

Classificaion for fMRI curves with similar mean, different variation

2024-04-16

1. 데이터 형태

- 191 subjects obsereved from 232 timepoints
- 각 subject마다 82개 brain region에서 관측된 데이터
 - 즉, 1개 subject가 82 x 232 형태의 데이터를 가짐 (82 curves)

문제점

- region을 모두 이어서 1개의 curve로 만들면 19,024개 timepoints가 됨
 - 차원 너무 커서, cov 계산 불가
- 82 functional covariate을 고려하는 multivariate functional classification 방법??

2. 기존 분석 과정

1. 각 subject마다 region 별 curve들의 length를 계산하여 $n \times 82$ 인 multivariate data로 변환
2. sparse PCA를 적용하여 20개로 차원축소
3. $n \times 20$ 데이터로 multivariate data classification 방법 적용

3. 결과

1. 1st region에서의 curve만 이용

- 80% training, 20% test 데이터로 고려하였고, 100번 반복하여 결과 요약

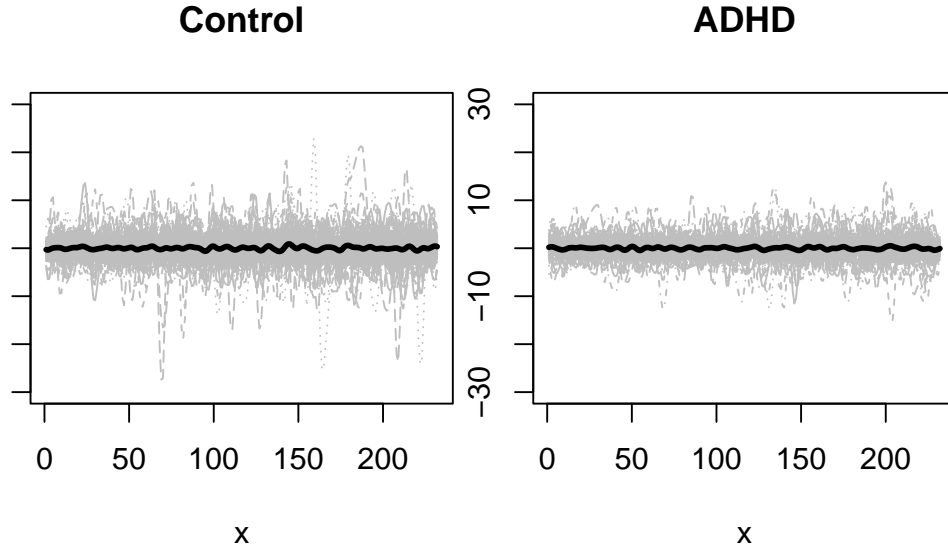


Figure 1: fMRI curves from 1st region.

- 다음의 비교 방법론 고려
 - VPC: Jiao, S., Frostig, R. D., & Ombao, H. (2023). Variation pattern classification of functional data. *Canadian Journal of Statistics*, 51(4), 943-958.
 - * feature function 개수는 $PVE \geq 0.90$ rule 적용
 - Fused: Baek, S., Kim, Y., Park, J., & Lee, J. S. (2022). Revisit to functional data analysis of sleeping energy expenditure. *Journal of Applied Statistics*, 49(4), 988-1002.
 - * 현재 knots 개수 = 50, penalty parameters $\lambda_1 = 0.1, \lambda_2 = 0.1$ 로 fixed된 결과
 - FGLM: Functional generalized linear model (logistic)
 - * 내부적으로 B-spline basis expansion을 적용

	VPC	Fused	FGLM
ACC (se)	0.443 (0.087)	0.464 (0.079)	0.572 (0.068)

- 전체적으로 결과가 매우 안 좋음
- VPC의 경우도, 시뮬레이션에서 매우 fluctuate하고 variance만 달라지는 경우를 고려하지는 않은 것처럼 보임

2. 82 region에서의 curve 모두 이용

- 80% training, 20% test 데이터로 고려하였고, 100번 반복하여 결과 요약

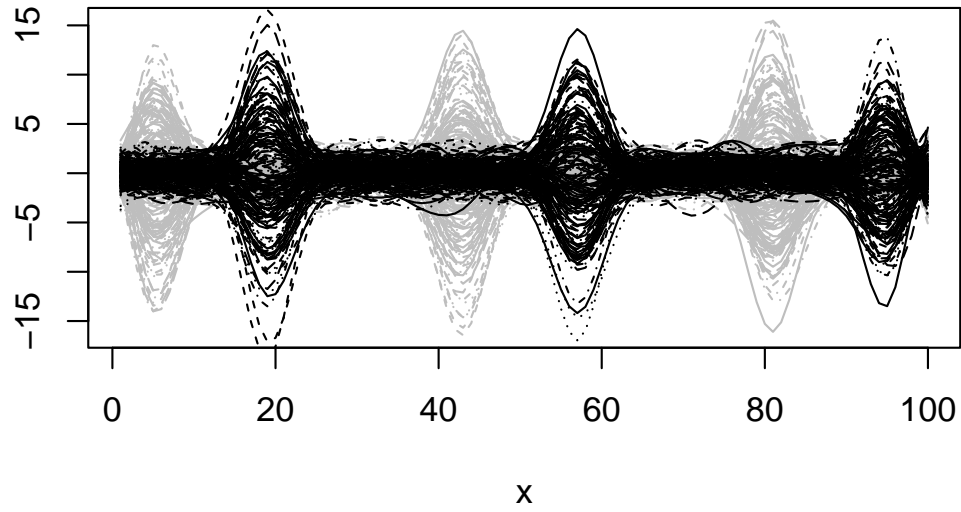


Figure 2: Simulated data from the VPC paper.

- Multivariate functional classification 적용
 - FGLM: Functional GLM (logistic)
 - * trainig data 수에 비해 parameter 수가 매우 많아져서 rank-deficient warning 발생
 - FGAM: Functional generalized additive model

	FGLM	FGAM
ACC (se)	0.494 (0.082)	0.514 (0.084)