

Figure 1. An example of a two-class problem where a maximum likelihood-derived decision surface is not optimal (adapted from [4]). In the exploded view, the shaded region indicates the error induced by modeling the separable data by Gaussians estimated using maximum likelihood. This case is common for data, such as speech, where there is overlap in the feature space or where class boundaries are adjacent.

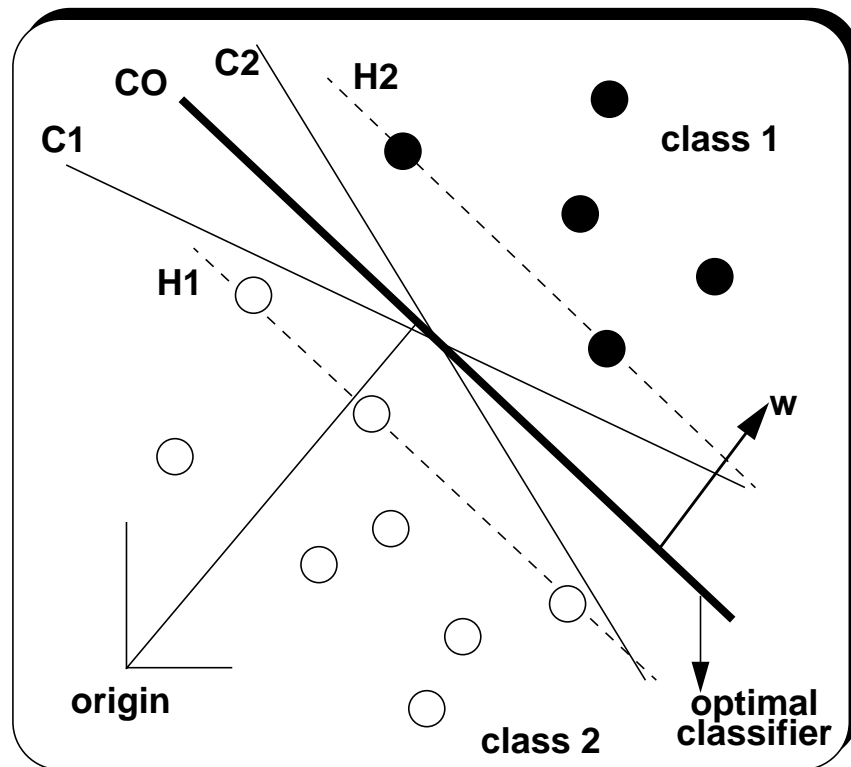


Figure 2. Difference between empirical risk minimization and structural risk minimization for a simple example involving a hyperplane classifier. Each hyperplane (C_0 , C_1 and C_2) achieves perfect classification and, hence, zero empirical risk. However, C_0 is the optimal hyperplane because it maximizes the margin — the distance between the hyperplanes H_1 and H_2 . Maximizing the margin indirectly results in better generalization.

