USC University of Southern California

USC Viterbi School of Engineering

Ming Hsieh Department of Electrical Engineering

GREEDY FEATURE CLUSTERING FOR CLASSIFIER ENSEMBLE TRAINING

Rahul Gupta, Kartik Audhkhasi, Shrikanth Narayanan

Signal Analysis and Interpretation Lab (SAIL)

Motivation & Introduction

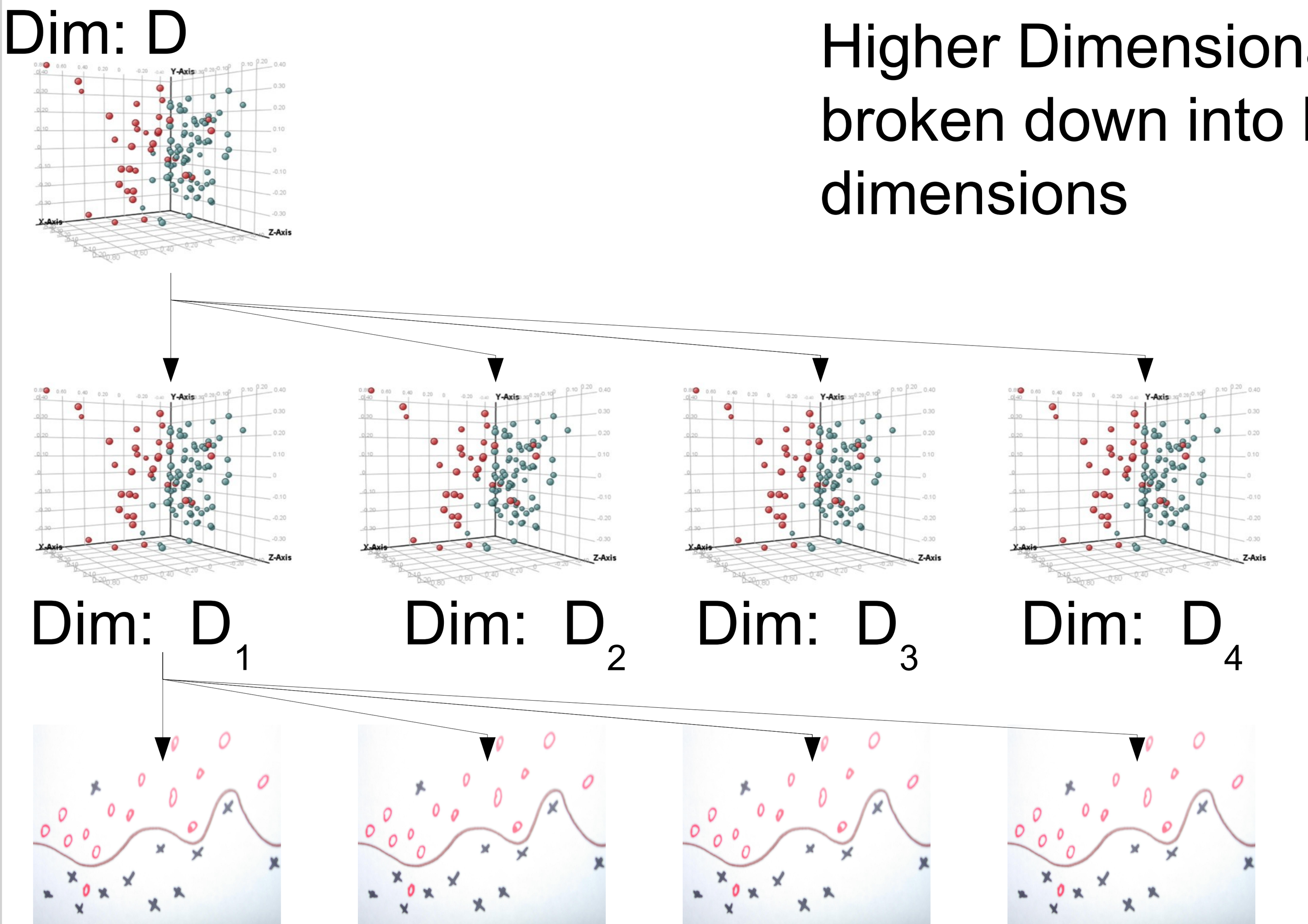
Classifier: f(X) → Y
X: features
Y: final label

