Rate Distortion Theory

Introduction

The description of an arbitrary real number requires an infinite number of bits, so a finite representation of a continuous random variable can never be perfect.

Define the “goodness” of a representation of a source: distortion measure which is a measure of distance between the random variable and its representation.

The basic problem in rate distortion theory can then be stated as follows:

Given a **source distribution** and a **distortion measure**, what is the **minimum expected distortion** achievable at a particular rate?

Or, equivalently, what is the **minimum rate description** required to achieve a particular distortion?

Quantization

In this section we motivate the elegant theory of rate distortion by showing how complicated it is to solve the quantization problem exactly for a single random variable.

Since a continuous random source requires infinite precision to represent exactly, we cannot reproduce it exactly using a finite-rate code.

We first consider the problem of representing a single sample from the source.

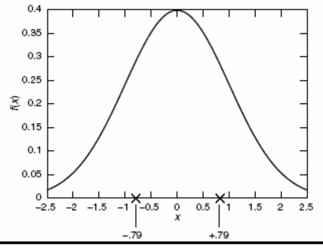
Let the random variable be represented be *X* and let the representation of *X* be denoted as ˆ*X(X)*. If we are given *R* bits to represent *X*, the function ˆ*X* can take on 2^R values. The problem is to find the optimum set of values for ˆ*X* (called the ***reproduction points*** or ***code points***) and the **regions** that are associated with each value ˆ*X*.

Example: Gaussian Distribution

For example, let *X* ~ *N(*0*,* σ^2*)*, and assume a squared-error distortion measure. In this case we wish to find the function ˆ*X(X)* such that ˆ*X* takes on at most 2^R values and minimizes *E(X* − ˆ*X(X))*^{2}.

If we are given one bit to represent *X*, it is clear that the bit should distinguish whether or not *X* > 0.

To minimize squared error, each reproduced symbol should be the mean of its region.



Example

Thus,

$$\hat{X}(x) = \begin{cases} \sqrt{\dfrac{2}{\pi}}\sigma & x \geq 0 \\ -\sqrt{\dfrac{2}{\pi}}\sigma & x < 0 \end{cases}$$

If we are given 2 bits to represent the sample, the situation is not as simple.

Clearly, we want to divide the real line into four regions and use a point within each region to represent the sample.

Optimal Regions and Reconstruction Points

But it is no longer immediately obvious what the representation regions and the reconstruction points should be. We can, however, state two simple properties of optimal regions and reconstruction points for the quantization of a single random variable:

- Given a set {ˆ*X(w)*} of reconstruction points, the distortion is minimized by mapping a source random variable *X* to the representation ˆ*X(w)* that is closest to it. The set of regions of χ defined by this mapping is called a *Voronoi* or *Dirichlet partition* defined by the reconstruction points.

- The reconstruction points should minimize the conditional expected distortion over their respective assignment regions.

The “Good” Quantizer

These two properties enable us to construct a simple algorithm to find a “good” quantizer:

1. We start with a set of reconstruction points, find the optimal set of reconstruction regions (which are the nearest-neighbor regions with respect to the distortion measure);
2. Then find the optimal reconstruction points for these regions (the centroids of these regions if the distortion is squared error),
3. Then repeat the iteration for this new set of reconstruction points.

The expected distortion is decreased at each stage in the algorithm, so the algorithm will converge to a local minimum of the distortion. This algorithm is called the *Lloyd algorithm*

Peculiarities

Instead of quantizing a single random variable, let us assume that we are given a set of *n* i.i.d. random variables drawn according to a Gaussian distribution.

These random variables are to be represented using *n*R bits.

Since the source is i.i.d., the symbols are independent, and it may appear that the representation of each element is an independent problem to be treated separately.

