ECE 231A Discussion 7

TA: Hengjie Yang

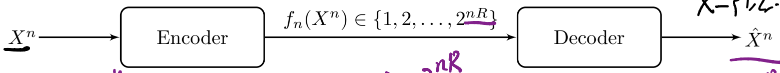
Email: hengjie.yang@ucla.edu

05/15/2020

Motivation and distortion measures

n

$\hat{X} = \{1, 2, \dots 2^R\}$



$|\mathcal{X}|^n$ $\gg 2^{nR}$ 2^{nR}

Source model $\mathcal{X}$: $\mathcal{X}$ with distribution $p(x)$ producing i.i.d. sequence X^n.

Distortion measure: $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$. $d(x, \hat{x}), x \in \mathcal{X}, \hat{x} \in \hat{X}$ is the cost of representing symbol x with symbol $\hat{x}$.

Distortion between two sequences:

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

Goal: To find the "minimal representation" of X^n using R bits/per source symbol to guarantee $\mathbb{E}[d(X^n, \hat{X}^n)] \leq D$.

Examples of distortion measures:

$\mathbb{E}[d(x, \hat{x})] = \Pr\{x \neq \hat{x}\}$

1. Hamming distortion: $d(x, \hat{x}) = \mathbb{I}_{\{x \neq \hat{x}\}}$.
2. Squared-error distortion: $d(x, \hat{x}) = (x - \hat{x})^2$.

Rate distortion theorem

Rate distortion theorem: For an i.i.d. source X with distribution $p(x)$ and bounded distortion function $d(x, \hat{x})$,

$$R(D) = R^{(I)}(D) = \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x}) \le D} I(X;\hat{X}).$$

Namely, all rates $R \ge R^{(I)}(D)$ are achievable; Conversely, any $(2^{nR}, n)$ code that achieves distortion D must have $R \ge R^{(I)}(D)$.

$X_{p(x)}$ $\Pr(Y|X)$ max $I(X;Y)$ $X \to Y$

Duality with channel capacity:

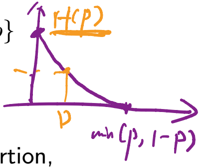
1. In solving for channel capacity, the channel $X \to Y$ is fixed and we are seeking an optimized input distribution;
2. In solving for rate-distortion function, the output X is fixed and we are seeking an optimized channel $\hat{X} \to X$.

$\hat{X} \longrightarrow X_{p(x)}$ min $I(X;\hat{X})$

Calculation of the rate distortion function $R(D)$

Bernoulli source: for a Bernoulli(p) source with Hamming distortion,

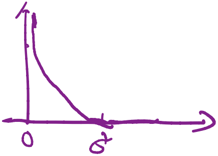
$$R(D) = \begin{cases} H(p) - H(D), & 0 \le D \le \min\{p, 1-p\} \\ 0, & D > \min\{p, 1-p\}. \end{cases}$$



Gaussian source: for a $\mathcal{N}(0, \sigma^2)$ source with squared-error distortion,

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \le D \le \sigma^2 \\ 0, & D > \sigma^2 \end{cases}$$

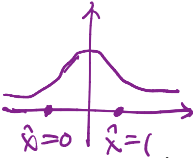
$R = \frac{1}{2}\log\frac{\sigma^2}{D} \Rightarrow D = \left(\frac{1}{4}\right)^R \sigma^2$



Remark:

1. Reducing distortion by considering long blocks: From the above result for Gaussian source, $D(R) = \left(\frac{1}{4}\right)^R \sigma^2$.

$$D(R) = 0.25\sigma^2, \quad \text{if } R = 1, n \to \infty$$
$$D = 0.36\sigma^2, \quad \text{if } R = 1, n = 1$$



Rate distortion for a parallel Gaussian source

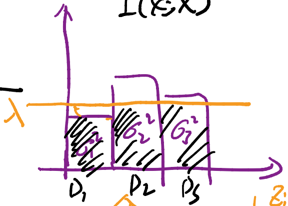
$\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$

Goal: To represent k independent normal random sources $X_1, \ldots, X_k$, $X_i \sim \mathcal{N}(0, \sigma_i^2)$ with squared error distortion using R bits per vector.

