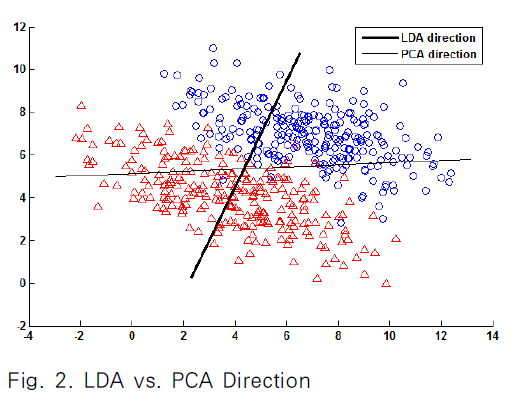
3월27일~3월31일

3월5주차 졸업논문 주제 선정 및 개념공부

- Principal Component Analysis (PCA)

PCA는 데이터에서 중요한 정보를 추출하여 주 구성 요소라는 새로운 직교 변수세트로 제시하는 데이터 압축 및 차원 감소를 위한 다변수 메소드이다. 첫 번째 주성분은 가능한 한 많은 데이터의 변동성을 설명하며, 나머지 각 구성 요소는 가능한 한 나머지 가변성의 많은 부분을 설명한다. 응용 분야에 따라 KarhunenLoèvetransform (KLT), 호스텔 변환 또는 proper orthogonal decomposition (POD)라고도 한다. PCA는 진정한 고유벡터 기반 다변수 분석 중 가장 단순하다. 이 작업은 데이터의 분산을 가장 잘 설명하는 방식으로 데이터의 내부 구조를 드러내는 것으로 생각할 수 있다. PCA는 요인 분석과 밀접한 관련이 있다. 실제로 일부 통계 패키지는 의도적으로 두 가지 기술을 결합한다. True factor analysis은 기본 구조에 대한 다른 가정을 하고 약간 다른 행령의 고유 백터를 해결한다.



-Linear Discriminant Analysis (LDA)

선형 판별분석은 클래스로 분류 되어있는 다차원공간의 데이터에 대해 클래스간의 분산이 최대가 되는 특징벡터를 찾는 방법이다. 선형 판별분석은 주성분분석과 비교될 수 있는데, 두 방법의 공통점은 다차원 데이터의 특성을 반영하는 특징벡터를 찾는다는 것이고, 차이점은 선형 판별분석이 데이터의 클래스 정보를 사용해 클래스 간의 분산이 최대가 되는 최적분류성질의 특징벡터를 찾는 반면, 주성분분석은 전체데이터의 분산이 최대가 되는 최적표현성질의 특징벡터를 찾는다는 것이다.

- K-means Clustering Algorithms

K-means는 기존 데이터를 두 개 이상의 그룹으로 나누기 위해 데이터를 그룹화하는 방법이다. 이 분할 방법은 특성에 따라 데이터를 차별화하기 위해 수행되고 비슷한 특성을 가진 데이터는 같은 그룹에 저장된다. 데이터 분류의 목적은 일반적으로 그룹 내에서의 변동을 최소화 하고 그룹 간의 편차를 극대화 하고자하는 그룹화 프로세스에서 설정된 목표기능을 최소화 하는 것이다.