Exercise - Information Theory

Exercises in Data Modeling (190.021)

[1]*dm@ai-lab.science* February 8, 2026

This is a cheat sheet describing the key concepts of this exercise on entropy and discretization of analogue signals.

1 General Information

Note that this is not a full explanation of all topics that are covered in the chapter "Information Theory". It mainly provides tips for the exercise and further explanations of the covered concepts.

2 Entropy Calculation

To calculate the number of needed bits to encode and represent data without losing information, entropy is needed as a measure of the unpredictability of data. First, you need to input a short sentence or sequence of numbers and letters. The input has to have the following format:

- **sentence** = "Put your sentence here"

In the descending code cell, you are supposed to hand over a sequence of symbols, based on these symbols, the entropy will be calculated. Enter the symbol dictionary as follows:

- **symbols** = 'a', 'a', 'a', 'a', 'a'

You can replace 'a' with any letter or number.

3 Signal Discretization and Reconstruction

To use this method in practise we showcase this method for signal discretization and reconstruction. First, the Nyquist-Shannon theorem will be introduced. This theorem describes the ratio between the highest upcoming frequency in a signal and the minimal sampling frequency required to discretize and reconstruct the signal without aliasing. This relationship is described by:

$$fs > 2 * f_{max}$$

The task within the Jupyter notebook is to explore different sample frequencies and to understand when aliasing appeared and which underlying frequencies can be reconstructed, also with a lower sampling frequency. To address this purpose, a signal containing different frequencies (5, 30, 80, & 130 Hz) is given.

Use the slider to adjust the variable f_s, which describes the number of signal samples within one second.

4 Entropy-based Signal Discretization

In this section, entropy-based signal discretization and reconstruction will be discussed and compared with uniform discretization. The objective of this task is to compare the two different methods for signal discretization and their behaviour using a different number of bins. To compare the quality of the discretization and reconstruction of the signal based on the number of bins, define the variable called bin_number with any integer between 2 and 20 and run the descending code cells.