Rate distortion for a parallel Gaussian source: Let $X_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, 2, \ldots, k$ be independent Gaussian r.v.'s. Let the distortion measure be $d(x^k, \hat{x}^k) = \sum_{i=1}^{k}(x_i - \hat{x}_i)^2$. The rate distortion function is

$I(X^k; \hat{X}^k)$

$$R(D) = \sum_{i=1}^{k} \frac{1}{2} \log \frac{\sigma_i^2}{D_i}$$



where

$$D_i = \begin{cases} \lambda, & \text{if } \lambda < \sigma_i^2 \\ \sigma_i^2, & \text{if } \lambda \geq \sigma_i^2 \end{cases}$$

where λ is chosen so that $\sum_{i=1}^{k} D_i = D$ (**Reverse water-filling**).

$$\frac{1}{2} \log \frac{\sigma_i^2}{D_i} = \frac{1}{2} \log\left(1 + \frac{\sigma_i^2 - D_i}{D_i}\right)$$

Derivation of reverse water-filling

$$R(D) = \min_{\mathbb{E}[|X_i|^2] \le D} I(X^k; \hat{X}^k)$$

$$I(X^k; \hat{X}^k) = H(X^k) - H(X^k | \hat{X}^k)$$

$$= \sum_{i=1}^{k} H(X_i) - \sum_{i=1}^{k} H(X_i | \hat{X}^k, X^{i-1})$$

$$\ge \sum_{i=1}^{k} H(X_i) - \sum_{i=1}^{k} H(X_i | \hat{X}_i)$$

$$= \sum_{i=1}^{k} I(X_i; \hat{X}_i) \quad \ge \left[\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right]^+$$

$$\ge \sum_{i=1}^{k} \frac{1}{2}\left[\log\frac{\sigma_i^2}{D_i}\right]^+$$

$$\sum_{i=1}^{k} D_i = D$$

$$\min \sum_{i=1}^{k} \frac{1}{2}\log\frac{\sigma_i^2}{D_i}$$

s.t. $\sum_{i=1}^{k} D_i \le D$

$D_i \ge 0$

$$J(D_1 \cdots D_k, \lambda) = \sum \frac{1}{2}\log\frac{\sigma_i^2}{D_i} + \lambda \sum_{i=1}^{k} D_i$$

Converse to the rate distortion theorem

Converse: Any $(2^{nR}, n)$ code that achieves distortion D must have $R \geq R(D)$, where

$$R(D) = \min_{p(\hat{x}|x): \sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D} I(X;\hat{X}).$$

Lemma: $R(D)$ is a nonincreasing, convex function of D.

Outline of the proof of the converse: $f_n : X^n \to \{1, 2, \cdots 2^{nR}\}$

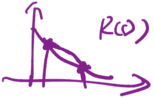
$X^n \to f_n(X^n) \to \hat{X}^n$

$$\begin{aligned}
nR \geq H(f_n(X^n)) &\geq I(X^n; f_n(X^n)) \\
&\geq I(X^n;\hat{X}^n) \\
&= H(X^n) - H(X^n|\hat{X}^n) \\
&\geq \sum_{i=1}^{n} \left(H(X_i) - H(X_i|\hat{X}_i) \right) \\
&\geq \sum_{i=1}^{n} R(\mathbb{E}[d(X_i,\hat{X}_i)]) \\
&\geq n \left(\sum_{i=1}^{n} \frac{1}{n} R(\mathbb{E}[d(X_i,\hat{X}_i)]) \right) \\
&\geq nR(\mathbb{E}[d(X^n;\hat{X}^n)]) \\
&\geq nR(D)
\end{aligned}$$

$- H(f_n(X^n)|X^n)$

$I(X_i;\hat{X}_i)$

$\geq R(\mathbb{E}[d(X_i,\hat{X}_i)])$



$\mathbb{E}[d(X^n;\hat{X}^n)] \leq D$

HW 7, Problem 2: rate distortion for uniform source

Consider a source X uniformly distributed on the set $\{1, 2, \ldots, m\}$. Find the rate distortion function for this source with Hamming distortion; that is,

$$d(x, \hat{x}) = \begin{cases} 0, & \text{if } x = \hat{x}, \\ 1, & \text{if } x \neq \hat{x}. \end{cases}$$

Hint: Fano's inequality.

