

# 원자력 발전소 상태 예측

Linear discriminant Analysis 의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

## 제공된 데이터 셋

한국수력원자력에서 제공받은 원자력발전소 모사 데이터이다. 주어진 데이터 내에서 발전소의 상태가 변하며 변하기 전 발전소 상태를 상태\_A, 변한 후 상태를 상태\_B 이다.

## Linear Discriminant Analysis - hyperparameter 값에 따른 성능분석

이 레포트에서는 solver 와 shrinkage, n\_components 의 변경에 따른 성능결과를 분석합니다.

no	solver	shrinkage	n_components	score
1	svd	-	None	0.59248
2	lsqr	None	None	0.59226
3	lsqr	auto	None	0.58074
4	eigen	None	None	0.59226
5	eigen	auto	None	0.58074
6	svd	-	1	0.59248
7	svd	-	2	0.59248
8	svd	-	10	0.59248

이 데이터 셋에서는

**solver** default 값은 svd 로, 다른 solver 인 lsqr 이나 eigen 에 비해 더 성능이 좋은 것으로 보인다.

**shrinkage** 이 파라미터는 solver 가 lsqr 과 eigen 일 때만 설정 값이 유효하여 두 solver 에만 적용해 보았다. lsqr 과 eigen 의 경우 모두 shrinkage 가 auto 일때보다 default 값인 None 일 때 성능이 더 우수한 것으로 측정되었다.

**n\_components** 여러값을 대입해 보았다가 성능에 변화가 없어 scikit-learn 의 매뉴얼을 확인해보니 *This parameter only effects the **transform** method.* 라는 말을 확인했다. 알고보니 transform 메소드에만 영향을 주는 파라미터 값이었다.

여러 파라미터 값을 넣어봤지만, baseline\_lda 와 같은 점수만 받을 수 있었다. 결국 더 좋은 점수를 받게 된 건 데이터 셋에서 예외처리를 하고 난 뒤 fit 한 모델이었다. parameter 를 잘 설정해보는 것도 중요하지만 전처리도 중요함을 알 수 있었다.