Homework 3

Due: 13:10, 11/14, 2024 (in class)

Homework Policy: (Read before You Start to Work)

- Copying from other students' solution is not allowed. If caught, all involved students get 0 point on that particular homework. Caught twice, you will be asked to drop the course.
- Collaboration is welcome. You can work together with **at most one partner** on the homework problems which you find difficult. However, you should write down your own solution, not just copying from your partner's.
- Your partner should be the same for the entire homework.
- Put your collaborator's name beside the problems that you collaborate on.
- When citing known results from the assigned references, be as clear as possible.

1. (Linear coding achieves the BSC capacity) [14]

In the lecture, we show that for the channel coding problem over a binary symmetric channel $\mathsf{BSC}(p)$, $\forall\,\delta > 0$ and $\forall\,\epsilon \in (0,1)$, there exists a codebook $\mathcal{C}$ of size 2^k such that

$$k > n\left(\mathsf{d_b}\left(p \,\middle\|\, \tfrac{1}{2}\right) - \delta\right) \quad \text{and} \quad \mathsf{P}^{(n)}_{\mathsf{e},\mathrm{ML}} \le \epsilon.$$

The key to the proof is a random coding argument, where the random ensemble given in the lecture is such that

$$X_i(w) \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(1/2) \quad \forall\, i = 1,2,\ldots,n,\ \forall\, w = 1,2,\ldots,2^k,$$

and as a result, a key inequality is established as follows:

$$\Pr\left\{\mathsf{w}(\boldsymbol{X}(w) \oplus \boldsymbol{Z} \oplus \boldsymbol{X}(\tilde{w})) \le n(p+\varepsilon)\right\} \le 2^{-n\mathsf{d_b}\left(p+\varepsilon \,\middle\|\, \frac{1}{2}\right)} \tag{1}$$

In this problem, let us consider an alternative random ensemble that comprises *linear codes*, that is, the codeword of a message bit vector $w \equiv \boldsymbol{b} = \begin{bmatrix} b_1 & b_2 & \ldots & b_k \end{bmatrix}$ is

$$\boldsymbol{x}(\boldsymbol{b}) = \boldsymbol{b}\mathbf{g},$$

for some binary matrix $\mathbf{g} \in \{0,1\}^{k\times n}$, and the above arithmetic is in the binary field. In other words, the encoding function is a linear transform governed by the matrix $\mathbf{g}$, and the codebook $\mathcal{C}_{\mathbf{g}} = \left\{ \boldsymbol{b}\mathbf{g} \,\middle|\, \boldsymbol{b} \in \{0,1\}^k \right\}$ can be viewed as a subspace in the vector space $\{0,1\}^n$ with k dimensions. The subscript associated to $\mathcal{C}$ is to emphasize its dependency with the *generator* matrix $\mathbf{g}$.

The alternative random ensemble is $\mathcal{C}_{\mathbf{G}}$, where the matrix $\mathbf{G}$ is random with

$$\mathbf{G}_{j,i} \overset{\text{i.i.d.}}{\sim} \text{Ber}(1/2) \quad \forall\, j = 1,2,\ldots,k,\ \forall\, i = 1,2,\ldots,n.$$

a) Show that the random codewords in $\mathcal{C}_{\mathbf{G}}$ are pairwise independent, that is,

$$\boldsymbol{X}(\boldsymbol{b}) \perp\!\!\!\perp \boldsymbol{X}(\tilde{\boldsymbol{b}}) \quad \forall\, \boldsymbol{b} \neq \tilde{\boldsymbol{b}}. \qquad [6]$$

b) Show that the random codewords in $\mathcal{C}_{\mathbf{G}}$ are NOT mutually independent. [2]

c) Show that (1) holds and conclude that linear coding achieves the BSC capacity. [6]

2. (Data processing) [10]

Recall the data processing inequality for information divergence in L2:

$$\mathrm{D}\left(\mathsf{P}_X \| \mathsf{Q}_X\right) \geq \mathrm{D}\left(\mathsf{P}_Y \| Q_Y\right) \tag{2}$$

where P_Y and Q_Y are the output distributions of a data processing block $\mathsf{W}_{Y|X}$ when the input X follows P_X and Q_X respectively.

Suppose now for some special $\mathsf{W}_{Y|X}$, the above inequality (2) can be *strengthened* to

$$\mathrm{D}\left(\mathsf{P}_X \| \mathsf{Q}_X\right) \geq \eta\, \mathrm{D}\left(\mathsf{P}_Y \| \mathsf{Q}_Y\right) \quad \forall\, \mathsf{P}_X, \mathsf{Q}_X, \tag{3}$$

where η is some constant and $\eta > 1$.