$X \to Y \to \hat{X}$ $\quad P_e = \Pr\{X \neq \hat{X}\}$

$\underline{H(P_e)} + \underline{P_e \log(|\hat{X}| - 1)} \geq H(X|\hat{X})$

Assume: $D = \Pr\{X \neq \hat{X}\}$

$I(X; \hat{X}) = \underbrace{H(X)}_{\log m} - H(X|\hat{X})$

$H(X|\hat{X}) \leq H(D) + D \log(m-1)$

$\Rightarrow \quad I(X; \hat{X}) \geq \underline{\log m - H(D) - D\log(m-1)}$

$= H_m\left(\frac{1}{m}, \frac{1}{m}, \cdots \frac{1}{m}\right) -$

$H_m\left(1-D, \underbrace{\frac{D}{m-1}, \frac{D}{m-1}, \cdots \frac{D}{m-1}}_{m-1}\right)$

$E = \mathbb{1}\{X \neq \hat{X}\}$

$H(X, E | \hat{X}) = H(X|\hat{X}) + \underbrace{\cancel{H(E|\hat{X}, X)}}_{=0}$

$= H(E|\hat{X}) + H(X|\hat{X}, E)$

$\Rightarrow H(X|\hat{X}) = \underline{H(E|\hat{X})} + H(X|\hat{X}, E)$

$\leq \underline{H(E)} + \Pr(E=1) \underline{H(X|\hat{X}, E=1)}$

$E \perp\!\!\!\perp \hat{X}$ $\quad \leq H(P_e) + \underline{P_e \log(m-1)}$

$\Pr(X|\hat{X}, X \neq \hat{X}) = \frac{1}{m-1}$

choose $p(\hat{x}) = \frac{1}{m}$ $\hat{x} = 1, 2, \cdots m$

$\Rightarrow \frac{1}{m} = 1-D \Rightarrow D = 1 - \frac{1}{m}$

$\Rightarrow D \geq 1 - \frac{1}{m}$ $I(X;\hat{X}) = 0$

$\Rightarrow D < 1 - \frac{1}{m}$ $I(X;\hat{X}) = \log m - H(D) - D \log(m-1)$

$$p(x|\hat{x}) = \begin{cases} 1-D & \text{if } x = \hat{x} \\ \frac{D}{m-1} & \text{if } x \neq \hat{x} \end{cases}$$

$p(\hat{x}) = \frac{1}{m}$

if $D < 1 - \frac{1}{m}$

$$\begin{array}{cc} & \hat{x} \\ & \begin{array}{cccc} 1 & 2 & \cdots & m \end{array} \\ X \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \end{array} & \begin{bmatrix} \frac{1}{m}(1-D) & & & \\ \frac{1}{m} \cdot \frac{D}{m-1} & \ddots & & \\ \vdots & & \ddots & \\ \frac{1}{m} \cdot \frac{D}{m-1} & & & \frac{1}{m}(1-D) \end{bmatrix} \end{array}$$

$E[d(x,\hat{x})]$

$\parallel$

$Pr(x \neq \hat{x}) = D$

if $D < 1 - \frac{1}{m}$

HW 7, Problem 4: rate distortion function with infinite distortion

Find the rate distortion function $R(D) = \min I(X;\hat{X})$ for $X \sim \text{Bern}(1/2)$ and distortion

$$d(x,\hat{x}) = \begin{cases} 0, & x = \hat{x}, \\ 1, & x = 1, \hat{x} = 0, \\ \infty, & x = 0, \hat{x} = 1. \end{cases}$$

$p(\hat{x}|x)$

$p(x=0, \hat{x}=1) = 0$

$I(X;\hat{X}) = H(X) - H(X|\hat{X})$

$$p(x,\hat{x}) = \begin{array}{cc} & \hat{x} \\ x & \begin{bmatrix} 1/2 & 0 \\ D & 1/2 - D \end{bmatrix} \end{array} \Rightarrow \begin{array}{l} p(x|\hat{x}=1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ p(x|\hat{x}=0) = \begin{bmatrix} \frac{1/2}{\frac{1}{2}+D} \\ \frac{D}{\frac{1}{2}+D} \end{bmatrix} \end{array}$$

Assume $D = \Pr\{x=1, \hat{x}=0\}$

$I(X;\hat{X}) = H(\frac{1}{2}) - \Pr(\hat{x}=0)\, H(X|\hat{X}=0)$

$\Pr(\hat{x}=0) = \frac{1}{2}+D$

$$R(D) = \begin{cases} H(\frac{1}{2}) - \Pr(\hat{x}=0)\, H(X|\hat{X}=0) & \text{if } 0 \le D \le \frac{1}{2} \\ 0 & D > \frac{1}{2} \end{cases}$$

$\hat{x} = 0 \quad \Pr(X \neq \hat{X})$

$= \Pr(X \neq 0) = \frac{1}{2} \quad \text{if } D > \frac{1}{2}$