Dacon_nuclear_study_1st_3조

3 조 3 주차 발표자료_차원축소

PCA

```
#차원축소 분류 예측 성능 평가

#원본 데이터세트
#랜덤 포레스트 이용해 타깃 값이 디폴트 값을 3개 교차 검증 세트로 분류 예측
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score

rcf=RandomForestClassifier(n_estimators=300, random_state=156)
scores = cross_val_score(rcf, X_train, y_train, scoring='accuracy', cv=3)
print('CV3인 경우 개별 fold 세트별 정확도 : ', scores)
print('평균 정확도 : {0:.4f}'.format(np.mean(scores)))
```

```
CV3인 경우 개별 fold 세트별 정확도 :  [0.7795827  0.80259345 0.79941585]
평균 정확도 : 0.7939
```

```
#PCA차원축소
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
#원본스케일링
scaler = StandardScaler()
df_scaled=scaler.fit_transform(X_train)
#컴포넌트 임의 6개 선정
pca=PCA(n_components=6)
df_pca = pca.fit_transform(df_scaled)
scores_pca = cross_val_score(rcf,df_pca, y_train, scoring='accuracy', cv=3)
print('CV3인 경우 PCA변환된 개별 fold 세트별 정확도 : ', scores_pca)
print('PCA 변환 데이터 세트 평균 정확도 : {0:.4f}'.format(np.mean(scores_pca)))
```

```
CV3인 경우 PCA변환된 개별 fold 세트별 정확도 :  [0.64587394 0.64872501 0.66316174]
PCA 변환 데이터 세트 평균 정확도 : 0.6526
```

- 원본 데이터세트(n_estimators=300) 평균 정확도 : 0.7939
- pca 차원축소결과 1 (n_components=6) 평균 정확도 : 0.6526
- pca 차원축소결과 2 (n_components=100) 평균 정확도 : 0.6579

- → pca 차원 축소효과가 크지 않음.(정확도가 약 15%떨어짐)

참고

전처리 과정 거친 후 차원축소 결과

- 원본 데이터세트 평균 정확도 : 0.7998

Dacon_nuclear_study_1st_3조

- pca 차원축소결과 1 (n_components=6) 평균 정확도 : 0.6572
- → 성능 변화 거의 없음

LDA

```
#LDA
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.preprocessing import StandardScaler
#원본스케일링
scaler = StandardScaler()
df_scaled=scaler.fit_transform(X_train)

#컴포넌트 임의 6개 선정
lda=LinearDiscriminantAnalysis(n_components=2)
lda.fit(df_scaled, train.label)
df_lda=lda.transform(df_scaled)
scores_lda = cross_val_score(rcf, df_lda, y_train, scoring='accuracy', cv=3)
print('CV3인 경우 lda변환된 개별 fold 세트별 정확도 : ', scores_lda)
print('lda 변환 데이터 세트 평균 정확도 : {0:.4f}'.format(np.mean(scores_lda)))
```

```
C:₩Users₩jih02₩AppData₩Local₩Continuum₩anaconda3₩lib₩site-packages₩sklearn₩discriminant_analysis.py:388: UserWar
ning: Variables are collinear.
  warnings.warn("Variables are collinear.")
```

```
CV3인 경우 lda변환된 개별 fold 세트별 정확도 :  [0.05183741 0.05201391 0.05213582]
lda 변환 데이터 세트 평균 정확도 : 0.0520
```

- 평균 정확도 : 0.0520
- 변수가 collinear 오류 발생