Based on (3), prove the following *strengthened* data processing inequality for mutual information: if $U - X - Y$ forms a Markov chain and the conditional law of Y given X is $\mathsf{W}_{Y|X}$, then

$$\mathrm{I}\left(U;X\right) \geq \eta\, \mathrm{I}\left(U;Y\right).$$

3. (Capacity of the permutation channel) [12]

A channel model in neural communication is the following:

- Input/ouput alphabet: $\mathcal{X} = \mathcal{Y} = \{0, 1\}^d$
- Channel law:

$$\mathsf{P}_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} 1/\binom{d}{\|\boldsymbol{x}\|_1}, & \text{if } \|\boldsymbol{y}\|_1 = \|\boldsymbol{x}\|_1 \\ 0, & \text{otherwise} \end{cases}$$

Note that for a d-dimensional *binary* vector $\boldsymbol{x}$, its ℓ_1-norm is the number of 1's in $\boldsymbol{x}$:

$$\|\boldsymbol{x}\|_1 = \textstyle\sum_{i=1}^{d} \mathbb{1}\{x_i = 1\}.$$

In words, the channel permutes the length-d binary vector uniformly at random. In this problem, let us compute the capacity of this channel, namely, find

$$\mathrm{C} = \max_{\mathsf{P}_{\boldsymbol{X}}} \mathrm{I}(\boldsymbol{X}; \boldsymbol{Y}).$$

a) (Warm-up) Let $L := \|\boldsymbol{X}\|_1$. Show that $\mathrm{I}(\boldsymbol{X}; \boldsymbol{Y}) = \mathrm{I}(\boldsymbol{X}; \boldsymbol{Y}|L) + \mathrm{H}(L)$. [2]

b) Show that

$$\mathrm{C} = \max_{\mathsf{P}_L} \left\{ \mathrm{H}(L) + \max_{\mathsf{P}_{\boldsymbol{X}|L}} \mathrm{I}(\boldsymbol{X}; \boldsymbol{Y}|L) \right\}$$

and compute the channel capacity C accordingly. What is the capacity achieving input distribution? [6]

c) Let α be a constant between 0 and 1, that is, $0 < \alpha < 1$. Now suppose the channel delivers $\boldsymbol{x}$ noiselessly with probability $(1 - \alpha)$, and permutes $\boldsymbol{x}$ uniformly at random with probability α (note: keeping $\boldsymbol{x}$ the same is also one possible permutation.).

Compute the channel capacity C of this channel. What is the capacity achieving input distribution? [4]

4. (List codes) [14]

In this problem, let us consider a variant of the channel coding problem over a DMC $\mathsf{P}_{Y|X}$ with Shannon capacity C.

Recall that in the formulation of channel coding, the decoder aims to uniquely decode the message. In practice, such an aim might be too stringent, and the decoder may just want to determine a *list* of plausible codewords from the channel output.

A $(n, \lceil n\mathrm{R}\rceil, \lceil n\mathrm{L}\rceil)$ *list code* consists of

- An encoding function that maps each message w to a length-n codeword $x^n(w)$, for each $w \in \{1, ..., 2^{\lceil n\mathrm{R}\rceil}\}$.
- A decoding function that maps a channel output length-n sequence y^n to a list of messages $\mathcal{L}(y^n) \subseteq \{1, ..., 2^{\lceil n\mathrm{R}\rceil}\}$ of size $|\mathcal{L}| \le 2^{\lceil n\mathrm{L}\rceil}$. An error occurs if the transmitted message is not contained in this list.

Hence, the definition of the *probability of error* becomes

$$\mathsf{P}_{\mathsf{e}}^{(n)} := \mathsf{Pr}\{W \notin \mathcal{L}(Y^n)\}.$$

A tuple (R, L) is said to be *achievable* if there exists a sequence of $(n, \lceil n\mathrm{R}\rceil, \lceil n\mathrm{L}\rceil)$ list codes with

$$\lim_{n\to\infty} \mathsf{P}_{\mathsf{e}}^{(n)} = 0.$$

a) Since the definition of the probability of error is changed, to prove the converse part of the coding theorem, a new Fano-type inequality is needed.

Consider a random variable $U \in \mathcal{U} = \{1, 2, \ldots, m\}$ and a random set $\mathcal{V} \subseteq \mathcal{U}$. For example, if $\mathcal{V} \sim \mathrm{Unif}(2^{\mathcal{U}})$ where $2^{\mathcal{U}}$ denotes the collection of all subsets of $\mathcal{U}$, then $\mathrm{H}(\mathcal{V}) = \log_2(2^{|\mathcal{U}|}) = |\mathcal{U}| = m$.

Suppoes $|\mathcal{V}| \le v$ with probability 1. Let $\mathsf{P}_{\mathsf{e}} := \mathsf{Pr}\{U \notin \mathcal{V}\}$. Show that

$$\mathrm{H}(U|\mathcal{V}) \le \mathsf{H}_{\mathsf{b}}(\mathsf{P}_{\mathsf{e}}) + \mathsf{P}_{\mathsf{e}} \log(m) + (1 - \mathsf{P}_{\mathsf{e}}) \log(v). \quad [4]$$

b) Show that for every sequence of $(n, \lceil n\mathrm{R}\rceil, \lceil n\mathrm{L}\rceil)$ list codes with vanishing $\mathsf{P}_{\mathsf{e}}^{(n)}$ as $n \to \infty$, (R, L) must satisfy

$$\mathrm{R} - \mathrm{L} \le \mathrm{C}. \quad [4]$$

c) Show that if $\mathrm{R} - \mathrm{L} < \mathrm{C}$, then (R, L) is achievable. [6]

Hint: You may design the encoder so that roughly $2^{n(\mathrm{R}-\mathrm{L})}$ distinct codewords need to be transmitted reliably over the channel.