



## Why

We consider the set of predictors from which we select.

## Definition

Let  $(X, \mathcal{X}, \mu)$  and  $f : X \rightarrow Y$  be probabilistic data-generation model. A *hypothesis class*  $\mathcal{H} \subset (X \rightarrow Y)$  is a subset of measurable functions. For a dataset  $D \in (X \times Y)^n$ , a *restricted empirical error minimizer* of  $\mathcal{H}$  is a hypothesis  $h \in \mathcal{H}$  with minimal (among elements of  $\mathcal{H}$ ) empirical error on  $D$ .

## Inductive bias

Many authorities call the hypothesis class a *inductive bias* and speak of “biasing” the “learning algorithm”. Since one specifies the hypothesis class prior to the data it often said to “encode prior knowledge about the problem to be learned.”

## Realizable classes

Some hypothesis classes are better than others. For example, a hypothesis class that includes the correct labeling function seems preferable to a class that does not.

We formulate a weaker condition that captures the case when  $f \in \mathcal{H}$ . A hypothesis class  $\mathcal{H}$  is *realizable* if there exists  $h^* \in \mathcal{H}$  with

$$\mu(\{x \in X \mid h^*(x) \neq f(x)\}) = 0.$$

This condition is natural because if  $f \in \mathcal{H}$ , then  $\mathcal{H}$  is realizable. Moreover, the condition has two natural corollaries

First, there exists  $h^*$  so that

$$\mu^n(\{x \in X^n \mid |\{i \in [n] \mid h^*(x_i) \neq f(x_i)\}| \neq 0\}) = 0.$$

We interpret this as follows: “there exists a hypothesis whose probability of achieving nonzero empirical risk is zero.”

Second, denote by  $M_x$  the empirical risk minimizers of  $x \in X^n$ . Then,

$$\mu^n(\{x \in X^n \mid \forall h \in M_x, \forall i, h(x_i) = f(x_i)\}) = 1.$$

We interpret this as follows: “all empirical risk minimizers will achieve zero empirical risk on the dataset, with probability 1.”

Roughly speaking, there exists a hypothesis for which the event that it achieves zero empirical risk has probability one. In this event, every hypothesis in the set of empirical risk minimizers must achieve zero empirical risk. This is a consequence of the hypothesis class being realizable.

