



Why

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

Definition

A *feature embedding* 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*. A feature embedding 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 embedding functions ϕ_i which operate component-wise on the fields of a . We concatenate these field embeddings 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*.

