Information Theory

Hadar Tal
hadar.tal@mail.huji.ac.il

This paper is a summary of the educational materials and lectures from

- **Wikipedia**
- **3Blue1Brown** YouTube channel

Winter 2024

Contents

Chapter 1

Definitions

Definition 1.0.1 (Entropy)
The entropy of a discrete random variable X with probability mass function $p(x)$ is defined as

$$H(X) = -\sum_x p(x) \log p(x) \tag{1.1}$$

Definition 1.0.2 (Joint Entropy)
The joint entropy of two discrete random variables X and Y with joint probability mass function $p(x, y)$ is defined as

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y) \tag{1.2}$$

Definition 1.0.3 (Conditional Entropy)
The conditional entropy of a discrete random variable X given another discrete random variable Y is defined as

$$H(X|Y) = -\sum_{x,y} p(x, y) \log p(x|y) \tag{1.3}$$

Definition 1.0.4 (Mutual Information)
The mutual information between two discrete random variables X and Y is defined as

$$I(X; Y) = H(X) - H(X|Y) \tag{1.4}$$

Definition 1.0.5 (Conditional Mutual Information)
The conditional mutual information between two discrete random variables X and Y given a third discrete random variable Z is defined as

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z) \tag{1.5}$$

Definition 1.0.6 (Kullback-Leibler Divergence)
The Kullback-Leibler divergence between two probability distributions p and q is defined as

$$D_{KL}(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \tag{1.6}$$

Definition 1.0.7 (Cross Entropy)
The cross entropy between two probability distributions p and q is defined as

$$H(p, q) = -\sum_x p(x) \log q(x) \tag{1.7}$$

Definition 1.0.8 (Hamming Distance)
The Hamming distance between two binary strings x and y of equal length is defined as

$$d_H(x, y) = \sum_i |x_i - y_i| \tag{1.8}$$