Module 7

PCA기반 얼굴인식

메가존 클라우드

2021

# PCA에 대해 서술하시오. (PCA가 언제 사용되며, 선형대수학의 어떤 개념들을 사용하고, DD^T가 아니라 D^TD를 사용하여 계산하는 이유)

* 구성 요소를 계산하는 과정으로써, 데이터의 기본적인 변화를 수행하기 위해 사용한다.
* 몇개만 사용하고 나머진 무시하는 경우도 있으며, 예측을 위한 모델을 만들 때, 탐구적 자료분석, 차원 감소를 위해 사용한다. 주된 구성 요소가 데이터의 공분산 매트릭스 eigenvector 라는 것을 보여준다.
* DDT -> DTD를 사용하는 이유는, 데이터의 사이즈가 너무 커지기 때문에

고유값, 고유벡터를 구함 -> 연산량이 많아짐.

# 각각의 등록된 사람에 대해 authentic/imposter histogram을 그리시오.

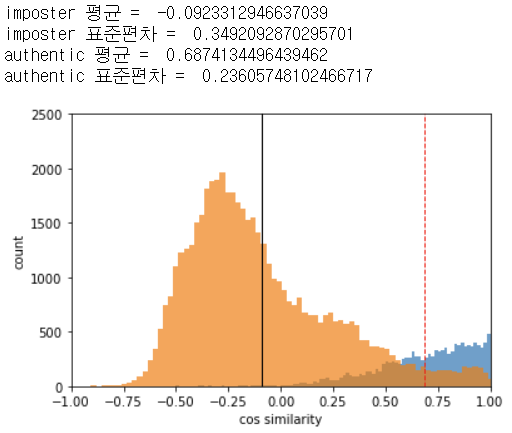


그림 Authentic / Imposter Histogram

# 전체 데이터에 대한 FAR/FRR을 축으로 하는 ROC curve를 그리시오. (이때, threshold는 0.05단위로 -1에서 1까지 그릴 것)

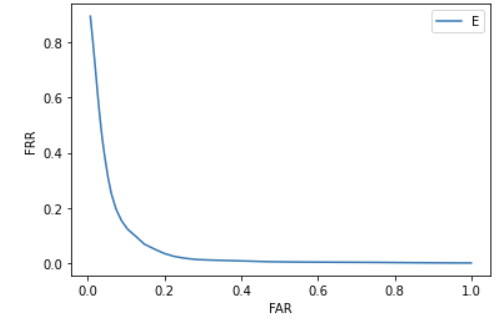


그림 FAR / FRR ROC Curve

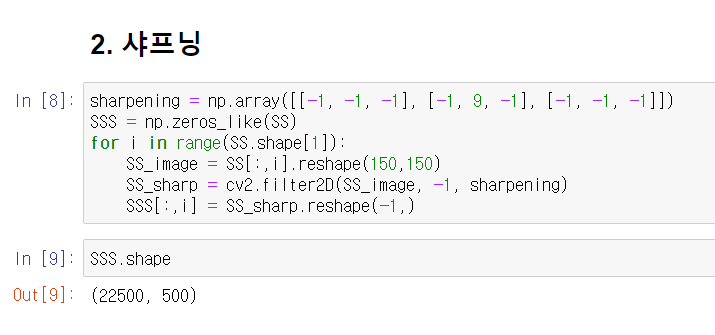
# Eigenvector가 15, 20, 25일 때의 authentic의 평균/표준편차, imposter의 평균/표준편차, threshold와 EER에 대한 결과표 작성 후 성능이 가장 높은 eigenvector와 threshold를 선택하고, 근거를 서술하시오.

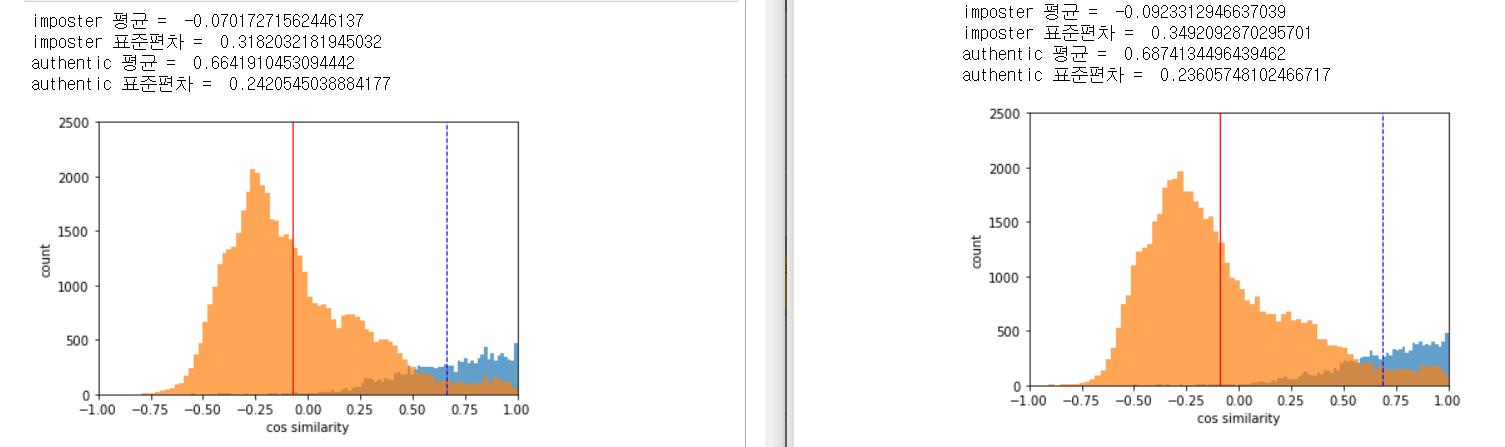
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Eigenvector** | | 15 | 20 | 25 |
| **Authentic** | **평균** | 0.7296 | 0.6874 | 0.6712 |
| **표준편차** | 0.2280 | 0.2361 | 0.2355 |
| **Imposter** | **평균** | -0.0776 | -0.0923 | -0.0706 |
| **표준편차** | 0.3651 | 0.3492 | 0.3348 |
| **Threshold** | | 0.45 | 0.4 | 0.35 |
| **EER** | | 0.0104 | 0.0190 | 0.0226 |

표 Eigenvector에 따른 각 값들의 변화

* EER이 가장 적을 때 성능이 최상이므로, eigenvector의 값이 15일 때 성능이 가장 좋은 것으로 나타난다.
* 그 때의 threshold 값은 0.45이다.

# 영상처리 기법을 사용해서 성능 향상된 내용을 지표로 보이고 자신이 생각하는 성능 향상의 근거를 서술하시오.





|  |  |
| --- | --- |
| **영상 처리 후 EER** | **기존 EER** |
|  |  |

표 샤프닝 처리 전후 threshold와 EER 비교

Eigenvector가 20일 때, sharpening 처리를 통해 EER값이 기존 0.0190에서 0.0012로 성능이 향상됨을 볼 수 있다.