알고리즘 마케팅 6강

2023. 3. 30. (목)

서울과학기술대학교 데이터사이언스학과
김 종 대

오늘의 강의

- 5차시 Review
- 일반선형모형: 로지스틱 / 프로빗 / 포아송 회귀분석
- 알고리즘 마케팅 개관
- 경제적 최적화
- 머신 러닝: 지도학습
- 머신 러닝: 비지도학습(표현학습)
- 고객 선택 모형
- 생존 분석
- 개인 프로젝트 관련 안내

상관관계분석

- **상관관계분석**: 두 변수 X, Y의 연관도, 즉, 선형관계에 대한 정도
  - 공분산(covariance)

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

    - 변수 간 선형관계의 방향과 유무를 판단할 수 있으나, 선형관계의 정도는 파악할 수 없음.
    - 예: 측정단위에 따른 공분산 값의 변화

  - 상관계수(correlation coefficient)

$$Corr(X,Y) = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

    - X,Y의 표준편차로 나누어 준 형태로서 언제나 $-1 \leq \rho(X,Y) \leq 1$

선형회귀분석

- **단순선형회귀분석**
  - 두 변수 사이에 존재하는 상호의존관계를 함수 관계로 표현하여 연관성을 검정하는 통계적 분석

  - $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
    - Y_i: i번째 관측값에 대한 종속변수의 값
    - X_i: i번째 관측값에 대한 독립변수의 값
    - β_0, β_1: 회귀 계수(regression coefficient)
    - ε_i: Y_i의 오차항을 나타내는 확률변수

선형회귀분석

- **다중선형회귀분석**
  - Y와 p개의 독립변수 $X_1,,,X_p$ 사이의 관계를 분석하는 통계적 방법론
  - 주요 이슈: 다중공선성(multicollinearity), 모형 선택(model selection)

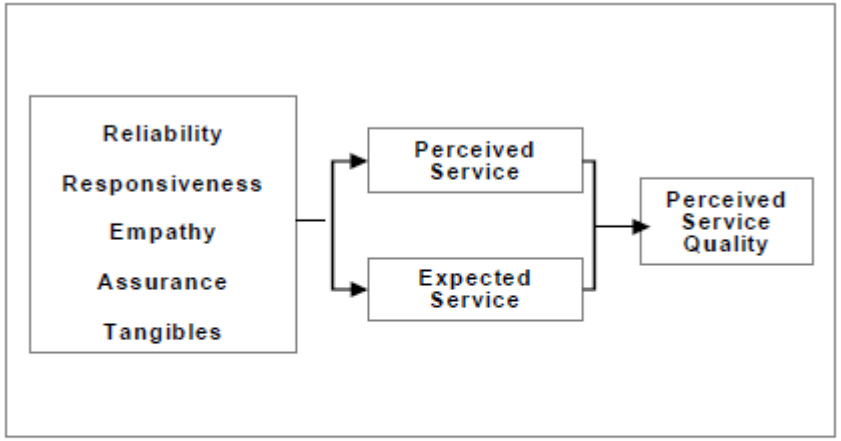
  - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i$
    - 단순회귀분석에서 독립변수 여러 개가 추가하여 확장된 형태
    - 검정 방법 등에서 근본적인 차이는 없음

선형회귀분석

- **선형회귀분석의 기본 가정**
  - 선형성(linearity)
  - 정규성(normality)
  - 외생성(exogeneity)
  - 조건부 독립성(conditional independence)
  - 등분산성(equal-variance)

발표 논문

- **서비스 품질의 측정: SERVQUAL 모형 (Parasuraman et al., 1988)**
  - 지각된 서비스 품질의 개념을 '서비스의 우수성과 관련된 소비자의 전반적인 판단이나 태도'로 정의
  - 포커스 그룹 인터뷰(FGI)를 통해 고객이 서비스 품질을 평가하는 10가지 차원을 추출한 뒤, 이중 중복되는 차원 등을 정리하여 최종적으로 5가지 차원 22개 항목을 추출
  - 신뢰성(reliability), 응답성(responsiveness), 공감성(empathy), 확신성(assurance), 유형성(tangibles)