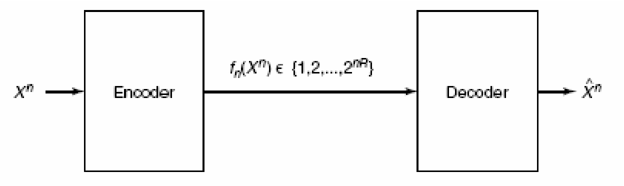
But this is not true. We will represent the entire sequence by a single index taking 2^{nR} values. This treatment of entire sequences at once achieves a lower distortion for the same rate than independent quantization of the individual samples.

Encoder and Decoder

Assume that we have a source that produces a sequence $X_1, X_2, \ldots, X_n$ i.i.d. ~ *p(x), x* ∈ *X*.

We assume that the alphabet is finite, but most of the proofs can be extended to continuous random variables.

The encoder describes the source sequence X_n by an index $f_n(X^n) \in \{1, 2, \ldots, 2^{nR}\}$. The decoder represents X^n by an estimate $\hat{}X^n \in \hat{}X$, as illustrated in Figure



Distortion Function

Definition A *distortion function* or *distortion measure* is a mapping

$d : X \times \hat{}X \rightarrow R^+$

from the set of source alphabet-reproduction alphabet pairs into the set of nonnegative real numbers. The distortion *d(x, ˆx)* is a measure of the cost of representing the symbol *x* by the symbol *ˆx*.

Bounded Distortion

Definition A distortion measure is said to be *bounded* if the maximum value of the distortion is finite:

$$d_{\max} \overset{def}{=} \max_{x\in\chi,\hat{x}\in\hat{X}} d(x,\hat{x}) < \infty$$

In most cases, the reproduction alphabet ˆ*X* is the same as the source alphabet *X*.

Hamming and Square Error Distortion

• *Hamming (probability of error) distortion.* The Hamming distortion is given by

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

which results in a probability of error distortion, since Ed*(X, ˆX)* = Pr*(X ≠ ˆX).*

• *Squared-error distortion.* The squared-error distortion,

$$d(x,\hat{x}) = (x-\hat{x})^2$$

is the most popular distortion measure used for continuous alphabets. Its advantages are its simplicity and its relationship to least-squares prediction.

Distortion between Sequences

Definition The *distortion between sequences* x^n and ˆx^n is defined by

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

So the distortion for a sequence is the average of the per symbol distortion of the elements of the sequence.

Rate Distortion Code

Definition A ***(2^nR^, n)-rate distortion code*** consists of an encoding function,

$f_n : X^n \to \{1, 2, \ldots, 2^{nR}\}$,

and a decoding (reproduction) function,

$g_n : \{1, 2, \ldots, 2^{nR}\} \to$ ˆX^n.

The distortion associated with the *(*2^{nR}*, n)* code is defined as

$$D = Ed(X^n, g_n(f_n(X^n)))$$

where the expectation is with respect to the probability distribution on X:

Some Terms

The set of n-tuples $g_n(1), g_n(2), \ldots, g_n(2^{nR})$, denoted by $\hat{X}^n(1), \ldots, \hat{X}^n(2^{nR})$, constitutes the ***codebook***, and $f^{-1}_n(1), \ldots, f^{-1}_n(2^{nR})$ are the associated ***assignment regions.***

Many terms are used to describe the replacement of X^n by its quantized version $\hat{X}^n(w)$.

It is common to refer to $\hat{X}^n$ as the ***vector quantization***, ***reproduction***, ***reconstruction***, ***representation***, ***source code***, or ***estimate*** of X^n.

Distortion-Related Definitions

Definition A rate distortion pair (R,D) is said to be ***achievable*** if there exists a sequence of $(2^{nR}, n)$-rate distortion codes (f_n, g_n) with:

$$\lim_{n\to\infty} Ed(X^n, g_n(f_n(X^n))) \le D$$

Definition The ***rate distortion region*** for a source is the closure of the set of achievable rate distortion pairs (R,D).

Definition The ***rate distortion function*** $R(D)$ is the infimum of rates R such that (R,D) is in the rate distortion region of the source for a given distortion D.

Definition The ***distortion rate function*** $D(R)$ is the infimum of all distortions D such that (R,D) is in the rate distortion region of the source for a given rate R.