6월 19일 (금) 랩미팅

박재영

1. **PLS와 PLS-DA 이해**
2. **PLS-DA 데이터 적용 및 VIP Score 계산**
   1. 0이 90% 이상인 feature를 제외한 데이터 (# of Compounds 106)
   2. 전체 데이터 (# of Compounds 191)
   3. 0-9일차 vs 11-18일차
   4. 0-5일차 vs 7-9일차 vs 11-18일차

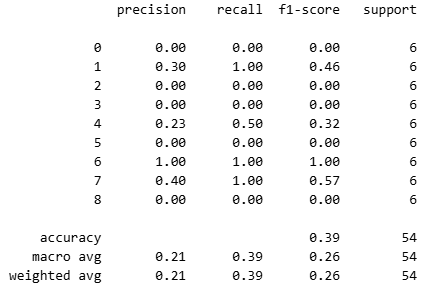
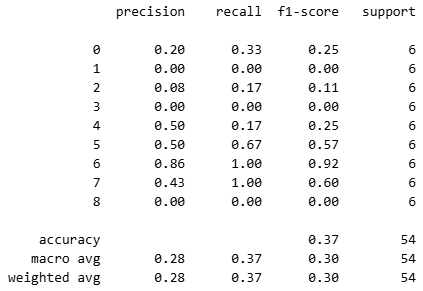
\*이지윤 연구원님이 임시 부패 기준에 따라 분류한 그룹을 참고.

1. **VIP Score 상위 10개 feature에 대한 ANOVA (일원분산분석) 및 사후검정**
2. **VIP Score>1 인 feature 중 A/B/C/D 모두에 나타나는 feature의 평균과 분산 변화**
3. **향후 분석 방향성**
4. **PLS와 PLS-DA 이해**

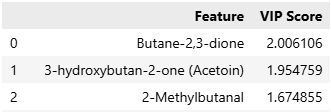
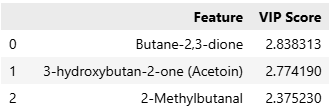
(git)

1. **PLS-DA 데이터 적용 및 VIP Score 계산**

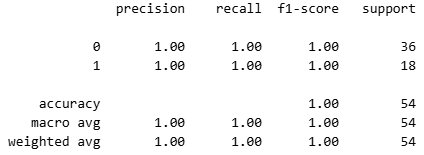
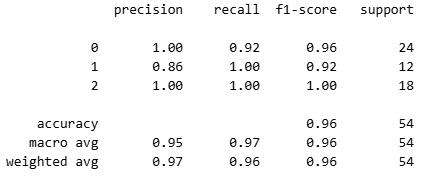
**A Compounds 106** **B Compounds 191 (total)**

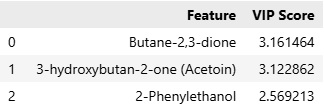
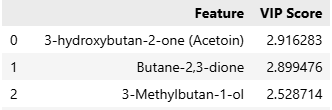
평균 CV accuracy: 0.6794 평균 CV accuracy: 0.7225

