



Why

We use the language of probability distributions to characterize predictions.¹

Definition

A *probabilistic classifier* is a function from precepts to probability distributions on the set of classes. Throughout this sheet, let A be a set of precepts and B a set of postcepts.

Point classifier as probabilistic classifier

Let $f : A \rightarrow B$ be a point classifier. One can always obtain a probabilistic classifier from f in the following way. Define the probabilistic classifier at a precept $a \in A$ to be the distribution that takes 1 on the predicted value of the point classifier at a and 0 otherwise.

Probabilistic classifier from point classifier

Now let $g : A \rightarrow (B \rightarrow [0, 1])$ be a probabilistic classifier. One can always obtain a point classifier from g in the following way. Assign to a precept a the value of the distribution $g(a)$ with the most probability mass. If there are ties, order the (finite) set B arbitrarily, and break ties accordingly. We call this the *maximum likelihood classifier* corresponding to the probabilistic classifier g .

¹Future editions will improve this.

