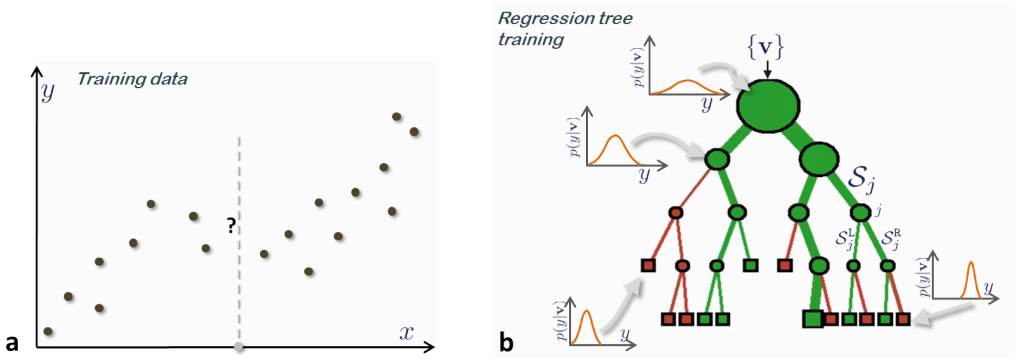

Fig. 4.1: **Regression: training data and tree training. (a)** Training data points are shown as dark circles. The associated ground truth label is denoted by their position along the y coordinate. The input feature space here is one-dimensional in this example ($\mathbf{v} = (x)$). x is the independent input and y is the dependent variable. A previously unseen test input is indicated with a light gray circle. **(b)** A binary regression tree. During training a set of labelled training points $\{\mathbf{v}\}$ is used to optimize the parameters of the tree. In a regression tree the entropy of the continuous densities associated with different nodes decreases (their confidence increases) when going from the root towards the leaves.

Given a labelled training set learn a general mapping which associates previously unseen independent test data with their correct continuous prediction.

Like classification the regression task is inductive, with the main difference being the continuous nature of the output. Figure 4.1a provides an illustrative example of training data and associated continuous ground-truth labels. A previously unseen test input (unavailable during training) is shown as a light grey circle on the x axis.

Formally, given a multi-variate input $\mathbf{v}$ we wish to associate a continuous multi-variate label $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^n$. More generally, we wish to estimate the probability density function $p(\mathbf{y}|\mathbf{v})$. As usual the