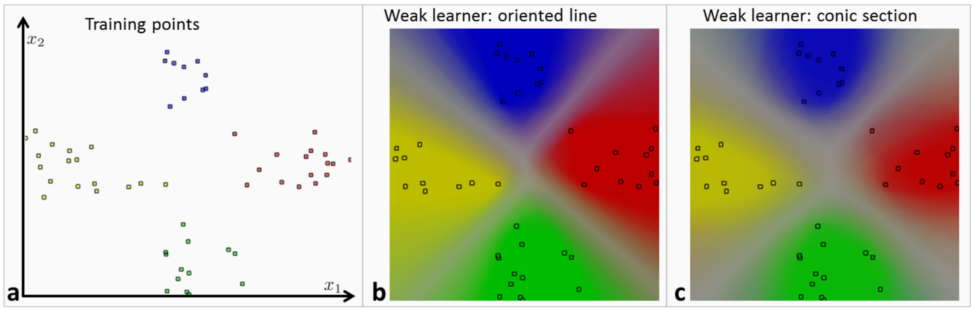

Fig. 3.11: **Forest's max-margin properties for multiple classes. (a)** Input four-class training points. **(b)** Forest posterior for oriented line weak learners. **(c)** Forest posterior for conic section weak learners. Regions of high entropy are shown as grey bands and correspond to loci of optimal separation. In these experiments we have used the following parameter settings $\rho = 50, D = 3, T = 400$.

subsets of training data. So, each tree sees a different training subset. Its node parameters are then fully optimized on this set. This means that specific "support vectors" may not be available in some of the trees. The posterior associated with those trees will then tend to move the optimal separating surface away from the maximum-margin one.

This is illustrated in fig. 3.12 where we have trained two forests with $\rho = 500, D = 2, T = 400$ and two different randomness models. The forest tested in fig. 3.12a uses randomized node optimization (RNO). The one in fig. 3.12b uses bagging (randomly selecting 50% training data with replacement) on exactly the same training data. In bagging, when training a node, there may be a whole range of values of a certain parameter which yield maximum information gain (*e.g.* the range $[\tau_1', \tau_1'']$ for the threshold τ_1). In such a case we could decide to always select one value out of the range (*e.g.* τ_1'). But this would probably be an unfair comparison. Thus we chose to randomly select a parameter value uniformly within that range. In effect here we are combining bagging and random node optimization together. The effect is shown in fig. 3.12b. In both cases we have used a large value of ρ to make sure that each tree achieves decent optimality in parameter selection.