



**Why**

A distribution/density selector picks a probabilistic model for a dataset by using the maximum likelihood principle. We want to generalize this idea.<sup>1</sup>

**1 Definition**

An *unsupervised data model* is a function  $\ell : \mathbf{R}^d \rightarrow \mathbf{R}$  where  $\ell(x)$  is the *surprise* of the vector  $x$ .<sup>2</sup>

---

<sup>1</sup>Future editions probably include a genetic approach, which moves through clustering and mixtures of Gaussians to illustrate the point more fully.

<sup>2</sup>Future editions will expand.