**C 0-9일차 vs 11-18일차** **D 0-5일차 vs 7-9일차 vs 11-18일차**

평균 CV accuracy: 0.764 평균 CV accuracy: 0.7452

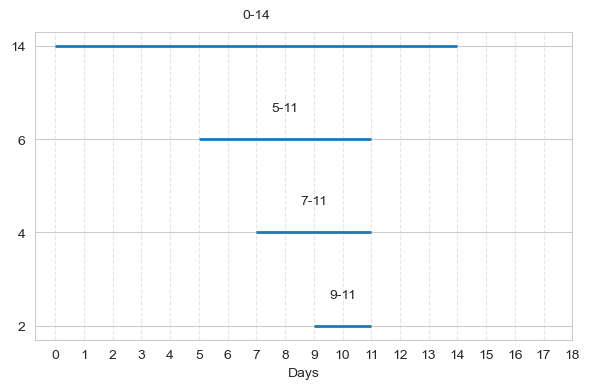
**** 

\* 모델과 교차검증 정확도 차이 – 모델에서는 train/test 데이터

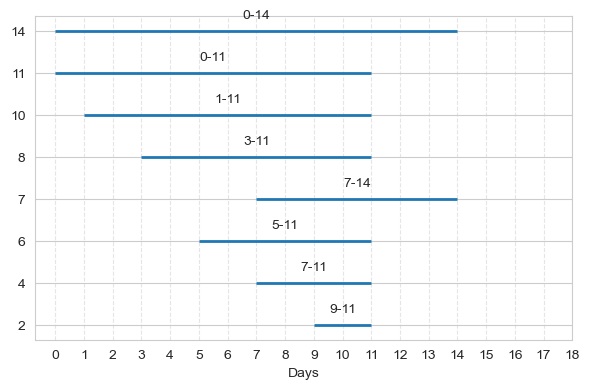
\* accuracy 너무 높다 – 특정 test 데이터에 너무 좋은 성능, cv는 낮음, 추가 검증 필요

1. **VIP Score 상위 10개 feature에 대한 ANOVA (일원분산분석) 및 사후검정**

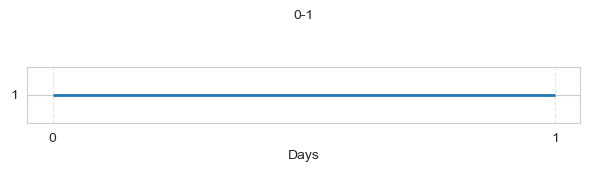
**A Compounds 106**



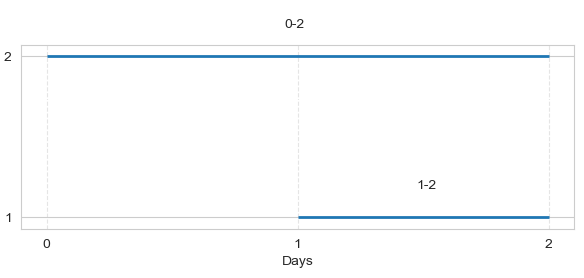
**B Compounds 191 (total)**



**C 0-9일차 vs 11-18일차**



**D 0-5일차 vs 7-9일차 vs 11-18일차**



\* A, B는 11일차를 기준으로 나누어지는 것처럼 보인다.

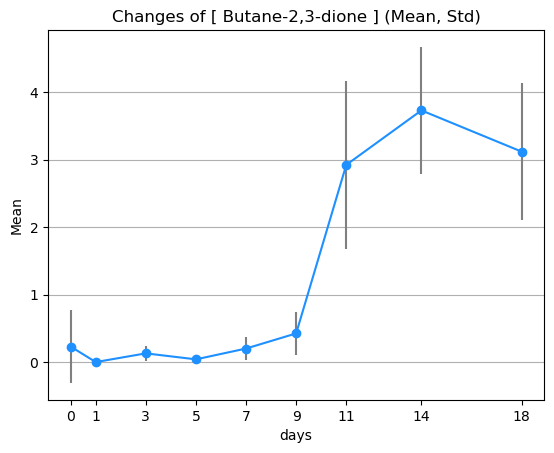
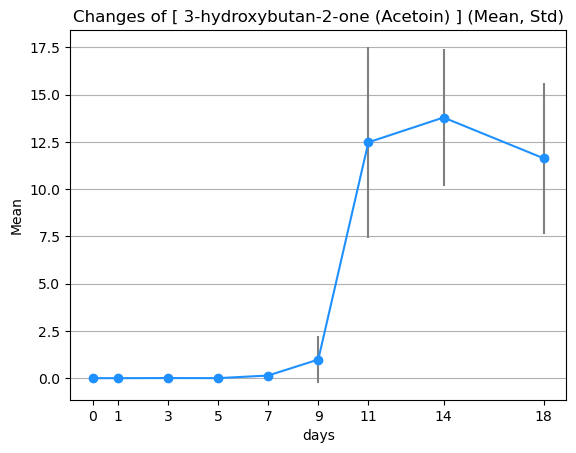
\* C에서는 상위 10개 feature가 모두 ‘0~9일차’ 와 ’11-18일차’ 두 그룹 간 유의한 차이가 있다.

