EE5585 Data Compression March 5, 2013

Lecture 12

Instructor: Arya Mazumdar *Scribe: Shreshtha Shukla*

1 Review

In the last class we studied the fundamental theorem of source coding, where R(D) is the optimal rate of compression, given the normalized average distortion D. Our task was to find this optimal rate, for which we had the theorem stated as

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

We also proved $R(D) \geq \min I(X,\hat{X})$, known as the converse part of the theorem. The direct part of the theorem known as achievability result is proved using Random Coding method. Also recall, For Bernoulli(p) Random Variables: $R(D) = h(p) - h(D)$ where $h(\cdot)$ is the binary entropy function, when $D = 0, R(D) = h(p)$ that is, the source can be compressed to $h(p)$ bits.

2 Continuous Random Variables:

Suppose we have a source that produce the i.i.d. sequence

$$X_1, X_2, X_3 \ldots X_n$$

these can be real or complex numbers.

2.1 1 Bit Representation:

For example: we have gaussian random variable, $X \sim \mathcal{N}(0,\sigma^2)$.

Its pdf is then:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$$

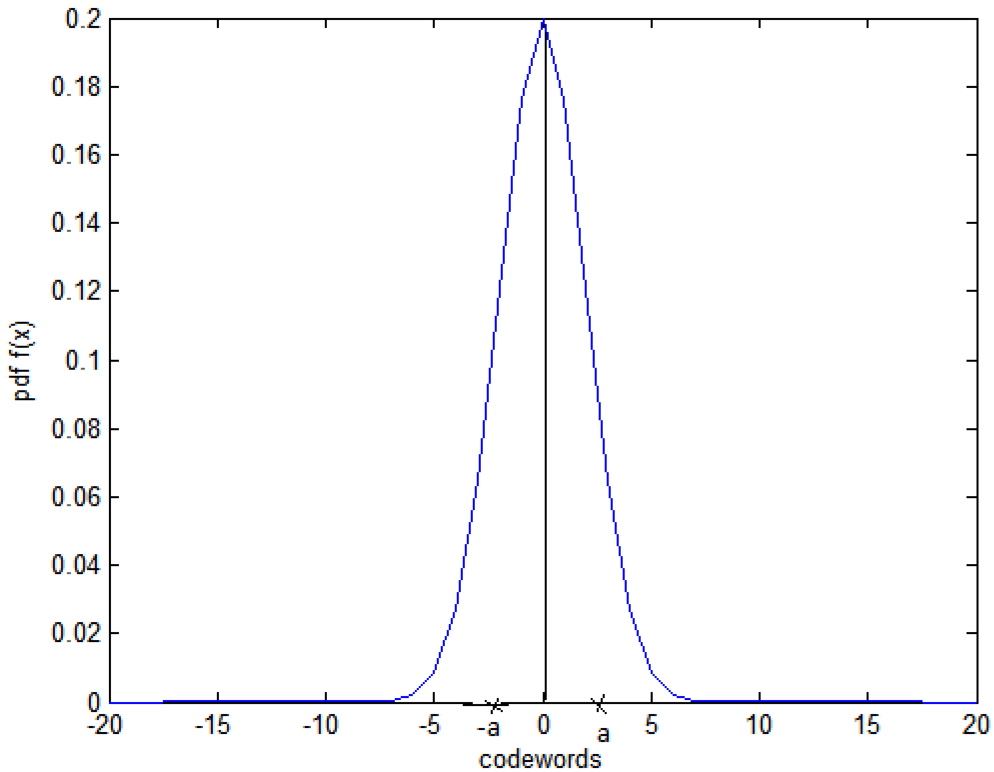

Figure 1: gaussian pdf with 1 bit partition

Now, how can you represent this using 1 bit? With 1 bit, at most we can get the information whether $x \geq 0$ or $x < 0$.

After compressing the file to 1 bit, the next question is: how do we decode it?

For this, some criteria has to be selected, example, we set our codewords based on MSE (mean square error) i.e find the value of 'a' that will minimize the expected squared error

$$\min\ E[(X - \hat{X})^2]$$

$$= \int_{-\infty}^{0} f(x)(x+a)^2 dx \quad + \quad \int_{0}^{\infty} f(x)(x-a)^2 dx$$

$$= 2\int_{0}^{\infty} f(x)(x-a)^2 dx$$

$$= 2\,[\int_{0}^{\infty} x^2 f(x)dx \;-\; 2a\int_{0}^{\infty} x f(x)dx \;+\; a^2\int_{0}^{\infty} f(x)dx\,]$$

$$= (a^2 + \sigma^2) \;-\; 4a\int_{0}^{\infty} x\,\frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}dx$$

Let $y = x^2/2\sigma^2$,
$\Rightarrow dy = \frac{xdx}{\sigma^2}$.
So,

$$-4a \int_0^\infty x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} dx \;=\; -2a\sqrt{2\sigma^2/\pi} \int_0^\infty e^{-y} dy$$

$$= -2a\sqrt{2\sigma^2/\pi}$$

hence, $E[(X - \hat{X})^2] \;=\; a^2 \;+\; \sigma^2 \;-\; 2a\sqrt{2\sigma^2/\pi}$
to minimize this, differentiate with respect to a and set it to 0.

$$\Longrightarrow 2a - 2\sqrt{2\sigma^2/\pi} = 0$$

$$or \quad a = \sigma\sqrt{\frac{2}{\pi}}$$

which is what we choose our codeword.

