



Why

Linear predictors are simple and we know how to select the parameters. The main downside is that there may not be a linear relationship between inputs and outputs.

Definition

A *feature map* for postcepts A is a mapping $\phi : A \rightarrow \mathbf{R}^d$. In this setting, we call $a \in A$ the *raw input record* and we call $\phi(a)$ an *embedding*, *feature embedding* or *feature vector*. We call the components of a feature vector the *features*.

A feature map is *faithful* if, whenever records a_i and a_j are in some sense “similar” in the set A , the embeddings $\phi(a_i)$ and $\phi(a_j)$ are close in the vector space \mathbf{R}^d .

Since it is common for raw input records $a \in A$ to consist of many fields, it is regular to have several feature maps ϕ_i which operate component-wise on the fields of a . These are sometimes called *basis functions*.¹ We concatenate these field feature maps and commonly add a constant feature 1. Since \mathbf{R}^d is a vector space, it is common to refer to it in this case as the *feature space*.

Given a dataset $a = (a^1, \dots, a^n)$ in A and a feature map $\phi : A \rightarrow \mathbf{R}^d$, the *embedded dataset* of a with respect to ϕ is the dataset $(\phi(a^1), \dots, \phi(a^n))$ in \mathbf{R}^d .

¹Future editions will clarify, and perhaps remove.

