

# Multivariate time series classification

Nov 25, 2015

Je Hyuk Lee

Dept of Industrial Engineering, SNU

# Contents

---

- **1. Proposed methods**
- **2. Experiments (On-going)**
  - Data
  - Results

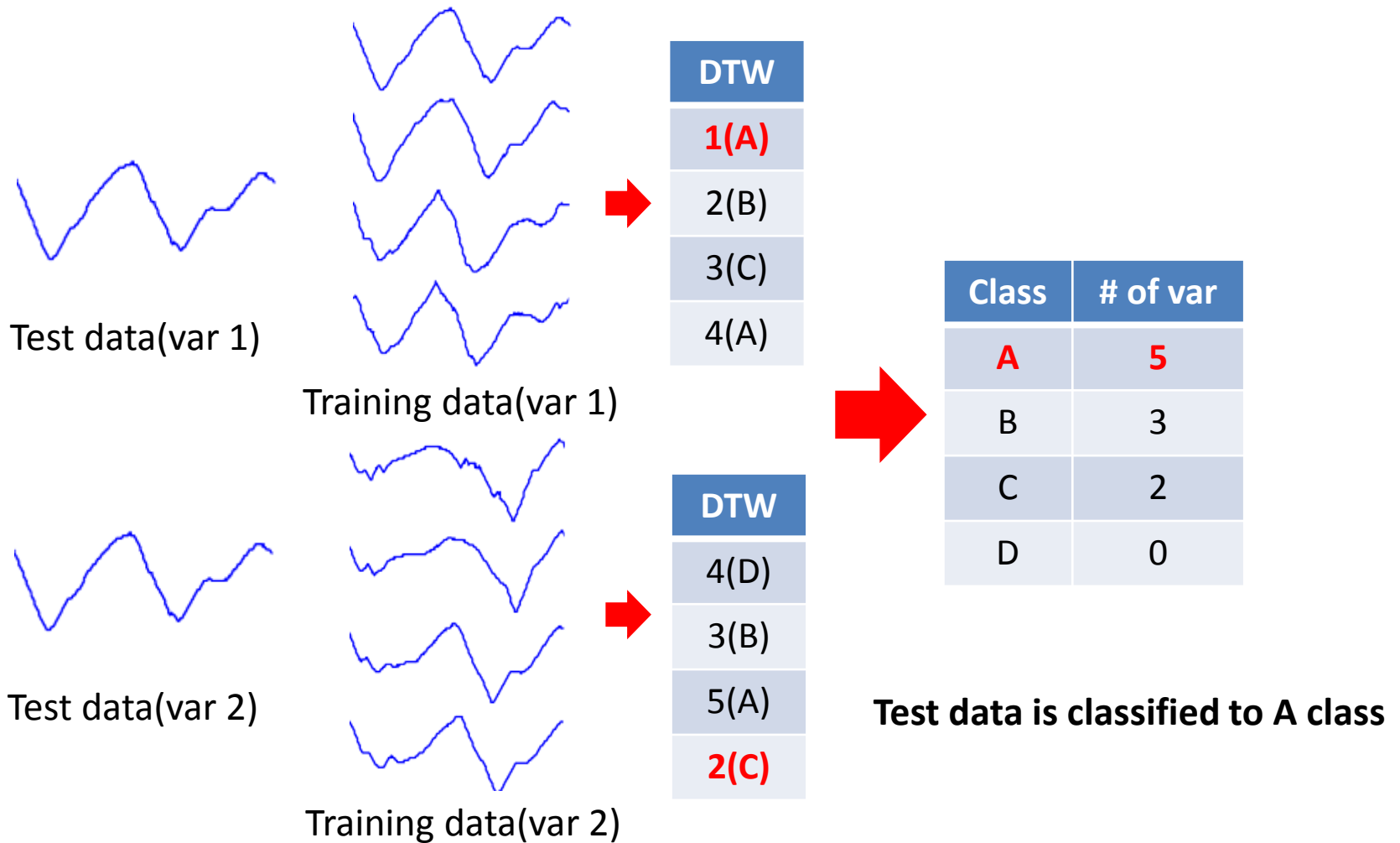
# Proposed Methods

---

- In this semester, I spent most of my time in shapelet classification approach
  - How to extend this concept to multivariate case?
- But, difficult to apply shapelet to multivariate classification experiment
  - It has very different time length even in the same dataset(7~29, 4~93)
  - Shapelet can be well made in the same length assumption
  - Also, some data points are too short to apply this concept

# Proposed Methods

- 1. DTW + 1NN classifier for each variable. Then vote.