자료원: Parasuraman, Zeithaml, and Berry(1988)

〈그림 1〉 SERVQUAL 모형

발표 논문

- **Emotion and Decision-making**
  - Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. Annual review of psychology, 66, 799-823.
  - The literature reveals that emotions constitute potent, pervasive, predictable, sometimes harmful and sometimes beneficial drivers of decision-making.
  - This paper proposes the emotion-imbued choice model, which accounts for inputs from traditional rational choice theory and from newer emotion research, synthesizing scientific models.
    - Bounded rationality (Herbert Simon, 1967): to refine existing normative models of rational choice to include cognitive and situational constraints

발표 논문

- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. Annual review of psychology, 66, 799-823.

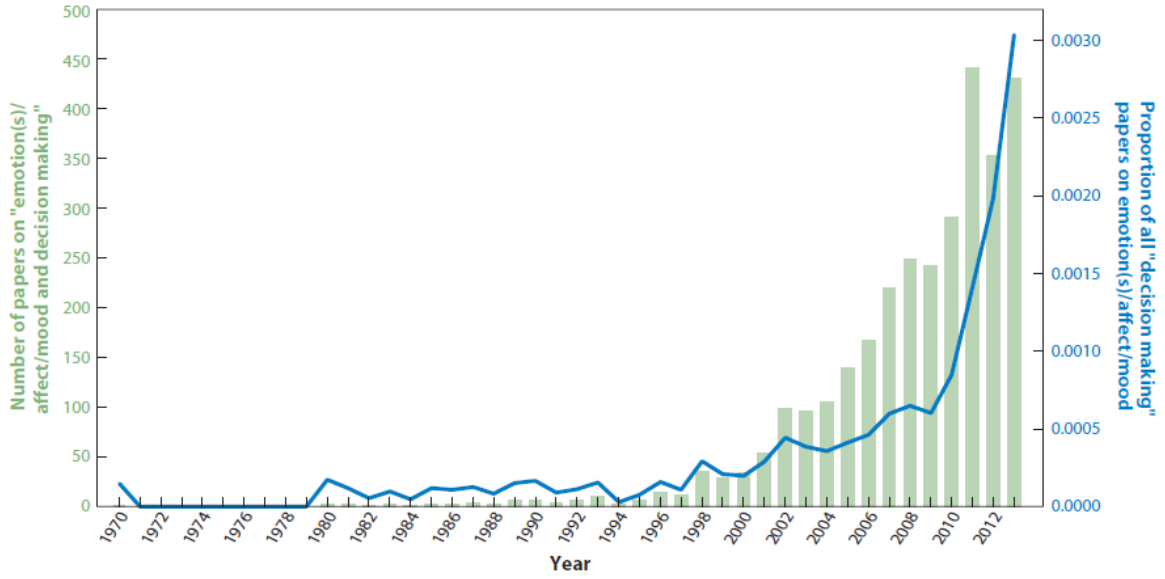
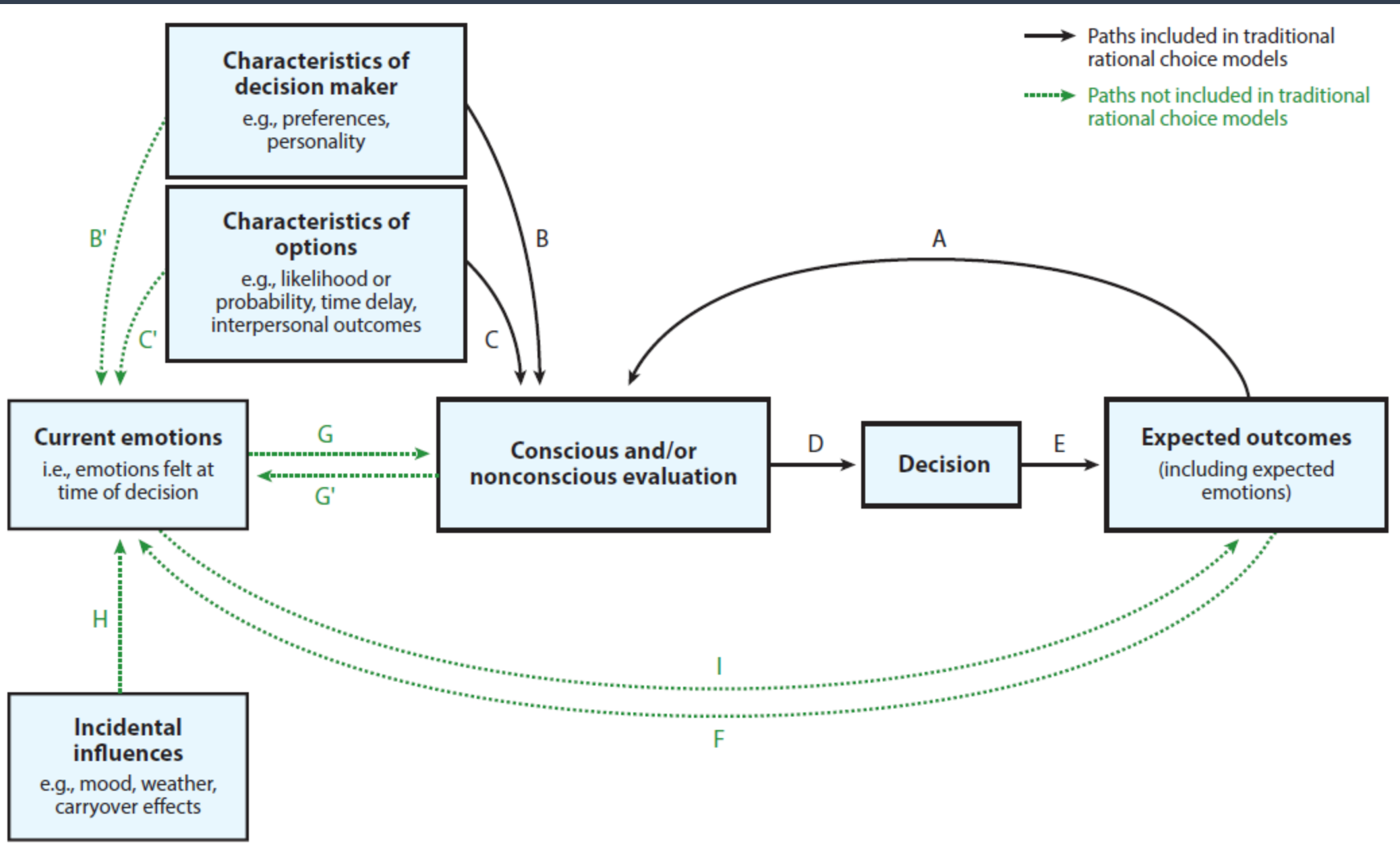


