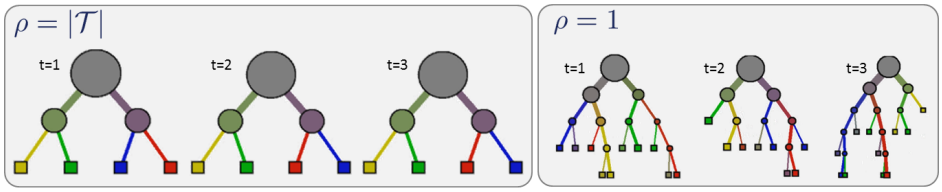

Fig. 2.7: **Controlling the amount of randomness and tree correlation. (a)** Large values of ρ correspond to little randomness and thus large tree correlation. In this case the forest behaves very much as if it was made of a single tree. **(b)** Small values of ρ correspond to large randomness in the training process. Thus the forest component trees are all very different from one another.

2.2.3 The randomness model

A key aspect of decision forests is the fact that its component trees are all randomly different from one another. This leads to de-correlation between the individual tree predictions and, in turn, to improved generalization. Forest randomness also helps achieve high robustness with respect to noisy data.

Randomness is injected into the trees during the training phase. Two of the most popular ways of doing so are:

- random training data set sampling [11] (*e.g.* bagging), and
- randomized node optimization [46].

These two techniques are not mutually exclusive and could be used together. However, in this paper we focus on the second alternative which: i) enables us to train each tree on the totality of training data, and ii) yields margin-maximization properties (details in chapter 3). On the other hand, bagging yields greater training efficiency.

Randomized node optimization. If $\mathcal{T}$ is the entire set of all possible parameters $\boldsymbol{\theta}$ then when training the j^{th} node we only make available a small subset $\mathcal{T}_j \subset \mathcal{T}$ of such values. Thus under the randomness