# Proposed Methods

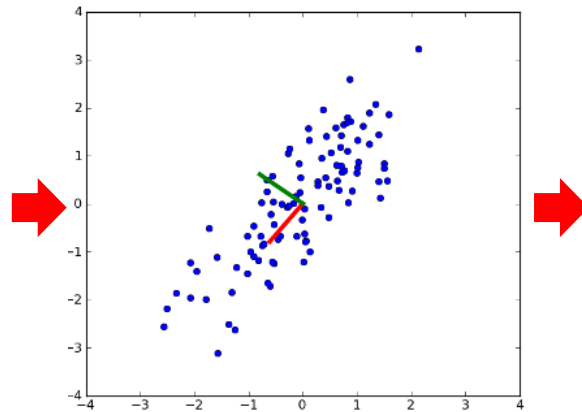
---

- Method (1) is simple and easy to understand
  - But it does not include anything about correlation structure
  - Also, if each variable have correlation structure
    - Some variables can overly cause influence to vote results
- We need two constraints
  - Sequences need to include correlation structure
  - Variable for voting should be nearly independent
  - How about using PCA?

# Proposed Methods

- 2. Use PCA projected sequence. Then, DTW+1NN+voting classifier

	v1	v2	...	vM
t1	a11	a12		a1M
t2	a21	a22		a2M
t3	a31	a32		a3M
t4	a41	a42		a4M
...				
tN	aN1	a2N		aMN

