



Fig. 3.6: **The effect of weak learner model.** The same set of 4-class training data is used to train 6 different forests, for 2 different values of D and 3 different weak learners. For fixed weak learner deeper trees produce larger confidence. For constant D non-linear weak learners produce the best results. In fact, an axis-aligned weak learner model produces blocky artifacts while the curvilinear model tends to extrapolate the shape of the spiral arms in a more natural way. Training has been achieved with $\rho = 500$ for all split nodes. The forest size is kept fixed at $T = 400$.

3.3.5 The effect of randomness

Figure 3.7 shows the same experiment as in fig. 3.6 with the only difference that now $\rho = 5$ as opposed to $\rho = 500$. Thus, much fewer parameter values were made available to each node during training. This increases the randomness of each tree and reduces their correlation.

Larger randomness helps reduce a little the blocky artifacts of the axis-aligned weak-learner as it produces more rounded decision boundaries (first column in fig. 3.7). Furthermore, larger randomness yields a