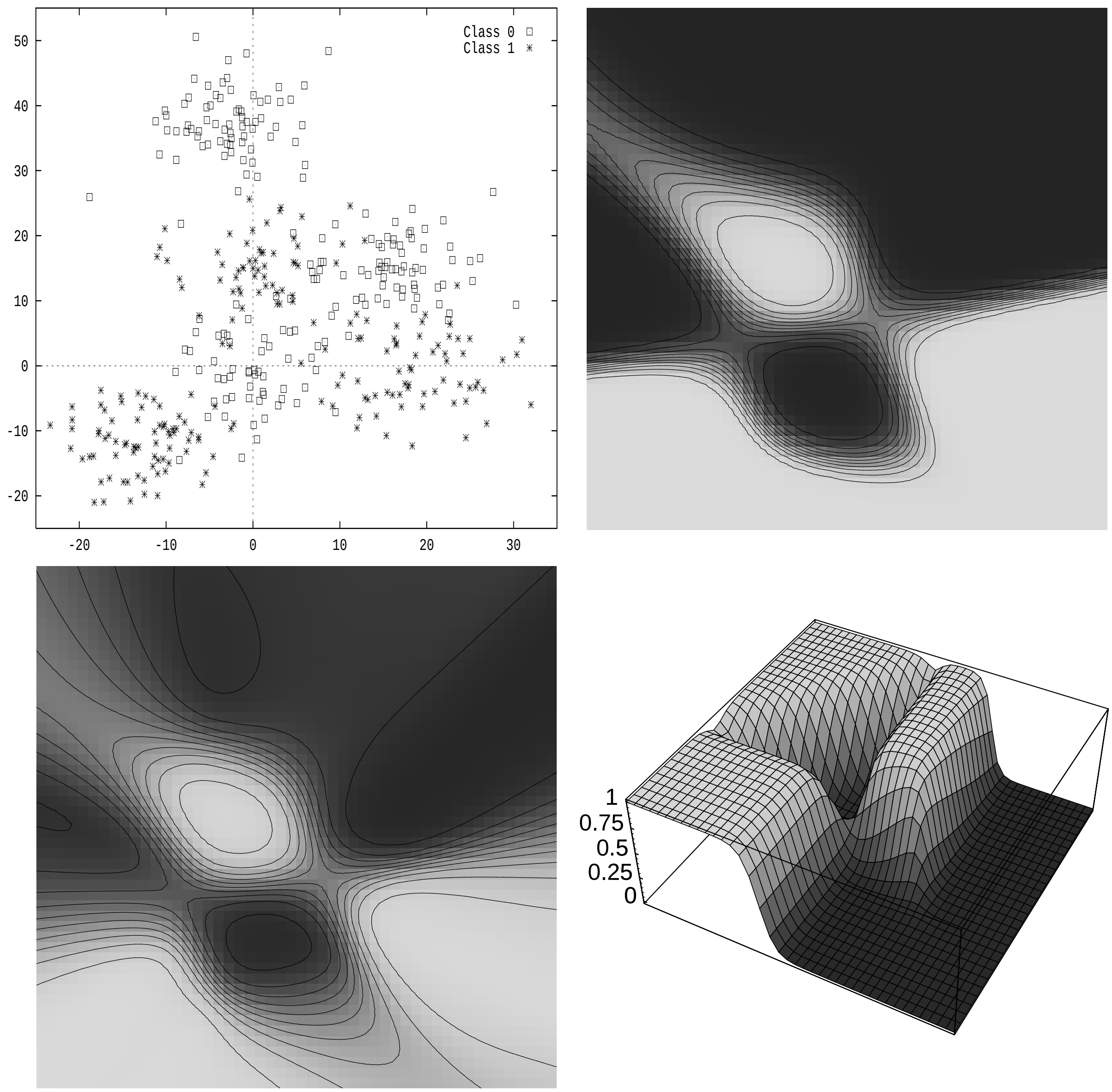

Figure 5.2: **Comparison of most probable outputs and moderated outputs for a toy problem**
a) The data set. The data were generated from six circular gaussian distributions, three gaussians for each class. The training sets for the demonstrations use between 100 and 1000 data points drawn from this distribution. b) 'Most probable' output of an eight hidden unit network trained on 100 data points. The contours are equally spaced between 0.0 and 1.0. c) 'Moderated' output of the network. Notice that the output becomes less certain compared with the most probable output as the input moves away from regions of high training data density. d) The true posterior probability, given the class densities that generated the data. The viewpoint is from the upper right corner of (a). In (b,c,d) a common grey scale is used, linear from 0 (dark grey) to 1 (light grey).