

Overview

In general, we have

Asymptotic Equipartition Property (AEP), theoretical data compression.

From this, we have that data compression is *possible*.

Source Coding / Entropy Coding, where we compress for a known source, and try to compress towards $H(S)$.

Universal Coding / Dictionary Coding, where we do not know / cannot find out¹ the probability distribution, thus the entropy, of our source before-hand, but still try to compress efficiently.

Of these, we then have

Lossless compression, where the compressed data can be reconstructed perfectly, and

Lossy compression, where an (acceptable) approximation of the original data can be reconstructed.

¹Due to feasibility, or practicality.

Why AEP?

Q: Isn't Shannon's Source Coding Theorem, as we learned it, enough?

A: Well, yes. But it was originally proven by use of AEP. So, AEP has historical significance. It is also mentioned a lot in Information Theory literature. *And*, it is applied widely, for instance, in physics.

You might also see it later in the course, in a different form. It helps if you are familiar with it already.

What AEP Says

Theorem (AEP)

If X_1, \dots, X_n are iid, then $-\frac{1}{n} \log p(x_1, \dots, x_n) \rightarrow H(X)$ as $n \rightarrow \infty$.

Multiplying both sides by $-n$, and putting both sides as a power of 2 tells us that $p(x_1, \dots, x_n)$ will be close to 2^{-nH} .

What AEP Says

Theorem (AEP)

If X_1, \dots, X_n are iid, then $-\frac{1}{n} \log p(x_1, \dots, x_n) \rightarrow H(X)$ as $n \rightarrow \infty$.

Multiplying both sides by $-n$, and putting both sides as a power of 2 tells us that $p(x_1, \dots, x_n)$ will be close to 2^{-nH} . Fix a small constant, ϵ . Let $A_\epsilon^{(n)}$, the “typical set”, be least set containing sequences (x_1, \dots, x_n) with

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}$$

What AEP Says

Theorem (AEP)

If X_1, \dots, X_n are iid, then $-\frac{1}{n} \log p(x_1, \dots, x_n) \rightarrow H(X)$ as $n \rightarrow \infty$.

Multiplying both sides by $-n$, and putting both sides as a power of 2 tells us that $p(x_1, \dots, x_n)$ will be close to 2^{-nH} . Fix a small constant, ϵ . Let $A_\epsilon^{(n)}$, the “typical set”, be least set containing sequences (x_1, \dots, x_n) with

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}$$

Theorem

- ➊ *If $\mathbf{x} \in A_\epsilon^{(n)}$, then $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, \dots, x_n) \leq H(X) + \epsilon$*
- ➋ *for n sufficiently large, $\Pr(\mathbf{x} \in A_\epsilon^{(n)}) > 1 - \epsilon$*
- ➌ *for n sufficiently large, $(1 - \epsilon)2^{n(H(X)-\epsilon)} \leq |A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$*

So, all elements of the typical set are nearly equiprobable, the typical set has probability near 1, and the number of elements in the typical set is nearly 2^{nH} .

What AEP Says

Theorem (AEP)

If X_1, \dots, X_n are iid, then $-\frac{1}{n} \log p(x_1, \dots, x_n) \rightarrow H(X)$ as $n \rightarrow \infty$.

Multiplying both sides by $-n$, and putting both sides as a power of 2 tells us that $p(x_1, \dots, x_n)$ will be close to 2^{-nH} . Fix a small constant, ϵ . Let $A_\epsilon^{(n)}$, the “typical set”, be least set containing sequences (x_1, \dots, x_n) with

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}$$

Theorem

- ➊ If $\mathbf{x} \in A_\epsilon^{(n)}$, then $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, \dots, x_n) \leq H(X) + \epsilon$
- ➋ for n sufficiently large, $\Pr(\mathbf{x} \in A_\epsilon^{(n)}) > 1 - \epsilon$
- ➌ for n sufficiently large, $(1 - \epsilon)2^{n(H(X)-\epsilon)} \leq |A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$

So, all elements of the typical set are nearly equiprobable, the typical set has probability near 1, and the number of elements in the typical set is nearly 2^{nH} . **Each $\mathbf{x} \in A_\epsilon^{(n)}$ representable with $nH(X)$ bits on average** (indexing). That's a good encoding (divide by $n \dots$).