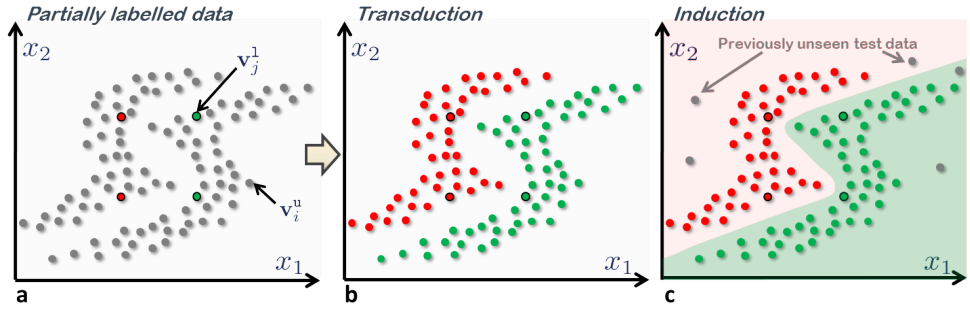

Fig. 7.1: **Semi-supervised forest: input data and problem statement. (a)** Partially labelled input data points in their two-dimensional feature space. Different colours denote different labels. Unlabelled data is shown in grey. **(b)** In transductive learning we wish to propagate the existing ground-truth labels to the many, available unlabelled data points. **(c)** In inductive learning we wish to learn a generic function that can be applied to previously unavailable test points (grey circles). Training a conventional classifier on the labelled data only would produce a sub-optimal classification surface, *i.e.* a vertical line in this case. Decision forests can effectively address both transduction and induction. See text for detail.

The training objective function. As usual, forest training happens by optimizing the parameters of each internal node j via

$$\boldsymbol{\theta}_j^* = \arg\max_{\boldsymbol{\theta}_j \in \mathcal{T}_j} I_j$$

Different trees are trained separately and independently. The main difference with respect to other forests is that here the objective function I_j must encourage both separation of the labelled training data as well as separating different high density regions from one another. This is achieved via the following mixed information gain:

$$I_j = I_j^{\mathrm{u}} + \alpha I_j^{\mathrm{s}}. \tag{7.1}$$

In the equation above I_j^{s} is a supervised term and depends only on the labelled training data. In contrast, I_j^{u} is the unsupervised term and