Bootstrap aggregated sparse FPCA for classification

Hyunsung Kim

Febrary 14, 2020

Department of Statistics
Chung-Ang University

How to perform sparse FPCA in `fdapace` package?

1. Estimate covariance function by **kernel smoothing**.
2. Conduct eigenanalysis.
3. Obtain FPC scores by **PACE** (Principal component Analysis through Conditional Expectation) method.

Simulation studies

Simulation studies

- *Probability-enhanced effective dimension reduction for classifying sparse functional data* (Yao *et al.*)
- 2 simulation models
    - Model II
    - Model IV
- 700 curves are generated with 200 training and 500 test set.
- Bagged classifers are obtained from 100 bootstrap resamples.
- 100 Monte Carlo repetitions for each model

Simulated data from model II

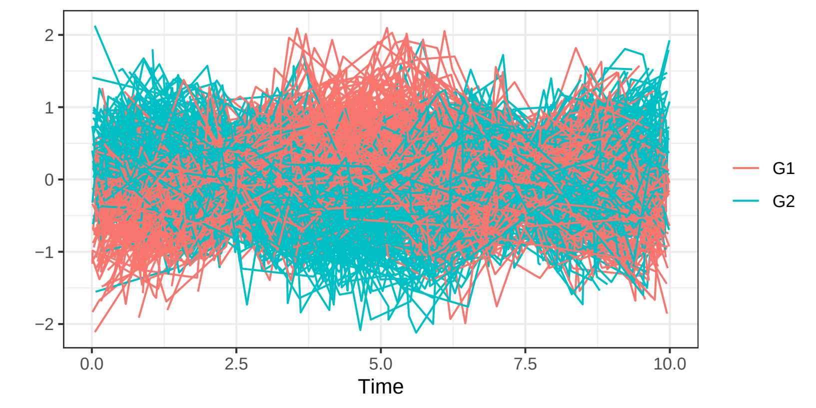

Figure 1: The simulated data obtained from Model 2

Results of model II

Table 1: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for sparse data (Model II)

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	16.7 (2.33)	16.8 (2.20)	17.5 (2.76)	16.6 (2.30)	17.8 (2.56)	18.4 (2.66)
Majority vote	15.6 (1.95)	15.9 (1.87)	16.2 (2.28)	15.8 (1.96)	16.5 (2.14)	17.3 (2.42)
OOB weight	16.0 (2.02)	16.2 (1.94)	16.6 (2.28)	16.1 (1.98)	16.9 (2.09)	17.7 (2.43)

Simulated data from model IV

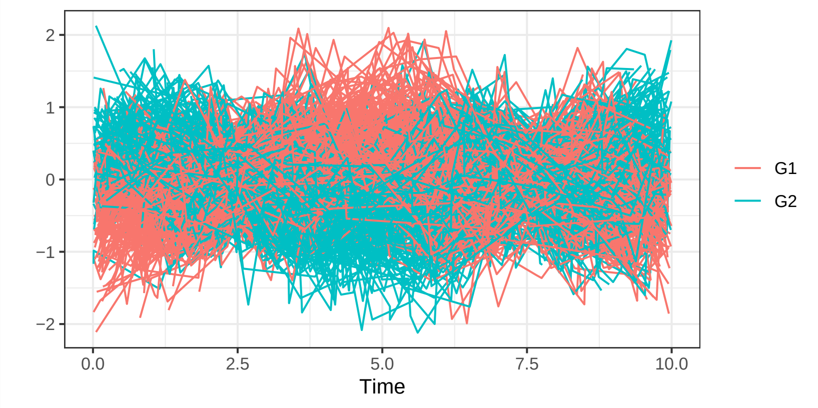

Figure 2: The simulated data obtained from Model 4

Results of model IV

Table 2: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for sparse data (Model IV)

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	12.8 (2.41)	12.8 (2.40)	13.3 (2.65)	12.8 (2.40)	13.8 (2.56)	14.8 (2.74)
Majority vote	11.2 (1.84)	11.1 (1.89)	11.5 (1.98)	11.2 (1.85)	11.9 (2.03)	13.3 (2.36)
OOB weight	11.6 (1.86)	11.5 (1.90)	12.0 (1.96)	11.6 (1.86)	12.3 (2.06)	13.6 (2.35)

Real data analysis

Real data analysis

- 2 real data applications
  - **Berkely growth data**
    - 93 curves with 54 girls and 39 boys
    - Split to 62 training and 31 test set randomly.
    - Dense data
    - Randomly sparsify with 12~15
  - **Spinal bone mineral density data**
    - 280 curves with 470 females and 390 males
    - Split to 187 training and 93 test set randomly.
    - Sparse data
- Gender classification
- Bagged classifers are obtained from 100 bootstrap resamples.
- 100 Monte Carlo repetitions for each data

Berkely growth data

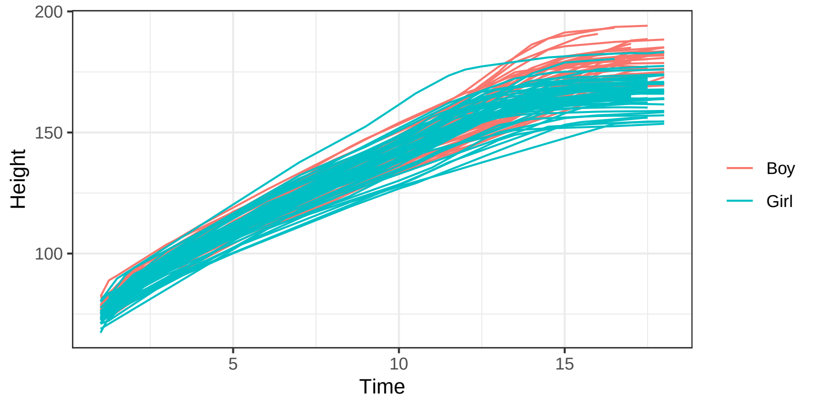

Figure 3: Berkely growth data sparsified randomly

Results of berkerly growth data

Table 3: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for berkerly growth data

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	7.3 (4.80)	5.3 (3.20)	5.7 (4.03)	5.8 (3.34)	5.6 (3.35)	5.6 (3.90)
Majority vote	5.9 (4.12)	4.9 (3.19)	5.3 (3.51)	5.4 (3.24)	4.9 (3.57)	5.5 (3.96)
OOB weight	5.9 (4.12)	5.0 (3.22)	5.4 (3.62)	5.4 (3.27)	4.9 (3.54)	5.5 (3.96)

Spinal bone mineral density data

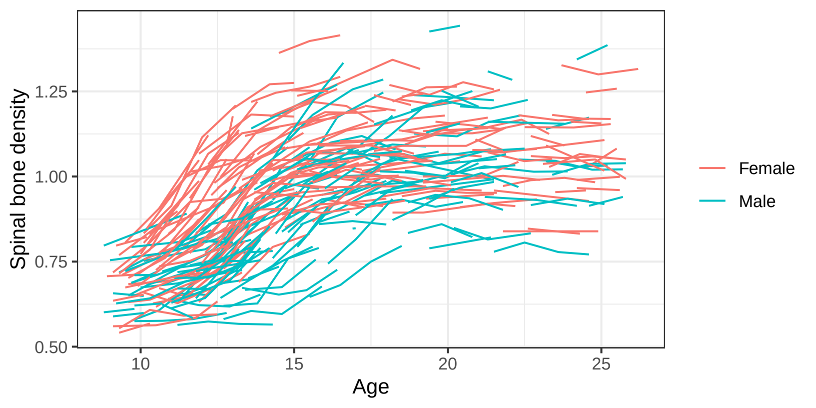

Figure 4: Spinal bone mineral density data

Results of spinal bone mineral density data

Table 4: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for spinal bone mineral density data

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	31.3 (4.30)	32.0 (4.27)	33.2 (4.71)	31.4 (4.44)	33.3 (4.10)	32.3 (4.33)
Majority vote	30.2 (3.72)	30.8 (4.18)	31.2 (3.88)	30.4 (3.77)	31.6 (3.78)	30.9 (3.83)
OOB weight	30.3 (3.71)	30.8 (4.07)	31.4 (3.81)	30.5 (3.82)	31.8 (3.71)	30.9 (3.86)