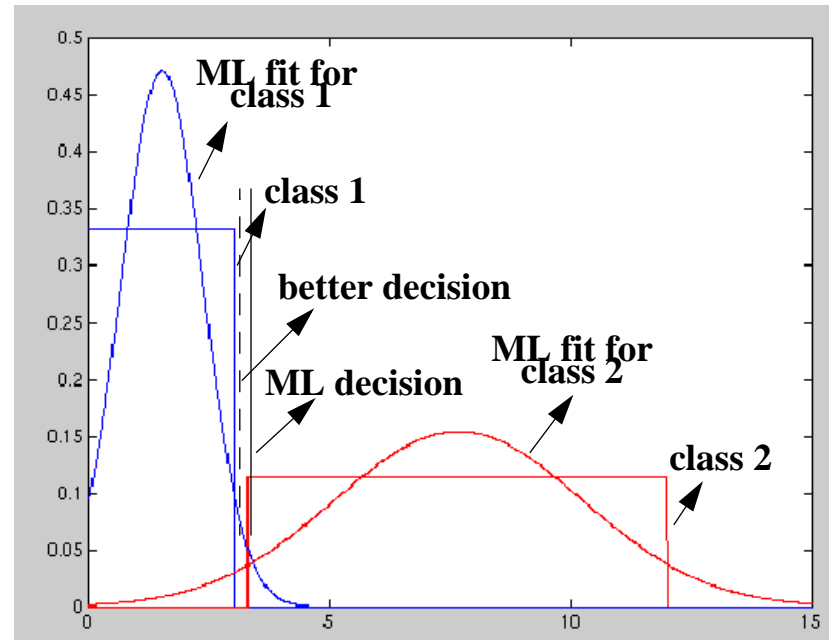


Motivation

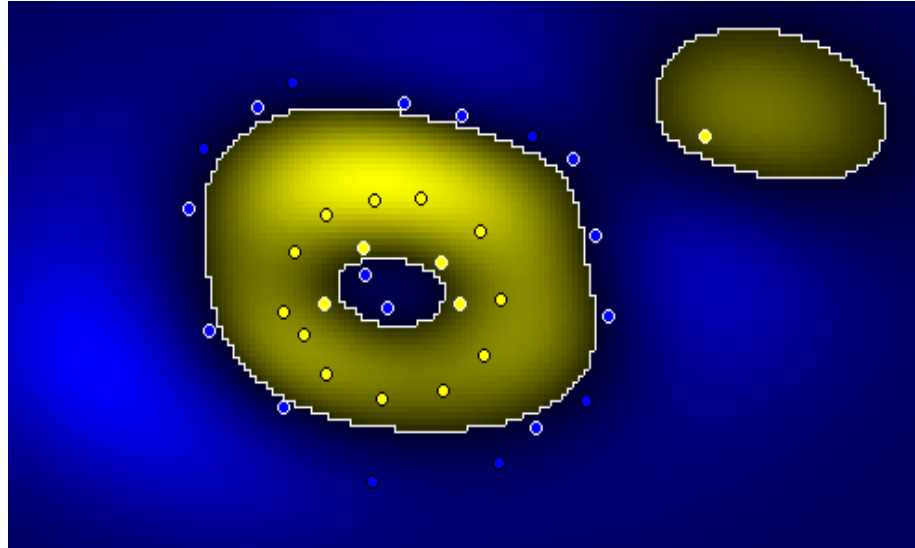
- ➡ **Need for discriminative techniques for enhanced acoustic modeling**
- ➡ **Maximum Likelihood based systems can be improved upon by discriminative machine learning techniques**
- ➡ **Support Vector Machines (SVM) have had significant success in several classification tasks**
- ➡ **A hybrid SVM/HMM system as a first step towards a SVM based recognizer**

