



Fig. 3.5: **The effect of tree depth.** A four-class problem with both mixing of training labels and large gaps. (a) Training points. (b,c,d) Testing posteriors for different tree depths. All experiments were run with $T = 200$ and a conic weak-learner model. The tree depth is a crucial parameter in avoiding under- or over-fitting.

be placed roughly in the middle of the gap.²

Finally, we notice that a large value of D ($D = 15$ in the example) tends to produce *overfitting*, *i.e.* the posterior tends to split off isolated clusters of noisy training data (denoted with white circles in the figure). In fact, the maximum tree depth parameter D controls the amount of overfitting. By the same token, too shallow trees produce washed-out, low-confidence posteriors. Thus, while using multiple

²This effect will be analyzed further in the next section.