2.2 2 Bit Representation:

Similarly, for a 2 bit representation,we need to divide the entire region into 4 parts:

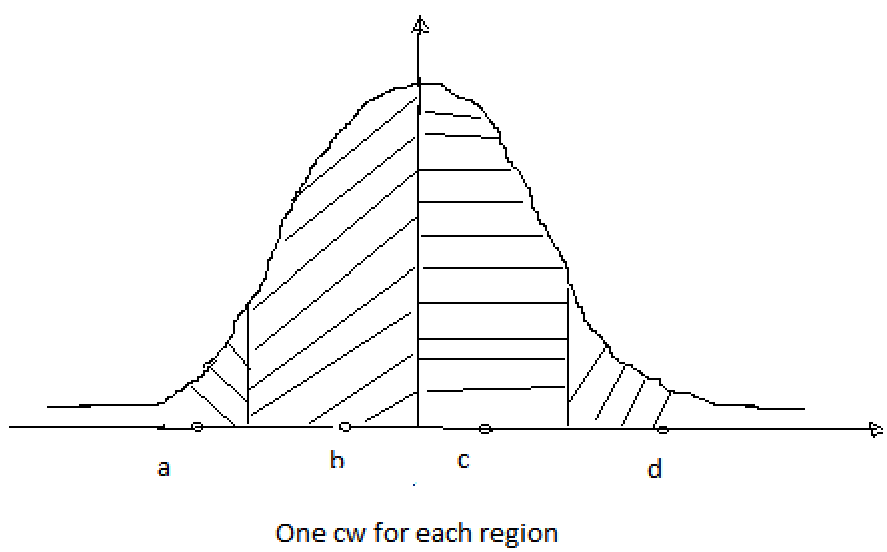

Figure 2: Divided into four parts

Further, it is always good to take a vector instead of a single random variable (i.e a scalar RV). The vector then, lives in an n-dimensional space. In this case also, we need to find the appropriate regions and their associated optimal reconstruction points.

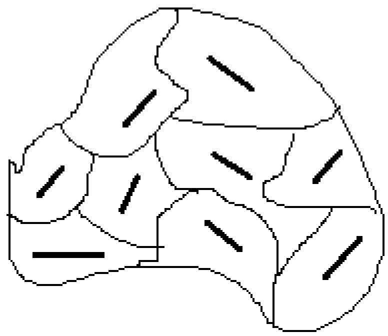

Figure 3: Voronoi Regions in R^nspace

These regions are called the VORONOI Regions and the partitions are known as DIRICHLET'S Partitions.To find the optimal partitions and associated centers, there is an algorithm known as the Llyod's Algorithm.

Briefly the Llyod's algorithm:

We start with some initial set of quantized points, Eg : for 10 bit; $2^{10} = 1024$ points and then find the Voronoi regions for these points. The expected distribution is then optimized. Update the points to those optimal points and again find the Voronoi regions for them. Doing this iteratively converges to optimal values.This algorithm has several names, such as Llyod's Algorithm also known as Expectation Maximization Algorithm. (we know the optimal values using the rate distortion theory so can compare to the convergence result). More detailed study on this will be done in later classes.

3 Entropy for continuous Random Variables

For discrete RVs we have:

$$H(x) = -\sum_{x \in \chi} p(x) \log p(x)$$

similarly for continuous RVs : instead of the pmf we have the pdf $f_X(x)$

Eg, $\mathcal{X} = \mathbf{R}$; $\quad X \sim \mathcal{N}(0, \sigma^2)$

We define Differential Entropy of a continuous random variable as

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \ln(f_X(x)) dx$$

All the other expressions of conditional entropy, mutual information can be written analogously.

Conditional Entropy: $H(X|Y) = \int_{-\infty}^{\infty} f(x,y) \ln f(x|y) dxdy$.

Mutual Information: $I(X;Y) = H(X) - H(X|Y)$

Also we can have the similar Rate Distortion Theorem for the continous case.

