

#### **PREDICTORS**

## Why

We discuss inferring (or learning) functions from examples.

#### **Definitions**

A predictor  $f: \mathcal{U} \to \mathcal{V}$  is a function from  $\mathcal{U}$  to  $\mathcal{V}$ . An inducer is a function from finite datasets in  $\mathcal{U} \times \mathcal{V}$  to predictors from  $\mathcal{U}$  to  $\mathcal{V}$ . A learner is a function family of inducers, indexed by n, each defined for datasets of size n. We call  $\mathcal{U}$  the inputs,  $\mathcal{V}$  the outputs, and f(u) the prediction of f on  $u \in \mathcal{U}$ .

## **Predicting relations**

A relation inducer is a function from finite datasets in  $\mathcal{U} \times \mathcal{V}$  to relations on  $\mathcal{U} \times \mathcal{V}$ . Since we can associate any relation R between  $\mathcal{U}$  and  $\mathcal{V}$  with a function  $f: \mathcal{U} \times \mathcal{V} \to \{0,1\}$ , f(u,v) = 1 if and only if  $(u,v) \in R$ , the predictor case can accommodate learning general relations, beyond functions.

#### Notation

Let D be a dataset of size n in  $\mathcal{U} \times \mathcal{V}$ . Let  $g : \mathcal{U} \to \mathcal{V}$ , a predictor, which makes prediction g(u) on input  $u \in \mathcal{U}$ . Let  $G_n : (\mathcal{U} \times \mathcal{V})^n \to (\mathcal{U} \times \mathcal{V})$  be an inducer, so that  $G_n(D)$  is the predictor which the inductor associates with dataset D. Then  $\{G_n : (\mathcal{U} \times \mathcal{V})^n \to \mathcal{P}((\mathcal{U} \times \mathcal{V}))\}_{n \in \mathbb{N}}$  is a learner.

# Consistent and complete datasets

Let  $D = ((u_i, v_i))_{i=1}^n$  be a dataset and  $R \subset X \times Y$  a relation. D is consistent with R if each  $(u_i, v_i) \in R$ . D is consistent if there exists a relation with which it is consistent. A dataset is always consistent (take  $R = \mathcal{U} \times \mathcal{V}$ ). D is functionally consistent if it is consistent with a function; in this case,  $x_i = x_j \Rightarrow y_i = y_j$ . D is functionally complete if  $\bigcup_i \{x_i\} = X$ . In this case, the dataset includes every element of the relation.

### Other terminology

Other terms for the inputs include independent variables, explanatory variables, precepts, covariates, patterns, instances, or observations. Other terms for the outputs include dependent variables, explained variables, postcepts, targets, outcomes, labels or observational outcomes. An input, output pair is sometimes called a record pair.

Other terms for a learner include learning algorithm, or supervised learning algorithm. Other terms for a predictor include input-output mapping, prediction rule, hypothesis, concept, or classifier.