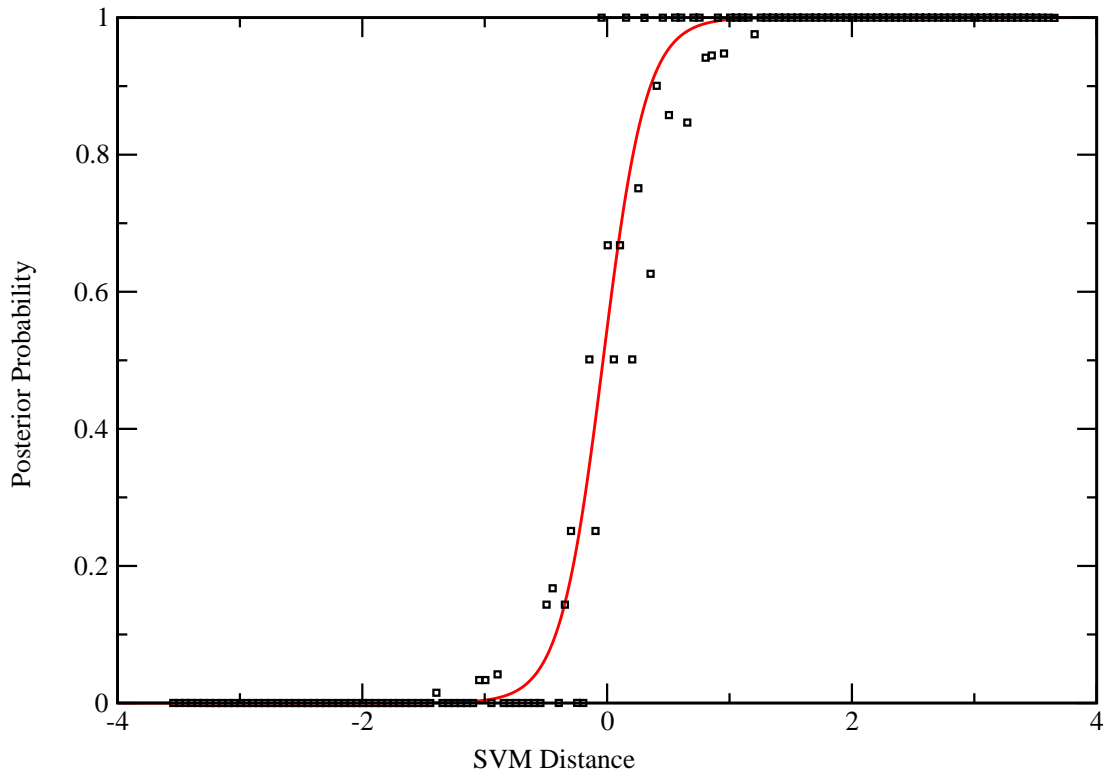


Figure 3. A sigmoid fit to the SVM distance-based posterior probability estimate.

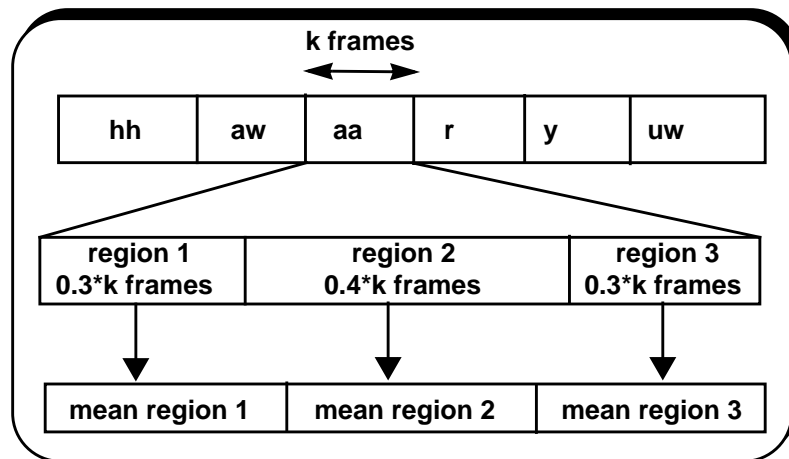


Figure 4. Composition of the segment level feature vector assuming a 3-4-3 proportion for the three sections.

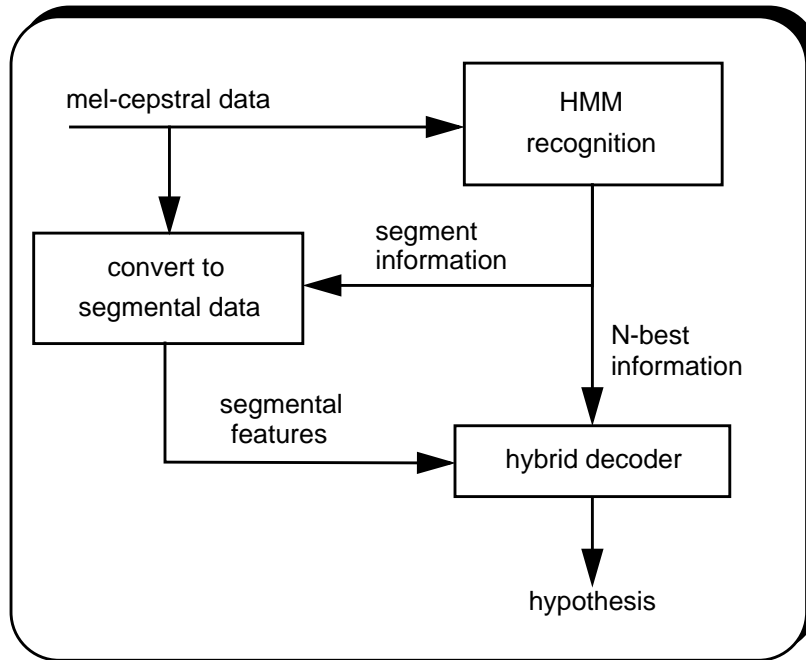


Figure 5. An overview of a hybrid HMM/SVM system based on an N-best list rescoring paradigm.