

LECTURE 8

Last time:

- Elias codes
- Slepian-Wolf
- Compression: pulling it together

Lecture outline

- Channel capacity
- Binary symmetric channels
- Erasure channels
- Maximizing capacity

Reading: Reading: Scts. 8.1-8.3.

Channel capacity

We describe a channel by a set of transition probabilities

Assume a discrete input alphabet $\mathcal{X}$ and a discrete output alphabet $\mathcal{Y}$

The transition probabilities are:

$$P_{\underline{Y}^n|\underline{X}^n}(\underline{y}^n|\underline{x}^n)$$

for all n

Let us first restrict ourselves to discrete memoryless channels (DMCs), for which

$$P_{\underline{Y}^n|\underline{X}^n}(\underline{y}^n|\underline{x}^n) = \prod_{i=1}^n P_{Y|X}(y[i]|x[i])$$

The capacity of a DMC channel is defined as

$$C = \max_{P_X(x)} I(X; Y)$$

We'll see later why this capacity is actually achievable when we study the coding theorem

Channel capacity

Channel capacity for a DMC is also for any $n > 1$

$$\frac{1}{n} \max_{P_{\underline{X}^n}(\underline{x}^n)} I(\underline{X}^n; \underline{Y}^n)$$

Indeed,

$$\begin{aligned} & \max_{P_{\underline{X}^n}(\underline{x}^n)} I(\underline{X}^n; \underline{Y}^n) \\ &= H(\underline{Y}^n) - H(\underline{Y}^n | \underline{X}^n) \\ &= \sum_{i=1}^n H(Y[i] | \underline{Y}^{i-1}) - \sum_{i=1}^n H(Y[i] | X[i]) \\ &\leq \sum_{i=1}^n H(Y[i]) - \sum_{i=1}^n H(Y[i] | X[i]) \\ &= \sum_{i=1}^n I(X[i]; Y[i]) \end{aligned}$$

Channel capacity

The inequality can be met with equality if we take the X s to be independent, because the Y s then are also independent

Moreover, by taking the X s to be IID, then we can maximize the last RHS if we select the PMF of X that maximizes each term of the sum

Thus, capacity of a DMC is the maximum average mutual information

Binary Symmetric Channel (BSC)

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{x=0,1} P_X(x) H(Y|X = x) \\ &= 1 - \mathbf{H}(\epsilon) \end{aligned}$$

where $\mathbf{H}(\epsilon) = -(\epsilon \log(\epsilon) + (1 - \epsilon) \log(1 - \epsilon))$

Binary Erasure Channel (BEC)

E indicator variable that is 1 if there is an error and is 0 otherwise

$$\begin{aligned} C &= \max_{P_X(x)} I(X; Y) \\ &= \max_{P_X(x)} (H(Y) - H(Y|X)) \\ &= \max_{P_X(x)} (H(Y, E) - H(Y|X)) \\ &= \max_{P_X(x)} (H(E) + H(Y|E) - H(Y|X)) \end{aligned}$$

Binary Erasure Channel (BEC)

$$H(E) = H(\epsilon)$$

$$\begin{aligned} & H(Y|E) \\ &= P(E = 0)H(Y|E = 0) + P(E = 1)H(Y|E = 1) \\ &= (1 - \epsilon)H(X) \end{aligned}$$

$$H(Y|X) = H(\epsilon)$$

$$\text{Thus } C = \max_{P_X(x)} (H(Y|E)) = 1 - \epsilon$$

Symmetric channels

Let us consider the transition matrix T the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix whose elements are $P_{Y|X}(y|x)$

A DMC is symmetric iff the set of outputs can be partitioned into subsets such that for all subsets the matrix T (using inputs as rows and outputs as columns) has the property that within each partition the rows are permutations of each other and the columns are permutations of each other

Maximizing capacity

A set of necessary and sufficient conditions on an input PMF to achieve capacity over a DMC is that

$$\begin{aligned} I(X = k; Y) &= C \text{ for all } k \text{ s.t. } P_X(k) > 0 \\ I(X = k; Y) &\leq C \text{ for all } k \text{ s.t. } P_X(k) = 0 \end{aligned}$$

where

$$\begin{aligned} &I(X = k; Y) \\ &= \sum_{j \in \mathcal{Y}} P_{Y|X}(j|k) \log \left(\frac{P_{Y|X}(j|k)}{\sum_{i \in \mathcal{X}} P_X(i) P_{Y|X}(j|i)} \right) \end{aligned}$$

C is the capacity of the channel

Maximizing capacity

We use the following theorem for the proof:

Let $f(\underline{\alpha})$ be a concave function of $\underline{\alpha} = (\alpha_1, \dots, \alpha_m)$ over the region R when $\underline{\alpha}$ is a probability vector. Assume that the partial derivatives $\frac{\partial f(\underline{\alpha})}{\partial \alpha_k}$ are defined and continuous over R with the possible exception that $\lim_{\alpha_k \rightarrow 0} \frac{\partial f(\underline{\alpha})}{\partial \alpha_k}$ is allowed to be $+\infty$. Then a necessary and sufficient condition on $\underline{\alpha}$ to maximize f over R is that:

$$\frac{\partial f(\underline{\alpha})}{\partial \alpha_k} = \lambda \quad \forall k \text{ s. t. } \alpha_k > 0$$

$$\frac{\partial f(\underline{\alpha})}{\partial \alpha_k} \leq \lambda \text{ for all other } \alpha_k$$

Maximizing capacity

$$\begin{aligned} & I(X; Y) \\ = & \sum_{k \in \mathcal{X}, j \in \mathcal{Y}} P_X(k) P_{Y|X}(j|k) \\ & \log \left(\frac{P_{Y|X}(j|k)}{\sum_{i \in \mathcal{X}} P_X(i) P_{Y|X}(j|i)} \right) \end{aligned}$$

thus

$$\frac{\partial I(X; Y)}{\partial P_X(k)} = I(X = k; Y) - \log(e) = \lambda$$

Note that

$$I(X; Y) = \sum_{k \in \mathcal{X} \text{ s.t. } P_X(k) > 0} P_X(k) I(X = k; Y)$$

$$\text{so } C = \sum_{k \in \mathcal{X} \text{ s.t. } P_X(k) > 0} P_X(k) I(X = k; Y)$$

since all the $I(X = k; Y)$ are the same and their convex combination is C , each of them individually is C

Symmetric DMC capacity

Consider BSC

If we take the inputs to be equiprobable, then we maximize $H(Y)$

$H(Y|X = k)$ is the same for $k = 0, 1$

For symmetric DMC, capacity is achieved by having equiprobable inputs

Check: for symmetric DMC selecting all the inputs equiprobable ensures that every output has the same probability

all of the $H(Y|X = k)$ are equal

since we have the same set of transition probability values, albeit over different outputs

Symmetric DMC capacity

Another view:

$$\begin{aligned} I(X = k; Y) \\ = \sum_{j \in \mathcal{Y}} P_{Y|X}(j|k) \log \left(\frac{P_{Y|X}(j|k)}{\frac{1}{|\mathcal{X}|} \sum_{i \in \mathcal{X}} P_{Y|X}(j|i)} \right) \end{aligned}$$

within a partition of outputs, each column of the T matrix is a permutation of each other column

thus, all $I(X = k; Y)$ are the same

DMC capacity

What about channels that are not symmetric?

All inputs may not be useful, but all outputs are

For any capacity-achieving input probability, the output probabilities are all positive (for reachable outputs)

For some $P_X(k) = 0$ (which must exist for there to be 0 probability outputs)

$$\sum_{j \in \mathcal{Y}} P_{Y|X}(j|k) \log \left(\frac{P_{Y|X}(j|k)}{P_Y(j)} \right) \leq C$$

but if there is $P_Y(j) = 0$, then LHS is ∞

DMC capacity

The output probability vector that achieves capacity is unique and all input probability vectors that give rise to the output vector yield capacity.

Because of concavity of mutual information in input probability, if two input probabilities were capacity-achieving, then any convex combination would also be capacity-achieving

Take a random variable Z to represent the value of θ

For all values of Z , $I(X; Y)$ are the same (corresponding to convex combinations of optimal input distributions), so Y and Z are independent, so the PMF of Y does not depend on Z and is unique.

DMC capacity

Let m be the smallest number of inputs that can be used with non-zero probability to achieve capacity and let $\mathcal{X}$ be such a set of inputs. Then, for output alphabet $\mathcal{Y}$, $m \leq |\mathcal{Y}|$ and the input probability to achieve capacity using inputs in $\mathcal{X}$ only is unique.

Proceed by contradiction. The output probabilities can be expressed as the solution of the $|\mathcal{Y}|$ equations

$$\sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x) = P_Y(y) \forall y \in \mathcal{Y}$$

(how about $\sum_{x \in \mathcal{X}} P_X(x) = 1$? Sum over all the y s)

and there exists at least one valid such $P_X(x)$

DMC capacity

If $m > |\mathcal{Y}|$ then there is a homogeneous solution with m

The sum of the elements of this homogeneous solution is 0, hence we can pick a sum of the solution to the equations and of a homogeneous solution (that sum is also a solution) so that one element is 0, which contradicts the fact that m is minimum

The same line of reasoning yields uniqueness. The equations cannot have a homogeneous solution by the above argument, so the solution must be unique.