

## 정보와지식 56121 2023년 봄학기 탐구 과제 2

제출 마감: 2022년 6월 24일(토) 23:59까지

탐구 내용: 어떤 데이터 세트에 PCA방법 (또는 목적에 따라 LDA)을 적용하여 새로운 축의 형태로 변환하고, 그 정보를 이용하여 목적에 맞는 파생 정보(결론)를 최종 생성한다.

1. kaggle.com에서 자신의 탐구 과제에 들어맞는 데이터 세트를 고른다. (또는 탐구과제 1에서 사용한 데이터를 사용하여서 결과를 비교하는 것도 좋음)
2. 이 데이터에 PCA를 적용하여 새로운 주축으로 변환하고 어떤 결론에 이를 수 있는 지식(또는 파생정보, 가령, 가치가 있는 정보의 형태)을 끌어낼 수 있는지 전체 과정을 정한다. (또는 LDA를 적용하여 패턴 인식의 문제를 해결해도 됨.)
3. 예를 들어, 우리나라 20~30세 남, 녀의 신체 수치 데이터를 선택했다면, PCA를 적용하여 신체 비례 지수 등의 몇 개의 새로운 도메인으로 변환하고, 이를 토대로 새로운 케주얼 상하의 세트를 맞출 때 기준이 되는 옷 치수의 표준을 만든다. 등등
4. 과정과 목적에 맞도록 PCA 또는 LDA를 적용하여 최종 결론을 얻고 결과의 장단점, 파생 가치 등을 논한다. 예를 들어 PCA로 변환된 (또는 차원 축소된) 데이터의 정보력을 판정할 기준을 정하고, 이를 토대로 파생 정보를 생성하여 그 성능을 평가한다. 또는 LDA를 적용한다면 N개에 대한 분류 판단의 오류 등으로 평가함.

제출할 내용: 다음의 내용을 자신의 학번으로 파일명으로 하는 하나의 압축 파일을 만들고 과제란에 마감 이전에 제출해야 합니다.

1. 실행 파일 및 소스 코드 (윈도우 또는 리눅스에서 실행할 수 있는 파일과 이를 만들 때 필요한 보조 파일도 함께 제출해야 합니다.)
2. 분석 결과를 도출하는 조건, 소프트웨어 구조와 실험 결과를 설명하는 논문 형식의 보고서 (A4 일반 여백, 폰트 10, 행간 130%, 5장 내외) 보고서에는 사용한 언어와 프로그램 실행 방법을 반드시 기재하며 알고리즘, 구현 방법, 결과 등을 상세히 기술한다.

평가 기준 및 방법: 다음의 항목을 기준으로 과제물을 평가합니다.

1. 탐구할 내용: 30% (탐구 과정의 완성도, 제출한 결과의 신뢰도)
2. PCA 적용: 30% (PCA의 성능 판정 정도와, 실행 시간 등. 프로그램 실행의 효율성은 각자 제출한 보고서에 기재된 프로그램 실행 방법을 기준으로 측정한다.)
3. 보고서 완성도: 40% (내용:20%, 형식:20%)

### \* 주의 사항

- 만일 두 사람의 과제 내용이나 데이터 세트의 구성이 일반적 수준 이상으로 같으면 두 학생의 성적 모두 낙제 처리합니다.
- 소프트웨어 패키지나 툴을 그대로 사용하면 50%만 인정합니다. PCA나 LDA의 라이브러리를 사용해도 되지만, 라이브러리 함수를 불러오기 위한 구조체등의 작성하면 100% 인정합니다.
- 제한된 마감 시간 이후에 제출되는 어떠한 형태의 것도 평가에서 제외됩니다. 반드시 제한 시간 이전에 업로드 완료해야 합니다. 마감에 인접하여 통신장애로 인해 업로드하지 못한 것도 인정되지 않습니다. 미리미리 업로드 하기 바랍니다. (수정이 필요하다면 다시 업로드하면 됩니다.)
- 과제 제출물을 제출하지 않으면 나머지 시험 점수와 관계없이 낙제(F)입니다. 단 과제 1이나 과제 2중 하나만 완료하면 됩니다. 둘다 제출하면 높은 점수를 인정합니다.