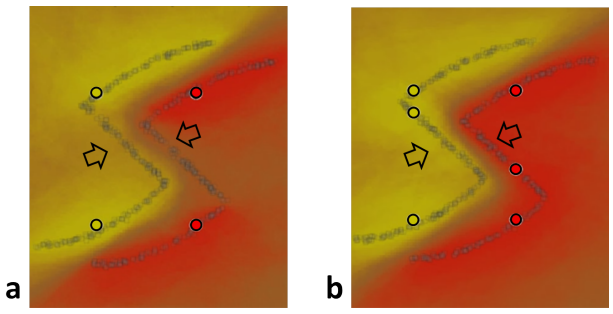

Fig. 7.4: **Active learning. (a)** Test forest posterior trained with only four labelled points and hundreds of unlabelled ones. The middle region shows lower confidence (pointed at by two arrows). **(b)** As before, but with two additional labelled points placed in regions of high uncertainty. The overall confidence of the classifier increases considerably and the overall posterior is sharper. Figure best seen on screen.

an economical way, the collection of additional training data. Next we compare semi-supervised forests with alternative algorithms.

Comparison with support vector machines. Figure 7.5 shows a comparison between semi-supervised forests and conventional SVM [97] as well as transductive SVM [47, 101], on the same two input datasets. [2]

In the figure we observe a number of effects. First, unlike SVM the forest captures uncertainty. As expected, more noise in the input data (either in the labelled or unlabelled sets, or both) is reflected in lower prediction confidence. Second, while transductive SVM manages to exploit the presence of available unlabelled data it still produces a hard, binary classification. For instance, larger amounts of noise in the training data is not reflected in the TSVM separating surface.

[2] In this example the SVM and transductive SVM results were generated using the "SVM-light" Matlab toolbox in `http://svmlight.joachims.org/`.