Lecture 35: Introduction to Lossy Compression

Instructor: Himanshu Tyagi
Scribe: Ganesh Ramachandra Kini
09-Nov-2015

Contents

1 Formal description 2
2 Basic example 3
3 A scheme: 4
4 Lower bound (Proof sketch): 4

Rate-distortion problem (Lossy Compression problem) is about finding the rate of a code which recovers the source with some distortion.

Goal: \( \hat{X}^{(n)} \) must not be very different from \( X^n \).
"distance" between \( X^n \) and \( \hat{X}^{(n)} \) is small, with probability 1.

"Distance" needs to be defined.

E.g.1 If \( X_i \in \{0, 1\} \), then we want the number of correctly reproduced bits \( \hat{X}_i^{(n)} \) to be at least \( \geq h(1 - \delta) \)

E.g.2 If \( X_i \in \mathbb{R} \), we want
\[
\sum_{i=1}^{n} (X_i - \hat{X}_i^{(n)})^2 = \|X^n - \hat{X}^{(n)}\|_2^2 \leq n\sigma^2
\]

Applications:

- Data quantization
- Classification/Clustering
1 Formal description

Given a DMS $X^n = X_1, \ldots, X_n$ with a common distribution $P \in \mathcal{P}(\mathcal{X})$, where $\mathcal{X}$ is finite, we want to construct lossy source codes/Rate distortion codes.

The “distortion criterion” is described in terms of a map

\[ d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+ \]
\[ d : x, xt \mapsto d(x, xt) \geq 0 \]

Assume: $\max_{x, xt} d(x, xt) = d_{\text{max}} < \infty$

Note that this assumption is not valid for the ”squared loss” criterion.

We allow the recovery of $x \in \mathcal{X}^n$ as $y \in \mathcal{Y}^n$ as long as $x$ and $y$ are ”close”.

\[ d : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+ \]

\[ X^n \xrightarrow{f(.)} \xrightarrow{\varphi(.)} Y^n \]

Rate-distortion code

Accept $y$ if,

\[ \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) < \delta \]

(The above is the average distortion criterion)

**Definition:** (Rate-distortion code)

Given $0 < \epsilon < 1$ and $0 \leq \delta \leq d_{\text{max}}$, a rate-distortion code for a DMS $\mathcal{X}^n$ (with reproduction alphabet $\mathcal{Y}$) of length $n$ and size $2^k$ consists of

an encoder

\[ f : \mathcal{X}^n \rightarrow \{0, 1\}^k \]

and a decoder

\[ \varphi : \{0, 1\}^k \rightarrow \mathcal{Y}^n \]

such that one of the following holds:

- Max-distortion criterion: (with large probability)

\[ \mathbb{P}\left( \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) > \delta \right) \leq \epsilon \]

where $Y_n = \varphi(f(X^n))$ (not IID)
• Average distortion criterion: (more relaxed than the max-distortion criterion)

\[ E \left[ \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) \right] \leq \delta \]

where \( Y_n = \varphi(f(X^n)) \)

Note that rate of this code = \( \frac{k}{n} \)

**Definition:** (Achievable rate)
A rate \( R > 0 \) is an \((\epsilon, \delta)\)-achievable max-distortion rate if for \( \eta > 0 \) there exists a rate distortion code of rate \( \leq R + \eta \) and with max-distortion \( < \delta \) with probability of error \( \leq \epsilon \), for all \( n \) large.

\[
R_{\epsilon}^{\max}(\delta) = \inf \{ R > 0 | R \text{ is } (\epsilon, \delta) - \text{achievable max-distortion rate} \}
\]

\[
R^{\max}(\delta) = \sup_{0 < \epsilon < 1} R_{\epsilon}^{\max}(\delta)
\]

\[
= \lim_{\epsilon \to 0} R_{\epsilon}^{\max}(\delta)
\]

2 **Basic example**

\( X_i \sim Ber(\frac{1}{2}) \), \( X = \{0, 1\} \), \( Y = \{0, 1\} \)

\[ d(x, y) = \mathbb{I}(x \neq y) \]

Then,

\[
\sum_{i=1}^{n} d(x_i, y_i) = d_H(x, y) \leq n\delta
\]
We are looking for the "minimal covering".

Remark: A rate-distortion code is a "covering" of a large probability subset of $\mathcal{X}^n$; the size of the code $\equiv$ size of the covering

3 A scheme:

Choose $\{c_1, \ldots, c_N\} \subseteq \mathcal{X}^n$

upon observing $x \in \mathcal{X}^n$, store it as $c_j$ where

$$j = \arg \min_{1 \leq i \leq N} \sum_{t=1}^{n} d(x_t, c_{i,t})$$

4 Lower bound (Proof sketch):

Let $A$ be the set of sequences that are recovered within allowed distortion:

$$A = \{x \mid y = \varphi(f(x)) \text{ satisfies } \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) \leq \delta\}$$
Then,

\[ P^n(A) \geq 1 - \epsilon \]

\[ \Rightarrow |A| \geq 2^{n(H(P) - \eta)} = 2^{n(1 - \eta)}, \text{ for all } n \text{ large} \quad (1) \]

\[ A \subseteq \bigcup_{i=1}^{2^n R} B_{n\delta}(y_i) \]

\[ |A| \leq 2^n R |B_{n\delta}(0)| \]

\[ \leq 2^n R 2^{n h(\delta)} \quad (2) \]

By Eq. (1) and Eq. (2),

\[ R \geq 1 - \eta - h(\delta) \]

for large \( n \),

\[ \Rightarrow R_{\epsilon}^{\max}(\delta) \geq 1 - h(\delta) \]