac{n^n}{\rho^\rho (n-\rho)^{n-\rho}}$$

The set $B_\rho(\mathbf{x})$ is usually called the *sphere* with radius ρ and center $\mathbf{x}$.

We shall use the following estimates:

$$\frac{\rho}{n}\log\frac{\rho}{n}=\frac{1}{n}\lfloor np+b\rfloor\log\frac{\lfloor np+b\rfloor}{n}=p\log p+O(n^{-1/2}),$$

$$(1-\frac{\rho}{n})\log(1-\frac{\rho}{n})=q\log q+O(n^{-1/2}),\ \ (n\to\infty).$$

Let $\mathbf{u} \in \{0,1\}^n, \ \mathbf{v} \in \{0,1\}^n$. Then

$$f(\mathbf{u},\mathbf{v}) := \begin{cases} 0, & \text{if } d(\mathbf{u},\mathbf{v}) > \rho \\ 1, & \text{if } d(\mathbf{u},\mathbf{v}) \leq \rho. \end{cases}$$

If $\mathbf{x}_i \in C$ and $\mathbf{y} \in \{0,1\}^n$ then

$$g_i(\mathbf{y}) := 1 - f(\mathbf{y},\mathbf{x}_i) + \sum_{j \neq i} f(\mathbf{y},\mathbf{x}_j).$$

Note that if $\mathbf{x}_i$ is the only codeword such that $d(\mathbf{x}_i,\mathbf{y}) \leq \rho$, then $g_i(\mathbf{y}) = 0$ and that otherwise $g_i(\mathbf{y}) \geq 1$.

Proof of Shannon's Theorem

We shall pick the codewords $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ at random (independently). We decode as follows. If $\mathbf{y}$ is received and if there is exactly one codeword $\mathbf{x}_i$ such that $d(\mathbf{x}_i, \mathbf{y}) \leq \rho$, then decode $\mathbf{y}$ as $\mathbf{x}_i$. Otherwise we declare an error (or if we must decode, then we always decode as $\mathbf{x}_1$).

Let P_i be defined as above. We have

$$\begin{aligned} P_i &= \sum_{\mathbf{y} \in \{0,1\}^n} P(\mathbf{y}|\mathbf{x}_i) g_i(\mathbf{y}) \\ &= \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}_i)\{1 - f(\mathbf{y}, \mathbf{x}_i)\} + \sum_{\mathbf{y}} \sum_{j \neq i} P(\mathbf{y}|\mathbf{x}_i) f(\mathbf{y}, \mathbf{x}_j). \end{aligned}$$

Here the first term on the right-hand side is the probability that the received word $\mathbf{y}$ is not in $B_\rho(\mathbf{x}_i)$. This probability is at most $\frac{1}{2}\epsilon$.

Hence we have

$$P_C \leq \frac{1}{2}\epsilon + M^{-1}\sum_{i=1}^{M}\sum_{\mathbf{y}}\sum_{j\neq i} P(\mathbf{y}|\mathbf{x}_i)f(\mathbf{y},\mathbf{x}_j).$$

The main principle of the proof is the fact that $P^*(M,n,p)$ is less than the expected value of P_C over all possible codes C picked at random. Therefore we have

$$\begin{aligned} P^*(M,n,p) &= \frac{1}{2}\epsilon + M^{-1}\sum_{i=1}^{M}\sum_{\mathbf{y}}\sum_{j\neq i}\mathcal{E}(P(\mathbf{y}|\mathbf{x}_i))\mathcal{E}(f(\mathbf{y},\mathbf{x}_j)) \\ &= \frac{1}{2}\epsilon + M^{-1}\sum_{i=1}^{M}\sum_{\mathbf{y}}\sum_{j\neq i}\mathcal{E}(P(\mathbf{y}|\mathbf{x}_i))\cdot\frac{|B_\rho|}{2^n} \\ &= \frac{1}{2}\epsilon + (M-1)2^{-n}|B_\rho|. \end{aligned}$$

So we have that

$$n^{-1}\log(P^*(M,n,p)-\frac{1}{2}\epsilon) \leq n^{-1}\log M-(1+p\log p+q\log q)+O(n^{-1/2}).$$

Substituting $M = M_n$ on the right-hand side we find, using the restriction on R,

$$n^{-1}\log(P^*(M,n,p) - \frac{1}{2}\epsilon) < -\beta < 0.$$

for $n > n_0$, i.e., $P^*(M,n,p) < \frac{1}{2}\epsilon + 2^{-\beta n}$.