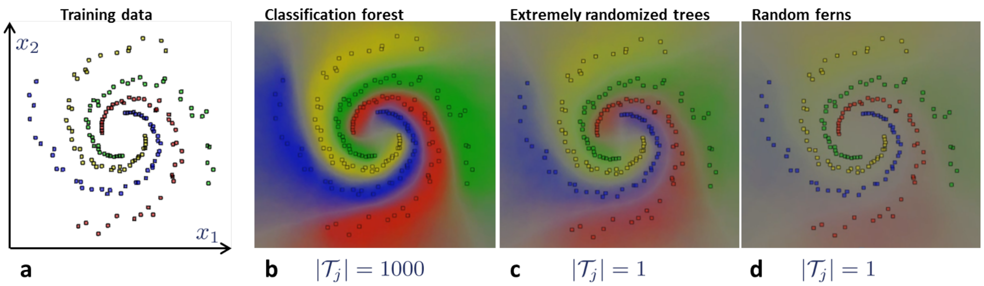

Fig. 8.1: **Forests, extremely randomized trees and ferns. (a)** Input training points for four classes. **(b)** Posterior of a classification forest. **(c)** Posterior of an ensemble of extremely randomized trees. **(d)** Posterior of a random fern. The randomness parameter is changed as illustrated. All other parameters are kept fixed. Extremely randomized trees are faster to train than forests but produce a lower-confidence posterior (in this example). The additional constraints of random ferns yield further loss of posterior confidence.

Figure 8.1 shows a comparison between classification forests and extremely randomized trees for a toy example. Some training points belonging to four different classes are randomly distributed along four spiral arms. Two decision forests were trained on the data. One of them with $\rho = 1000$ and another with $\rho = 1$ (extremely randomized). All other parameters are kept identical ($T = 200, D = 13$, weak learner = conic section, predictor = probabilistic). The corresponding testing posteriors are shown in fig. 8.1b, and fig. 8.1c, respectively. It can be observed that the increased randomness produces lower overall prediction confidence. Algorithmically higher randomness yields slower convergence of test error as a function of the forest size T.

8.2 Random ferns

Random ferns can also be thought of as a specific case of decision forests. In this case the additional constraint is that the same test parameters are used in all nodes of the same tree level [66, 68].

Figure 8.2 illustrates this point. As usual training points are indi-