Representation vs. Discrimination



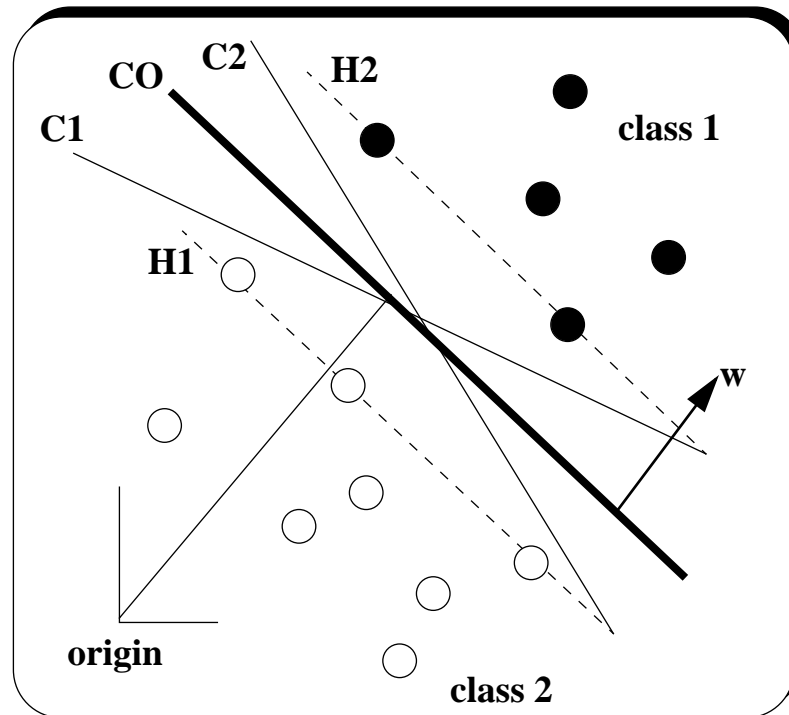
- ➡ **Efficient estimation procedures for classifiers based on maximum likelihood (ML)**
- ➡ **Convergence in ML does not guarantee optimal classification**

SVM Classifiers



- ➔ **Based on Structural Risk Minimization (SRM)**
- ➔ **Discriminative learning technique**
- ➔ **Models non-linear decision regions by transformation to higher dimension**

