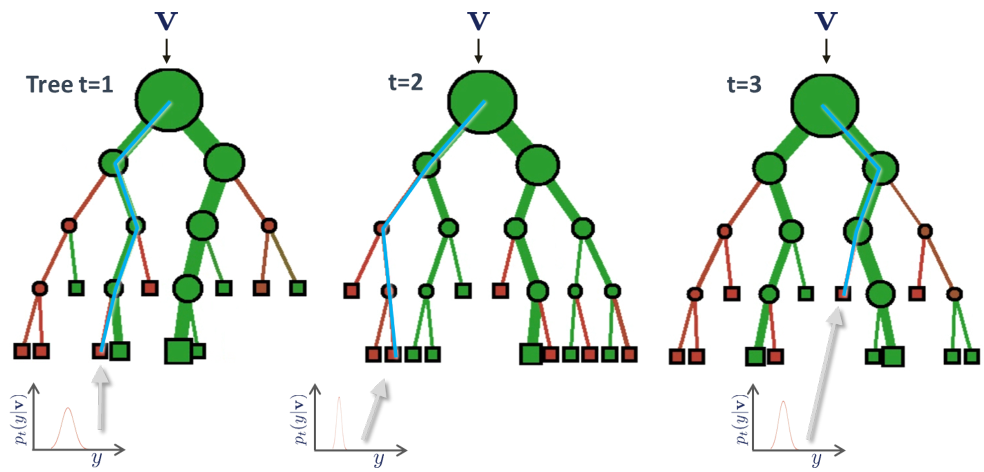

Fig. 4.3: **Regression forest: the ensemble model.** The regression forest posterior is simply the average of all individual tree posteriors $p(\mathbf{y}|\mathbf{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(\mathbf{y}|\mathbf{v})$.

actual value. Thus for prediction we can use a probability density function over the continuous variable $\mathbf{y}$. So, given the t^{th} tree in a forest and an input point $\mathbf{v}$, the associated leaf output takes the form $p_t(\mathbf{y}|\mathbf{v})$. In the low-dimensional example in fig. 4.2c we assume an underlying linear model of type $y = w_0 + w_1 x$ and each leaf yields the conditional $p(y|x)$.

The ensemble model. Just like in classification, the forest output is the average of all tree outputs (fig. 4.3):

$$p(\mathbf{y}|\mathbf{v}) = \frac{1}{T}\sum_{t}^{T} p_t(\mathbf{y}|\mathbf{v})$$

A practical justification for this model was presented in section 2.2.5.

Randomness model. Like in classification here we use a randomized node optimization model. Therefore, the amount of randomness is controlled during training by the parameter $\rho = |\mathcal{T}_j|$. The random subsets of split parameters $\mathcal{T}_j$ can be generated on the fly when training