Figure 1

Number of scholarly publications from 1970 to 2013 that refer to "emotion(s)/affect/mood and decision making" (*green bars*) and proportion of all scholarly publications referring to "decision making" that this number represents (*blue line*).

발

- of


Paths included in traditional rational choice models
Paths not included in traditional rational choice models
Characteristics of decision maker
e.g., preferences, personality
Characteristics of options
e.g., likelihood or probability, time delay, interpersonal outcomes
B'
C'
B
C
A
Current emotions
i.e., emotions felt at time of decision
G
G'
Conscious and/or nonconscious evaluation
D
Decision
E
Expected outcomes
(including expected emotions)
H
I
F
Incidental influences
e.g., mood, weather, carryover effects

일반선형모형: 로지스틱 회귀분석

- **로지스틱 회귀분석**
  - 반응변수가 더미변수(dummy variable)일 경우
  - 일종의 classification으로 해석 가능
- **오즈(odds)**
  - 그룹 1에 속할 확률을 p, 그룹 2에 속할 확률을 1-p라고 했을 때,
  - $Odds = \frac{p}{1-p}$
  - 로지스틱 회귀분석에서 오즈를 활용하는 이유는 오즈가 취할 수 있는 값의 범위가 0~무한대
    - 음수의 값을 가지지 않는 등의 문제가 있기 때문에, 최종적으로는 Log(Odds)를 사용