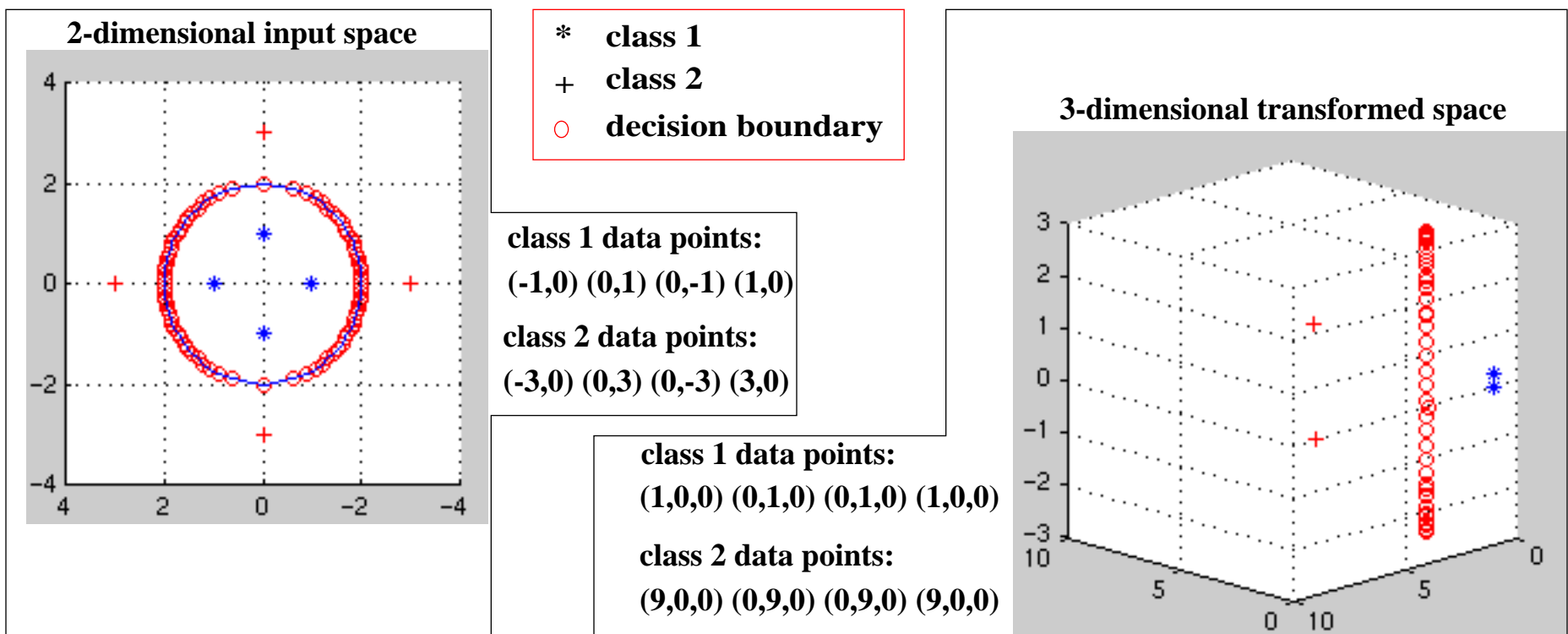
Optimal Hyperplanes



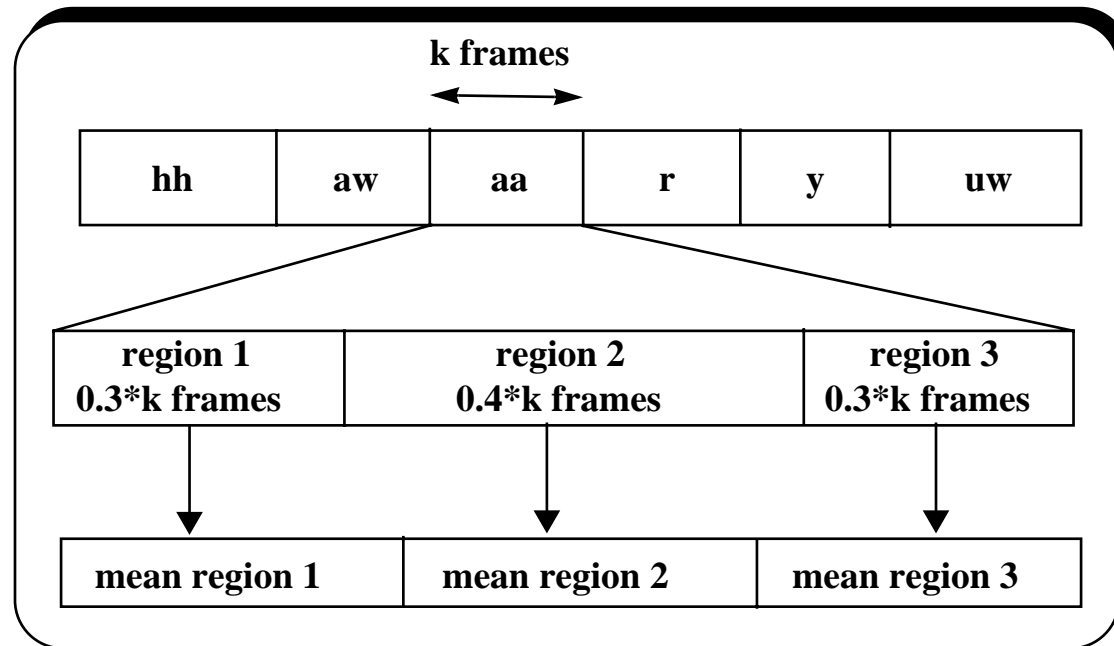
- ➔ **Empirical Risk Minimization is the most common form of optimization — does not guarantee good generalization**
- ➔ **SRM minimizes the bound on the risk**
- ➔ **Effectively chooses a classifier from a set of classifiers with the same empirical risk**

Kernels and Non-linear Classifiers

- ➡ **Non-linear classifiers estimated by implicitly transforming data to a high-dimensional space**
- ➡ **Kernels facilitate this transformation**
- ➡ **Polynomial and radial basis functions commonly used**

