Relative Entropy

1 Why

2 Definition

Consider two distributions on the same finite set. The *entropy* of the first distribution *relative* to the second distribution is the difference of the cross entropy of the first distribution relative to the second and the entropy of the second distribution. We call it the *relative entropy* of the first distribution with the second distribution. People also call the relative entropy the *Kullback-Liebler divergence*.

2.1 Notation

Let A be a non-empty finite set. Let $p : A \to \mathsf{R}$ and $q : A \to \mathsf{R}$ be distributions. Let $H(q, p)$ denote the cross entropy of p relative to q and let $H(q)$ denote the entropy of q. The entropy of p relative to q is

$$H(q, p) - H(p).$$

Herein, we denote the entropy of p relative to q by $d(q, p)$.

3 A similarity function

The relative entropy is a similarity function between distributions.

Proposition 1. *Let q and p be distributions on the same set. Then $d(q, p) \geq 0$ with equality if and only if $p = q$.*

So, d has a few of the properties of a metric. However, d is not a metric; for example, it is not symmetric.

Proposition 2. *There exist distributions $p : A \to \mathsf{R}$ and $q : A \to \mathsf{R}$ (with A a non-empty finite set) such that*

$$d(q, p) \neq d(p, q).$$

3.1 Optimization Perspective

if we want to find a distribution p to

$$\text{minimize} \quad d(q, p)$$

then $p = q$ is a solution.