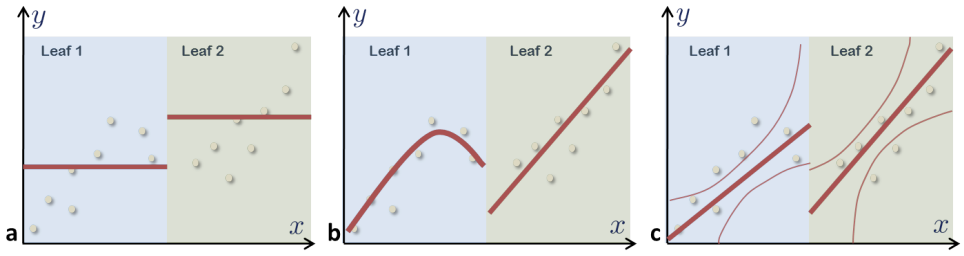

Fig. 4.2: **Example predictor models.** Different possible predictor models. **(a)** Constant. **(b)** Polynomial and linear. **(c)** Probabilistic-linear. The conditional distribution $p(y|x)$ is returned in the latter.

input is represented as a multi-dimensional feature response vector $\mathbf{v} = (x_1, \cdots, x_d) \in \mathbb{R}^d$.

Why regression forests? A regression forest is a collection of randomly trained regression trees (fig. 4.3). Just like in classification it can be shown that a forest generalizes better than a single over-trained tree.

A regression tree (fig. 4.1b) splits a complex nonlinear regression problem into a set of smaller problems which can be more easily handled by simpler models (*e.g.* linear ones; see also fig.4.2). Next we specify the precise nature of each model component.

The prediction model. The first job of a decision tree is to decide which branch to direct the incoming data to. But when the data reaches a terminal node then that leaf needs to make a prediction.

The actual form of the prediction depends on the prediction model. In classification we have used the pre-stored empirical class posterior as model. In regression forests we have a few alternatives, as illustrated in fig. 4.2. For instance we could use a polynomial function of a subspace of the input $\mathbf{v}$. In the low dimensional example in the figure a generic polynomial model corresponds to $y(x) = \sum_{i=0}^{n} w_i x^i$. This simple model also captures the linear and constant models (see fig. 4.2a,b).

In this paper we are interested in output confidence as well as its