

Information Theory for Data Science

Last modified: 2026-10-12 02:31:20 (PDT)

Welcome

Information theory measures uncertainty and information: how many bits a message needs, how much one random variable tells us about another, and how fast data can cross a noisy channel. Data science uses its quantities directly: cross-entropy is the loss in logistic regression and neural networks, Kullback–Leibler divergence links maximum likelihood to model fit, and mutual information drives the information gain of decision trees.

These notes collect the information theory that data science courses use. They are new. The pages written so far are:

- [Entropy and mutual information](#): surprisal, joint and conditional entropy, mutual information, and the information gain of decision trees
- [Source coding](#): the source coding theorem, and cross-entropy as the cost of coding with the wrong distribution

The planned topics still to come are:

- Kullback–Leibler divergence and its link to maximum likelihood
- Huffman and arithmetic codes
- channel coding: channel capacity and the noisy-channel coding theorem
- Kolmogorov complexity, and how it relates to Shannon entropy

Entropy, cross-entropy and KL divergence are defined in [Probability for Data Science](#); pages here link to those definitions rather than repeating them. Computational complexity has its own library, [Computational Complexity for Data Science](#).

Using these notes in another site

Course sites link to these pages by URL; they do not include this repository as a git submodule. A host site that keeps a copy of this repository at its root, named `ids`, can still include fragments with paths that start with `ids/`. This site includes its own fragments the same way, through a `ids` symlink that points at the repository root.

Quarto resolves `@id` cross-references only within one rendered page, so a host site that links to a result here uses an explicit link to the page and anchor.