Source Coding using Typical Sets

The concept of typical sets is useful for designing codes for a source P_X. Instead of encoding the source symbol-by-symbol, we will encode the source symbols in blocks of n symbols at a time.

This strategy can be used to design either a lossy or a lossless code. For a lossy code, we notice that with overwhelming probability, a sequence of n iid samples from P_X is typical, so it suffices to assign binary labels of length (at most) $\lceil n(H(X)+\epsilon)\rceil$ to the elements of $A_\epsilon^{(n)}$, and assign some constant (dummy) codeword to all elements outside of the set. Decoding this dummy codeword will result in an error (data loss), but this error occurs with probability at most ϵ.

The above scheme can be extended to a lossless version by assigning longer labels to the elements outside of $A_\epsilon^{(n)}$, for example binary labels of length $\lceil \log |\mathcal{X}|^n \rceil = \lceil n \log |\mathcal{X}| \rceil$. An extra 'flag' bit is needed to indicate whether the element is inside or outside the typical set. For large enough n, this code is quite efficient:

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. real random variables with respect to the set $\mathcal{X}$, and distributed according to P_X. Let $\epsilon > 0$. Then there exists a lossless code $\mathcal{X}^n \to \{0,1\}^*$ such that, for sufficiently large n, $\mathbb{E}[\frac{1}{n}\ell(X^n)] \leq H(X) + \epsilon$.

Proof

Consider the code described above: the code consist of a flag bit (indicating whether or not the element is inside the typical set), followed by either a short label (for elements in the typical set) or a longer one (for elements outside of it). Let $\epsilon' > 0$ (we will specify the value of ϵ' later). Let n be large enough such that $P[A_{\epsilon'}^{(n)}] > 1 - \epsilon'$ (see the second property of typical sets). Then

$$
\begin{aligned}
\mathbb{E}[\ell(X^n)] &= \sum_{\vec{x}\in\mathcal{X}^n} P_{X^n}(\vec{x})\ell(\vec{x}) \\
&= \sum_{\vec{x}\in A_{\epsilon'}^{(n)}} P_{X^n}(\vec{x})\ell(\vec{x}) + \sum_{\vec{x}\notin A_{\epsilon'}^{(n)}} P_{X^n}(\vec{x})\ell(\vec{x}) \\
&\leq P[A_{\epsilon'}^{(n)}]\cdot(\lceil n(H(X)+\epsilon')\rceil + 1) + P[\overline{A_{\epsilon'}^{(n)}}]\cdot(\lceil n\log|\mathcal{X}|\rceil + 1) \\
&\leq P[A_{\epsilon'}^{(n)}]\cdot(n(H(X)+\epsilon') + 2) + P[\overline{A_{\epsilon'}^{(n)}}]\cdot(n\log|\mathcal{X}| + 2) \\
&\leq n(H(X)+\epsilon') + \epsilon'\cdot n\log|\mathcal{X}| + 2 \\
&= n(H(X)+\epsilon),
\end{aligned}
$$

where $\epsilon = \epsilon' + \epsilon'\log|\mathcal{X}| + \frac{2}{n}$ (note that ϵ can be made arbitrarily small by choosing ϵ' and n wisely). The +1 in the first inequality is a consequence of the `flag' bit.

For large enough blocks of symbols, the flag bit becomes irrelevant. Typical sets thus allow the construction of an efficient code without the 1 bit of overhead that symbol codes may necessarily have. However, this efficiency is only guaranteed for 'sufficiently large n', a rather theoretical condition that may not be achievable in practice.

We conclude this chapter by showing that the typical set is in a sense 'optimal', i.e. that picking a smaller set instead of the typical set does not allow for much shorter codewords on average in a lossy setting, not even if we allow rather large error probabilities by allowing about half of the elements to lie outside of the typical set.

Let $B_\delta^{(n)}$ denote the smallest subset of $\mathcal{X}^n$ such that $P[B_\delta^{(n)}] > 1-\delta$ (for some parameter $\delta > 0$). $B_\delta^{(n)}$ can be explicitly constructed by, for example, ordering $\mathcal{X}^n$ in order of decreasing probability, and adding elements to $B_\delta^{(n)}$ until the probability threshold of $1-\delta$ is reached. The following theorem states that even for large values of δ, we still need almost $nH(X)$ bits to denote an element from $B_\delta^{(n)}$.

Theorem

Let $X_1,\ldots,X_n$ be i.i.d. random variables distributed according to P_X. For any $\delta < \frac{1}{2}$, and any $\delta' > 0$, if $P[B_\delta^{(n)}] > 1-\delta$, then

$$
\frac{1}{n}\log|B_\delta^{(n)}| > H(X) - \delta',
$$

for sufficiently large n.

Proof

Let $\delta, \epsilon < \frac{1}{2}$, and consider some $B_\delta^{(n)}$ such that $P[B_\delta^{(n)}] > 1 - \delta$. By the second property of typical sets, $P[A_\epsilon^{(n)}] > 1 - \epsilon$, for large enough n. Thus, by the union bound,

$$\begin{aligned}
1 - \epsilon - \delta &< 1 - P[\overline{A_\epsilon^{(n)}}] - P[\overline{B_\delta^{(n)}}] \\
&\leq 1 - P[\overline{A_\epsilon^{(n)}} \cup \overline{B_\delta^{(n)}}] \\
&= P[A_\epsilon^{(n)} \cap B_\delta^{(n)}] \\
&= \sum_{\vec{x} \in A_\epsilon^{(n)} \cap B_\delta^{(n)}} P_{X^n}(\vec{x}) \\
&\leq \sum_{\vec{x} \in A_\epsilon^{(n)} \cap B_\delta^{(n)}} 2^{-n(H(X)-\epsilon)} \\
&= |A_\epsilon^{(n)} \cap B_\delta^{(n)}| \cdot 2^{-n(H(X)-\epsilon)} \\
&\leq |B_\delta^{(n)}| \cdot 2^{-n(H(X)-\epsilon)}.
\end{aligned}$$

Rearranging this expression and taking the logarithm, we get

$$H(X) - \epsilon + \frac{1}{n}\log(1 - \epsilon - \delta) < \frac{1}{n}\log|B_\delta^{(n)}|.$$

If we now set $\delta' := \epsilon - \frac{1}{n}\log(1 - \epsilon - \delta)$, then

$$H(X) - \delta' < \frac{1}{n}\log|B_\delta^{(n)}|,$$

as desired. Observe that we can make the expression for δ' as small as desired by choosing a small enough $\epsilon > 0$ and a large enough n, even if δ is rather large.