일반선형모형: 로지스틱 회귀분석

- **로지스틱 회귀분석**

$$Log(Odds) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

  - 로지스틱 함수

$$p = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)]}$$

  - 일종의 비선형회귀분석: p와 predictor들의 관계
  - 연결 함수(link function): p와 predictor의 관계를 이어주는 사전에 정의된 함수
    - 로지스틱 회귀분석에서는 로지스틱 함수가 연결 함수의 역할

일반선형모형: 로지스틱 회귀분석

- **Independence of Irrelevant Alternatives (IIA)**
  - 로지스틱 회귀분석을 할 때 반드시 확인하여야 하는 가정
    - 특히 다중로짓 모형 등에서는 반드시 확인하여야 함.
  - “선택지 k에 비해 선택지 j를 선택할 오즈(odds)가 다른 선택지의 존재 여부에 영향을 받지 않아야 한다.”
  - “범주(category)의 추가나 제외가 남아있는 범주에 대한 설명변수의 상대적 위험도(relative risks)에 영향을 주지 않아야 한다.”

일반선형모형: 로지스틱 회귀분석

- **Independence of Irrelevant Alternatives (IIA)**
  - 사례: 파란 버스 – 빨간 버스 딜레마
    - 1. 출근할 때 이용할 수 있는 교통 수단으로 “자가용”과 “파란 버스” 두 가지가 있다고 가정
    - 2. “자가용”을 이용할 확률이 0.8일 때, “파란 버스”에 대한 오즈 = 4
    - 3. “파란 버스”와 노선이 똑같은 “빨간 버스” 도입 (새로운 범주의 추가)
    - 4. 버스의 색깔은 사실 별 상관이 없기 때문에, 각 버스를 이용할 확률이 0.1씩으로 쪼개질 것
    - 5. “자가용”과 “파란 버스”의 오즈는 8이 되므로 IIA 가정을 위배

일반선형모형: 프로빗 회귀분석

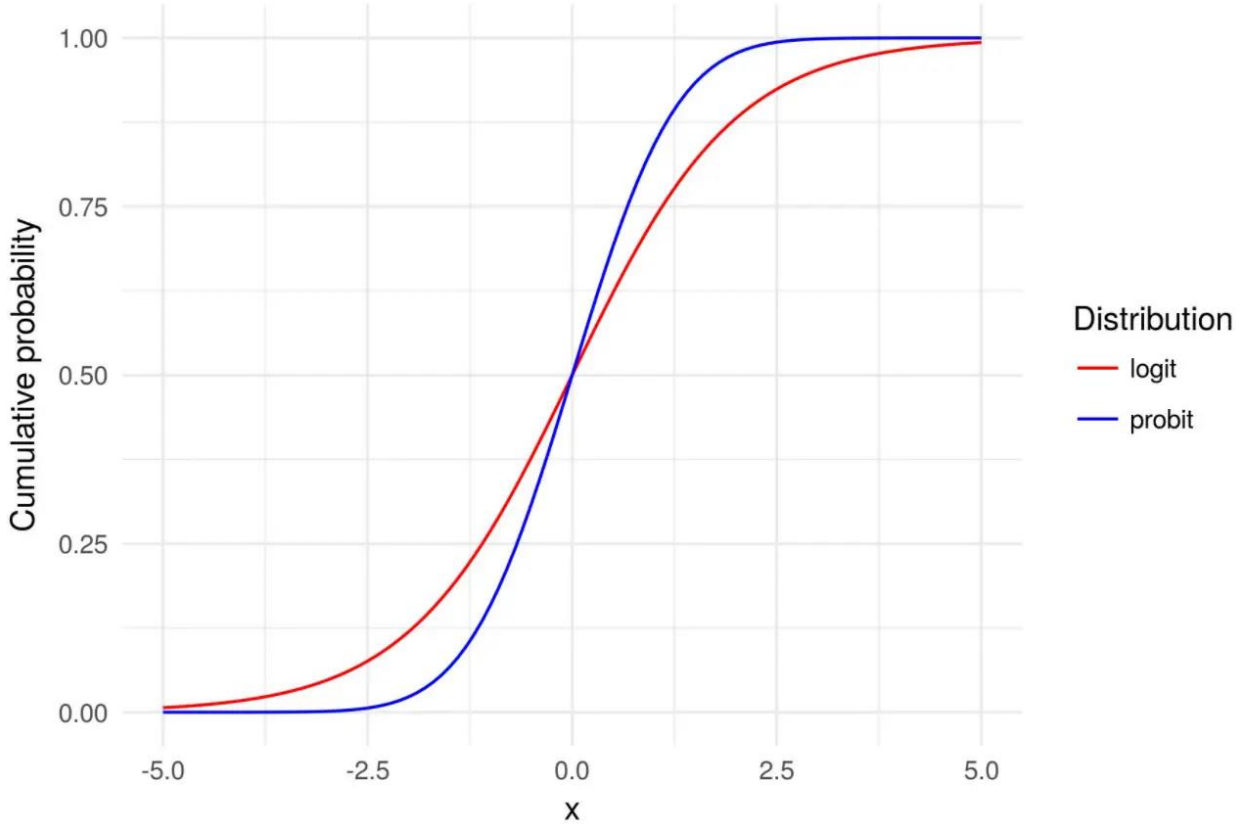
- **프로빗 모형(Probit model)**
  - 로지스틱 모형의 한계를 극복하기 위해 대안적 연결 함수로서 프로빗 함수를 제안