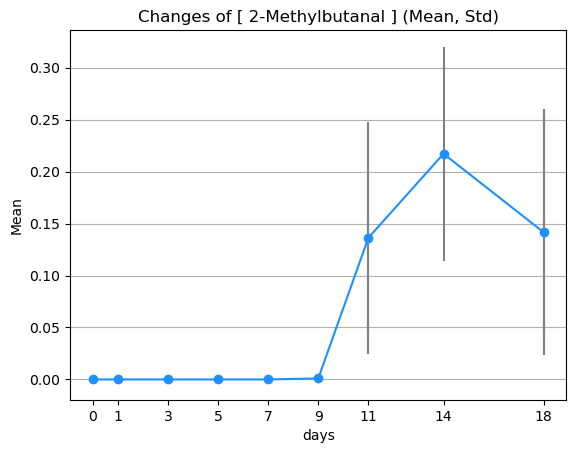
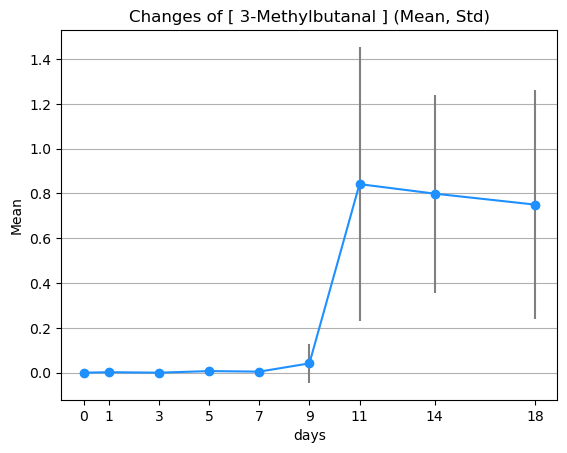
\* D에서는 ‘0-5일차’ 와 ‘7-9일차’ 두 그룹은 차이가 적다.

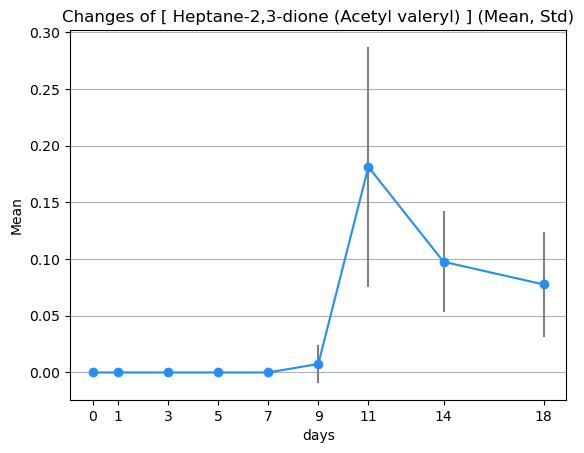
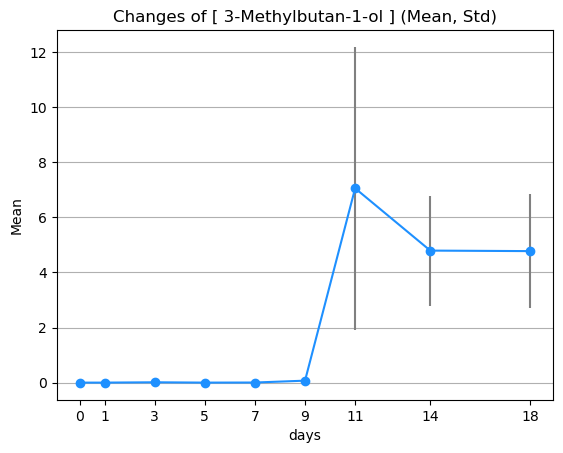
(A 결과값 예시)



1. **VIP Score>1 인 feature 중 A/B/C/D 모두에 나타나는 feature(23)의 평균과 분산 변화**

1. **향후 분석 방향성**

* 일자별 분류보다는 2-클래스, 3-클래스 분류가 유의한 것으로 보이므로, 다른 조합 찾기
* PCA 분석과 비교
* 군집 분류 (K-Means Clustering 등)
* 베이지안 최적화 적용?