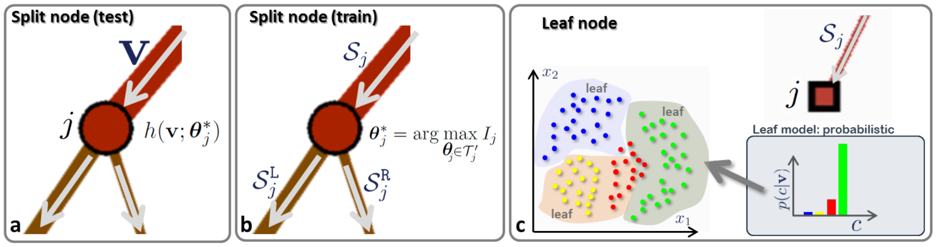


Fig. 2.5: **Split and leaf nodes.** **(a)** Split node (testing). A split node is associated with a weak learner (or split function, or test function). **(b)** Split node (training). Training the parameters $\boldsymbol{\theta}_j$ of node $j$ involves optimizing a chosen objective function (maximizing the information gain $I_j$ in this example). **(c)** A leaf node is associated with a predictor model. For example, in classification we may wish to estimate the conditional $p(c|\mathbf{v})$ with $c \in \{c_k\}$ indicating a class index.

## 2.2 The decision forest model

A random decision forest is an ensemble of randomly trained decision trees. The forest model is characterized by a number of components. For instance, we need to choose a family of split functions (also referred to as "weak learners" for consistency with the literature). Similarly, we must select the type of leaf predictor. The randomness model also has great influence on the workings of the forest. This section discusses each component one at a time.

### 2.2.1 The weak learner model

Each split node $j$ is associated with a binary split function

$$h(\mathbf{v}, \boldsymbol{\theta}_j) \in \{0, 1\}, \tag{2.1}$$

with *e.g.* 0 indicating "false" and 1 indicating "true". The data arriving at the split node is sent to its left or right child node according to the result of the test (see fig.2.5a). The weak learner model is characterized by its parameters $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\tau})$ where $\boldsymbol{\psi}$ defines the geometric primitive used to separate the data (*e.g.* an axis-aligned hyperplane, an oblique hyperplane [43, 58], a general surface *etc.*). The parameter vector $\boldsymbol{\tau}$