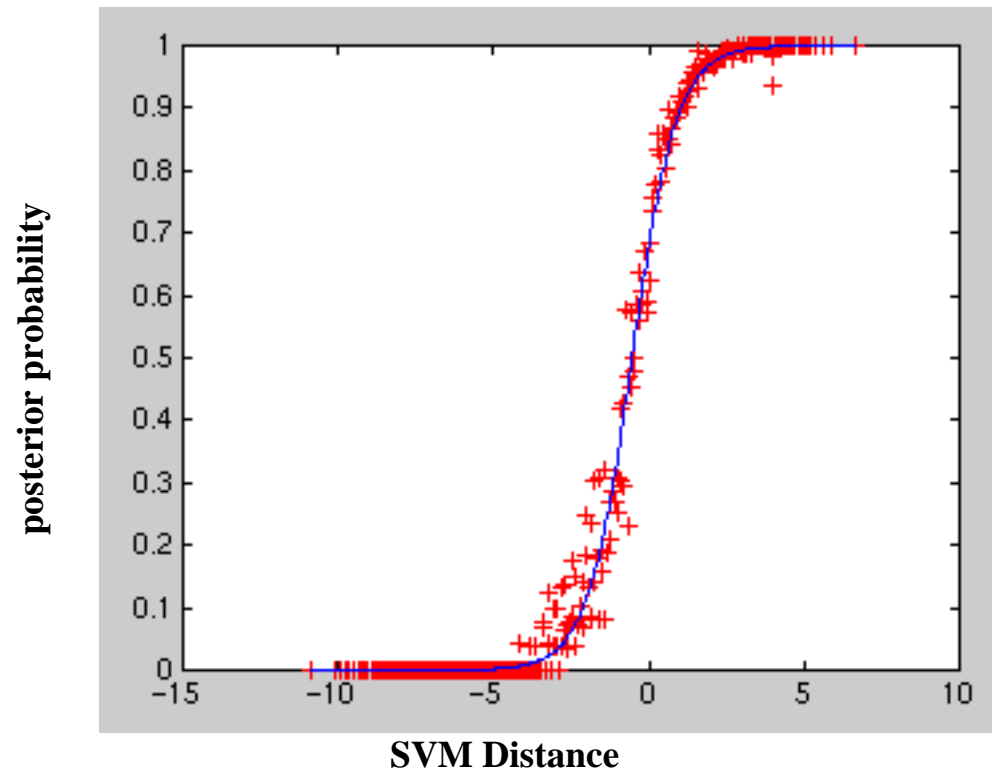


Composite Feature Vectors



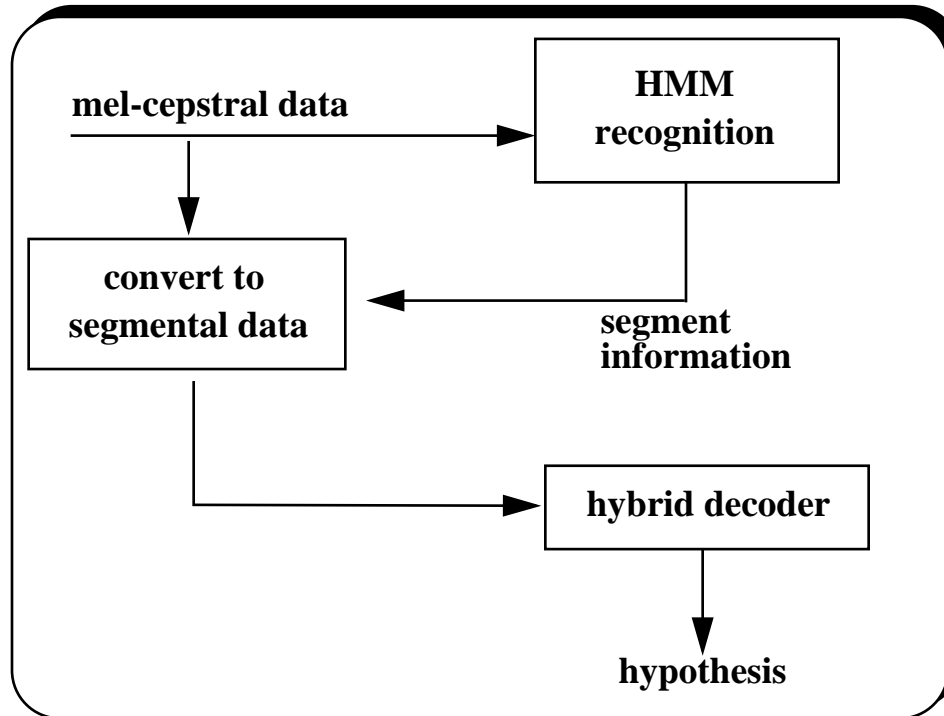
- ➡ **Two-fold need for composite feature vectors**
- ➡ **Allows for limiting data seen by each classifier**
- ➡ **Captures wider contextual variation**