$$p = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

  - 표준정규분포의 누적분포함수를 이용하여 모형화
  - 종속변수 p에 미치는 한계 효과(marginal effect)는 설명변수에 대해 편미분함으로써 도출 가능

일반선형모형: 프로빗 회귀분석

- **프로빗 모형(Probit model)**
  - 로지스틱 함수와 프로빗 함수의 형태는 유사하지만, p=0.5 근처에서 로지스틱 함수가 더 완만
  - 데이터가 p=0.5에 몰려 있으면 두 모형의 차이가 거의 없지만, 그보다 멀리 떨어진 곳에 많이 있을 경우 두 모형의 분석 결과가 상이할 수 있음.



일반선형모형: 포아송 회귀분석

- **포아송 회귀분석**
  - 종속변수가 0, 1, … 와 같은 가산자료(count data)일 경우
  - 사건(event)의 기대 빈도(expected frequency)를 모형화
  - 가정
    - 1. 동일한 길이의 두 구간에서 사건 발생의 확률은 동일
    - 2. 어떤 구간에서 사건이 발생할 확률은 다른 구간에서 사건이 발생할 확률과 독립
    - 3. 매우 짧은 시간 내에 두 개 이상의 사건이 발생할 확률은 0
  - 포아송 분포: $\Pr(\mathrm{Y} = \mathrm{y}) = \frac{\lambda^{y}\exp(-\lambda)}{y!}$

알고리즘 마케팅

- **알고리즘 마케팅의 중요성**
  - 마케팅의 역사: 특정 비즈니스를 최적화하기 위한 원리, 기술, 프랙티스의 진화 과정
  - 수학적, 통계학적, 공학적 접근 방식을 통해 마케팅 문제를 해결하고자 하지만,
    - 한계: 데이터의 불완전성, 현실 마케팅 환경의 복잡성, 비즈니스 프로세스의 융통성 부족 등
  - 그럼에도 또 한 가지 트렌드의 등장: 디지털 마케팅 채널의 진화
    - 타깃 광고, 온라인/오프라인 매장에서의 동적 가격 책정, 추천 서비스, 온라인 광고 등
    - 고객은 개인화된 경험을 원하고, 따라서 수백만 개의 서로 다른 의사 결정이 필요
  - 전례 없는 수준의 자율성, 규모, 깊이를 갖춘 의사 결정과 이를 위한 마케팅 시스템 구축 필요
  - 마케팅 자동화와 통합적 관리를 위한 다양한 비즈니스 솔루션 필요