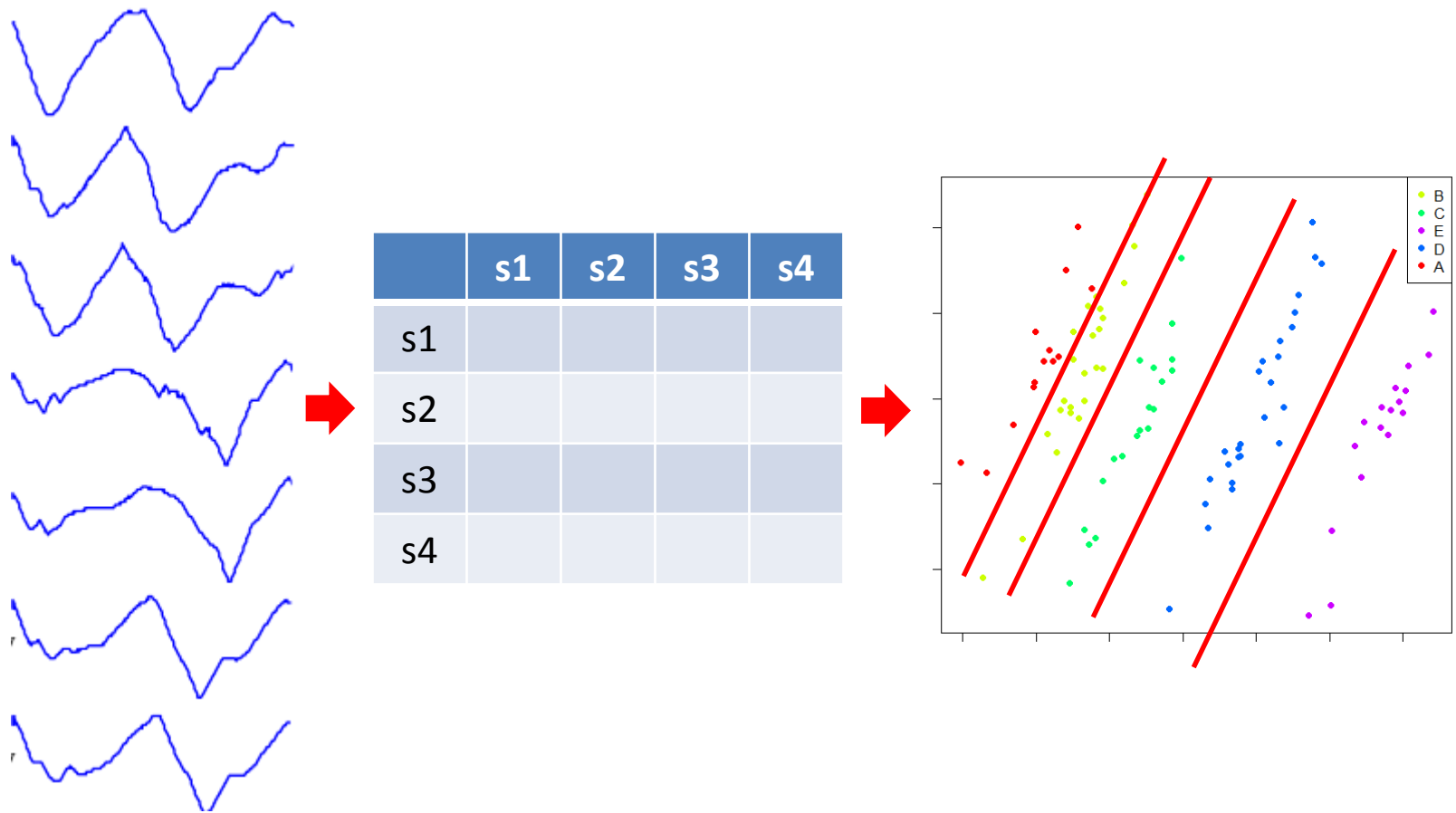


	v1	v2	...	vM
t1	a11'	a12'		a1M'
t2	a21'	a22'		a2M'
t3	a31'	a32'		a3M'
t4	a41'	a42'		a4M'
...				
tN	aN1'	a2N'		aMN'



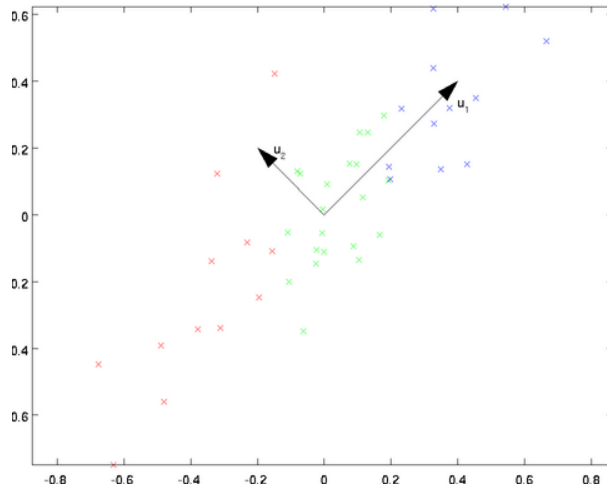
# Proposed Methods

- **3. Project to the coordinates and use conventional classifier (Not yet)**
  - First, calculate the DTW distance matrix for training data
  - Projected to the coordinate space (How? MDS??)
  - Use conventional classifier

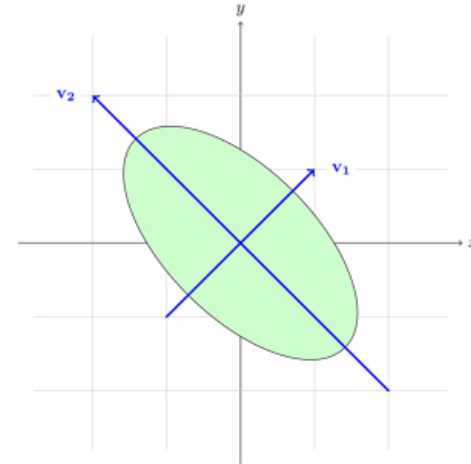


# Proposed Method

- (Sub) How about using Krzanowski distance(1-SPCA)?
  - The distance of two hyper plane which is made by PCs?



PCs from data 1



PCs from data 2

**Subspace1:** subspace spanned by PCs from data1

**Subspace2:** subspace spanned by PCs from data2

**Krzanowski distance:** Distance between Subspace1 and Subspace2



# Experiment

- Dataset

	Name	# of classes	# of Variables	Length	Training size	Test size
UCI	AUSLAN	95	22	45~136		
	Pendigits	10	2	8	300	10692
	Japanese Vowels	9	12	7-29	270	370
	Arabic Digits	10	13	4~93	6600	2200
	Character Trajectories	20	3	109~205	(2058)	(800)
	ECG	2	2	39~152	100	100
	Wafer	2	6	104~198	298	896

# Experiment

- Accuracy Results

Name	DTW+1NN	DTW+PCA	PCA coeff
AUSLAN			
Pendigits			
Japanese Vowels	73.78%	28.92% (1 PCs)	42.7% (2PCs)
Arabic Digits			17.36% (1PC)
Character Trajectories	84.13%	20.75% (1PC)	12.43% (1PC)
ECG			
Wafer			

# Experiment

- Classification Time Results

Name	DTW+1NN	DTW+PCA	PCA coeff
AUSLAN			
Pendigits			
Japanese Vowels	160.60sec	11.02sec (2 PCs)	2.20sec (2PCs)
Arabic Digits	(>1.5days)		487.40sec (1PC)
Character Trajectories	7043.54sec	1527.75sec (1PC)	30.99sec (1PC)
ECG			
Wafer			

# To be

---

- Do the unfinished experiment
- Compare the results to the pre-studied case
- For speed up, apply the constrained DTW method
- How about 2-class cases?
- Would variable selection cause influence to the performance in method 1?
- Would the 1-NN classifier is not suitable?