Distances to Posteriors



- ➡ **SVM classification based on distances**
- ➡ **Search based on probabilities**
- ➡ **Sigmoid-based transformation estimated using class-conditional histograms**

Hybrid System Architecture



- ➡ **Gaussian computations replaced with SVM based probabilities in the hybrid decoder**
- ➡ **Composite feature vectors generated based on traditional HMM-based alignments**

Experimental Setup

☞ Training Data

☞ 8500 sentences from OGI Alphadigits for SVM classifiers

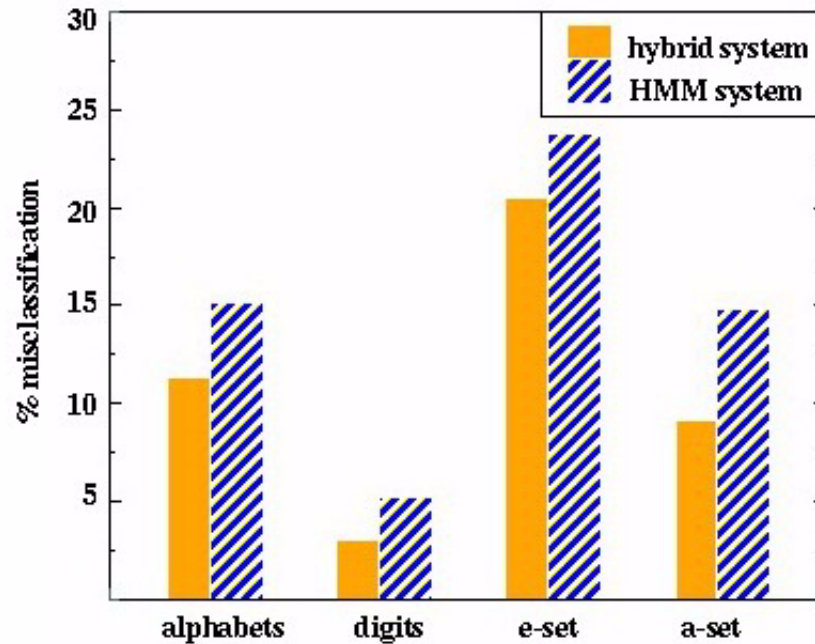
☞ 26000 sentences from OGI Alphadigits for the 12 mixture crossword HMM system

☞ Test data

☞ 1000 sentences from OGI Alphadigits — speaker independent open-loop test set

☞ Number of classifiers — 30 phone classifiers

Experiments



- **Crossword triphone HMM system — 12.7% WER**
- **Hybrid system with segmentations from the HMM system — 11.6% WER**
- **Composite feature vectors with a 3-4-3 proportion used**
- **Radial basis function kernels**

Further Experiments

- ➡ **Composite vectors based on the reference alignment for each utterance**
- ➡ **Hybrid system reduces error rate by 30% for OGI alphanumerals and 20% for SWB (relative to the baselines)**
- ➡ **Gains only moderate when hypothesis based alignments used for composite vectors**
- ➡ **Results suggest that:**
 - ➡ **bootstrapping to HMM alignments not optimal**
 - ➡ **online segmentation in a N-best framework may be better suited for SVM estimation**

Summary and Conclusions

- ➡ **First attempt at integrating SVMs into a full-fledged speech recognition system**
- ➡ **Hybrid architecture developed — Gaussian classifiers replaced by SVMs in an HMM-based system**
- ➡ **Encouraging results using composite vectors based on HMM-based hypotheses — 9% relative improvement on OGI Alphadigits**
- ➡ **Need to integrate segmentation into the SVM estimation loop — based on results from the cheating experiments where reference-based composite vectors are used**
- ➡ **Large scale classifier estimation is necessary**