- Combination of all feature all at once not always optimal
- We determine " Feature Clusters" in the data space (as opposed to data clusters in feature space)

Description

Dim: D

Higher Dimensional space broken down into lower dimensions



Train individual classifier and fuse the result

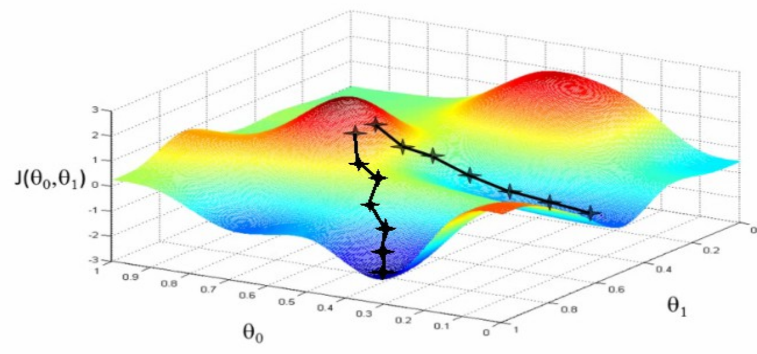
Algorithm

Input : Features $\{x_1, x_2, ..., x_N\}$,
Accuracy Threshold = α,
Number of Clusters = K
Output: Feature Clusters $(X_1, X_2, ..., X_{k+1})$
begin
$\Pi(x_1, x_2, ..., x_N) \xrightarrow{\text{ordered by accuracy}} (y_1, y_2, ..., y_N)$;
for $i = 1, 2, ..., K$ **do**
$X_i = \{y_i\}$;
$X_i^c = \{x_1, x_2, ..., x_N\} - \{y_i\}$;
$AccNew = f(DS, X_i)$;
$AccOld = 0$;
while $AccNew - AccOld \geq \alpha$ **do**
for $x_j \in X_i^c$ **do**
$X_i^t = X_i \cup x_j$;
$AccFeat(x_j) = f(DS, X_i^t)$;
end
$x_j^b = \underset{x_j}{\operatorname{argmax}}(AccFeat(x_j))$;
$X_i = X_i \cup x_j^b$;
$X_i^c = X_i^c - x_j^b$;
end
end
end
$X_{K+1} = \{x_1, x_2, ..., x_N\} - \cup_{i=1}^{K} X_i$;

- Forward feature selection with different initial "seeds"
- Feature clusters optimize accuracy
- Last feature cluster collection of all unselected features
- Feature clusters span entire feature space

Classifier fusion schemes

- Simple plurality
- Entropy based selection
- Log-likelihood summation
- Weighted log-likelihood summation
  - Weights determined by gradient ascent on accuracy function



- Entropy based weighted log-likelihood summation
- Confidence region based log-likelihood summation

Results

Database	#Data-Points #Classes #Features	Class Labels	Baseline Accuracies		Fusion Accuracies					
			Forward Feature Selection	Complete Feature Set	Simple Plurality	Entropy Feature Selection	Weighted Log Likelihood Summation			
							$w_k^n = 1$	w_k Data-tuned	$w_k^n = 1/H_d^k$	$w_k^n = Conf_k (D_d)$
Wine Quality Database [13]	4898 / 2 / 12	Quality $\gtrless 6$	71.70	75.26	76.79	74.44	73.89	74.88	73.95	71.52
MAGIC Gamma Telescope Data set [14]	19020 / 2 / 11	Signal vs Back-ground	86.06	85.67	86.71	86.36	86.53	84.24	86.49	80.85
Letter Recognition Data Set [15]	10000 / 3 / 8	(a,b,...,g) (h,i,...,o) (p,q,...,z)	52.30	50.80	54.00	53.90	55.20	56.00	55.90	47.10
IEMOCAP Database[16]	5498 / 4 / 11	Happy, Angry, Sad, Neutral	46.87	46.34	48.62	48.60	50.05	22.54	49.73	44.69

Conclusion and Future work

- Performance of fusion scheme depends on the database
- Feature clusters have better classification capability as compared to all the features taken together

Future work

- More feature clustering schemes (based on correlation, mutual information)
- Finding fusion schemes given the database characteristics.

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering