Bootstrap aggregated sparse FPCA for classification

Hyunsung Kim

January 10, 2020

Department of Statistics
Chung-Ang University

Simulation

- 2 different simulations
  - The classifiers applied sparse FPCA for 1st simulated data
  - The classifiers with bootstrap aggregating for 1st simulated data
- The number of FPCs are selected by the proportion of variance explained(PVE) with over 99%.

Simulation results

Table 1: The accuracy of classifiers after FPCA

No. of obs	Logistic Regression	SVM (Linear)	SVM (Gaussian)	SVM (Sigmoid)	KNN	LDA	QDA	Naive Bayes
2	0.700	0.700	0.690	0.610	0.880	0.680	0.700	0.650
3	0.740	0.750	0.750	0.740	0.810	0.730	0.690	0.720
4	0.770	0.750	0.740	0.740	0.780	0.760	0.730	0.720
5	0.710	0.700	0.710	0.740	0.810	0.750	0.770	0.700
6	0.770	0.780	0.790	0.770	0.810	0.780	0.740	0.740
7	0.830	0.840	0.840	0.850	0.820	0.830	0.850	0.780
8	0.770	0.760	0.820	0.780	0.870	0.780	0.810	0.770
9	0.790	0.830	0.770	0.820	0.850	0.790	0.820	0.760
10	0.810	0.840	0.820	0.840	0.860	0.830	0.840	0.800
11	0.860	0.850	0.830	0.880	0.870	0.870	0.880	0.750
12	0.840	0.850	0.820	0.820	0.860	0.840	0.860	0.780
13	0.810	0.830	0.810	0.830	0.850	0.810	0.850	0.800
14	0.820	0.790	0.820	0.820	0.860	0.830	0.870	0.780
15	0.800	0.810	0.800	0.760	0.880	0.810	0.840	0.760
16	0.800	0.770	0.800	0.800	0.860	0.810	0.840	0.770
17	0.810	0.770	0.800	0.770	0.850	0.810	0.860	0.770
18	0.800	0.810	0.810	0.810	0.860	0.800	0.860	0.770
Average	0.790	0.790	0.789	0.787	0.846	0.795	0.812	0.754

Simulation results

Table 2: The accuracy with bootstrap aggregated curves

No. of obs	Logistic Regression	SVM (Linear)	SVM (Gaussian)	SVM (Sigmoid)	KNN	LDA	QDA	Naive Bayes
2	0.750	0.740	0.740	0.740	0.920	0.750	0.730	0.730
3	0.740	0.730	0.730	0.730	0.850	0.740	0.720	0.710
4	0.770	0.780	0.760	0.780	0.940	0.780	0.790	0.740
5	0.710	0.680	0.710	0.690	0.860	0.690	0.710	0.720
6	0.760	0.760	0.760	0.770	0.900	0.760	0.760	0.740
7	0.860	0.840	0.810	0.850	0.890	0.860	0.860	0.780
8	0.770	0.780	0.820	0.780	0.890	0.760	0.810	0.790
9	0.780	0.790	0.790	0.780	0.860	0.780	0.820	0.760
10	0.810	0.820	0.800	0.810	0.880	0.820	0.830	0.760
11	0.860	0.850	0.840	0.850	0.880	0.870	0.870	0.760
12	0.840	0.840	0.820	0.830	0.880	0.850	0.860	0.780
13	0.800	0.810	0.810	0.810	0.900	0.820	0.830	0.790
14	0.820	0.810	0.820	0.840	0.870	0.830	0.860	0.770
15	0.800	0.800	0.820	0.800	0.910	0.800	0.840	0.740
16	0.800	0.790	0.800	0.790	0.880	0.810	0.830	0.760
17	0.810	0.800	0.810	0.810	0.890	0.810	0.840	0.760
18	0.800	0.810	0.820	0.790	0.880	0.800	0.850	0.760
Average	0.793	0.790	0.792	0.791	0.887	0.796	0.812	0.756

Simulation

- 5 different aggregation methods
  - Majority vote
  - LSE-based weighting(2003, Kim *et al.*)
  - Training accuracy
  - Out-of-bag accuracy
  - The proportion of response class($P(Y = 1)$)
- The data has different sparsity with 6~12 time points randomly from 1st simulated data.
- The number of FPCs are selected by the proportion of variance explained(PVE) with over 99%.

Simulation results

Table 3: The accuracy of classifiers between different aggregation methods

Aggregation method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	SVM (Sigmoid)	KNN	LDA	QDA	Naive Bayes
Majority vote	0.80	0.78	0.77	0.77	0.87	0.78	0.80	0.76
LSE(-1 vs 1)	0.22	0.23	0.24	0.24	0.20	0.21	0.22	0.27
LSE(0 vs 1)	0.77	0.84	0.80	0.81	0.88	0.81	0.86	0.76
LSE(1 vs 2)	0.77	0.84	0.80	0.82	0.89	0.81	0.85	0.76
LSE(Normalize)	0.80	0.78	0.77	0.77	0.87	0.78	0.80	0.76
Train accuracy	0.80	0.78	0.79	0.77	0.87	0.79	0.80	0.75
OOB accuracy	0.80	0.79	0.80	0.78	0.87	0.79	0.81	0.75
$P(Y = 1)$	0.79	0.78	0.73	0.77	0.71	0.82	0.82	0.74
Bagging X	0.81	0.79	0.78	0.79	0.86	0.78	0.81	0.74