3.1 Examples:

Example 1: Entropy of a gaussian RV:

$X \sim \mathcal{N}(0, \sigma^2)$ $\qquad f(x) = 1/\sqrt{2\pi\sigma^2} e^{-x^2/2\sigma^2}$

then

$$H(X) = -\int_{-\infty}^{\infty} 1/\sqrt{2\pi\sigma^2}e^{-x^2/2\sigma^2} \log_e(1/\sqrt{2\pi\sigma^2}e^{-x^2/2\sigma^2})dx$$

$$= -\int_{-\infty}^{\infty} f(x)(\log_e(1/\sqrt{2\pi\sigma^2}) \; - x^2/2\sigma^2)dx$$

$$= -[\log_e(1/\sqrt{2\pi\sigma^2}) - \; 1/2\sigma^2 \int_{-\infty}^{\infty} x^2 f(x)dx]$$

$$= \; -[\log_e(1/\sqrt{2\pi\sigma^2}) \; - \; 1/2\sigma^2\sigma^2]$$

$$= 1/2(\log_e 2\pi\sigma^2) \; + \; 1/2$$

$$= \frac{1}{2}\ln 2\pi e\sigma^2.$$

i.e., $H(X) = \frac{1}{2}\log(2\pi e\sigma^2)$ bits. which is the differential entropy of a gaussian random variable.

3.2 Theorem

CLAIM: A gaussian Random Variable $X \sim \mathcal{N}(0, \sigma^2)$ maximizes the differential entropy among all continuous random variables that have variance of σ^2.

PROOF:
Suppose Z is any random variable with $var(Z) = \sigma^2$ and pdf = g(z),
then $H(X) - H(Z)$ can be written as

$$= -\int f(x)\ln(f(x))dx \; + \quad \int g(x)\ln(g(x))dx$$

$$= -\int f(x)(\ln(1/\sqrt{2\pi\sigma^2}) \; - \; x^2/2\sigma^2)dx \quad + \int g(x)\ln(g(x))dx$$

$$= -\int g(x)\ln(f(x))dx \; + \quad \int g(x)\ln(g(x))dx$$

$$= \int g(x)\ln(g(x)/f(x))dx$$

$$= D(f||g)$$

$$\geqslant 0$$

Thus $H(X) - H(Z) \geqslant 0$, i.e., H(X)⩾H(Z).
This method of proof, is called the Maximum Entropy Method.

4 Rate Distortion for gaussian Random Variables:

Suppose we have a gaussian source X : $X_1, X_2X_3...X_n$ are iid.

To compress this data, according to the Rate distortion Theorem:

$$R(D) = \min_{f(\hat{x}|x)} I(X;\hat{X}) \qquad s.t. \int_x \int_{\hat{x}} f(\hat{x}|x)f(x)d(x,\hat{x}) \leq D$$

where $d(\hat{x}, x)$ is the euclidean distance. (Proof for this is similar to the discrete case.)

To evaluate, R(D),

Step1: Find a lower bound for $I(X;\hat{X})$

Step2: Find some f(x) that achieves the bound and hence is optimal.

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

now, $H(X/\hat{X}) = H(X - \hat{X}|\hat{X})$

$$= \frac{1}{2}\ln 2\pi e\sigma^2 - H(X - \hat{X}|\hat{X})$$

as conditioning always reduces entropy,

$$\geq \frac{1}{2}\ln 2\pi e\sigma^2 - H(X - \hat{X})$$

and $var(X - \hat{X}) = E[(X - \hat{X})^2]$

$$\geq \frac{1}{2}\ln 2\pi e\sigma^2 - H(N(0, E[(X - \hat{X})^2]))$$

$$\geq \frac{1}{2}\ln 2\pi e\sigma^2 - \frac{1}{2}\ln 2\pi eE[(X - \hat{X})^2]$$

$$\geq \frac{1}{2}\ln(\sigma^2/E[(x - \hat{x})^2]$$

and since we know that $E[(X - \hat{X})^2] \leq D$, always

it implies that for Gaussian,

$$R(D) \geq \frac{1}{2}\ln(\sigma^2/D) \qquad for \quad D \leq \sigma^2$$

$$= 0 \qquad\qquad for\, D > \sigma^2$$

After finding a lower bound, we now show that ∃ one f(x) for which this bound is achievable (which is the limit of compression for the gaussian RV)i.e we would like to back track and find how the inequalities meet the equality condition.

Suppose we have,

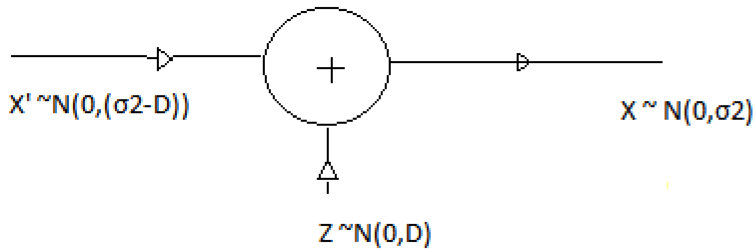

Figure 4: Generating X

where X' is $\hat{X}$

for this case $I(X;\hat{X})= \frac{1}{2}\ln\frac{\sigma^2}{D}$. i.e it achieves the bound.
Hence if there is a Gaussian source producing iid symbols then to encode a vector of length n with resulting quantization error $\leq nD$,
we need at least : $\frac{n}{2}\ln(\sigma^2/D)$ nats or $\frac{n}{2}\log(\sigma^2/D)$ bits to represent it.