알고리즘 마케팅

- **알고리즘 마케팅의 주제**
  - 전통적인 마케팅: 마케팅 믹스(marketing mix; 4P) 조합의 변형이 중심
    - 전략: 회사가 제공하는 가치를 정의하고 마케팅 프로세스를 위한 방향을 정하는 최상위 단계의 장기적인 비즈니스 의사 결정
    - 프로세스: 회사의 연속적인 운영을 지원하기 위한 전술적 결정에 집중하는 전략의 실행
  - 알고리즘 마케팅: 마케팅 소프트웨어 시스템 안에서 비즈니스 목표를 자동으로 결정할 수 있을 정도로 자동화된 마케팅 프로세스
    - 마케팅 전략 수립과 프로세스 실행을 위한 데이터 기반 방법론의 제공이 목적

알고리즘 마케팅의 역사

- **마케팅 과학**
  - 사례: 온라인 광고
    - 스팸 메일, 단순한 배너 광고

    → 최적화 알고리즘을 바탕으로 한 개인화 광고

경향신문

PICK

야놀자·여기어때·부킹닷컴…맨 위 숙박상품 추천 아닌 광고

입력 2023.03.21. 오후 2:28 · 수정 2023.03.21. 오후 3:00 기사원문

13 13

[숙박플랫폼별 광고 상품 표시 실태]

구분	광고 유무	광고 비율*	광고 표시	광고 유형
야놀자	O	93개/100개(93%)	'AD' 클릭 시 광고 안내 화면이 나타남	'야놀자초이스', '지역초이스 플러스', '지역초이스' 등

알고리즘 마케팅의 역사

- **마케팅 과학**
  - 사례: 항공사 매출 관리
    - 1978년 미국 내 항공 관련 규제가 풀리면서 항공사들이 가격과 항로를 자유롭게 조정 가능
    - 각 항공사들이 고객 데이터베이스 관리에서의 최적화를 통해 서비스를 제안
    - 성수기 고가 전략/비성수기 저가 전략 등을 이용해 매출 극대화

  - 이처럼 마케팅 과학은 여러 산업에 걸쳐 다양하게 응용 가능
  - 1960년대 마케팅 믹스의 조합이 강조된 것도 마케팅 과학의 발달이 그 배경

프로그램 기반 서비스

- **프로그램 기반 마케팅 시스템**
  - 가격 책정이나 프로모션 관리와 같은 특정 비즈니스 프로세스를 구현하는 하나 이상의 서비스 제공자
  - 여섯 가지 분야의 프로그램 기반 서비스에 대해 다룰 것
    - 1. 판매 촉진(promotion)
    - 2. 광고(advertising)
    - 3. 검색(search)
    - 4. 추천(recommendation)
    - 5. 가격 책정(pricing)
    - 6. 상품 구성(assortment)

기술적, 예측적, 처방적 분석

- **기술적 분석**
  - 데이터 요약, 데이터 품질 측정, 관련성 분석
  - 예: 판매량 데이터 분석, 마켓 바스켓 분석(특정 제품과 같이 구매되는 다른 제품을 찾아내는 분석)
- **예측적 분석**
  - 관찰된 데이터 또는 결과 이전의 확률을 사용해 가능한 결과들을 예측
  - 예: 수요 예측, 프로모션에 따른 고객의 구매 확률 예측
- **처방적 분석**
  - 최적 의사 결정을 위해 의사 결정과 미래의 경과 사이의 의존성을 모델링
  - 예: 가격 할인 전략이 가져다 줄 이익의 예측을 통해 최적 가격 책정

경제적 최적화

- **경제적 모델 설계**
  - 경제적 모델: 비즈니스 목표에 관한 목표 함수와 제약 조건 등을 수식화
    - 예: $p_{opt} = argmax\ \pi(Q, p, x)$
  - 비즈니스 목표: 최적화할 수 있는 수리적 지표로 표현
    - 예: 회사의 이익과 고객의 효용 사이에서 절충해야 하는 경우
  - 데이터 수집: 모형의 복잡성 등에 영향
  - 모델의 세분화 수준

경제적 최적화

- **라그랑지안 함수와 최적화**
  - 제한조건(constraint)이 있는 최적화 문제
    - M개의 제한조건식을 모두 만족시키면서 $f(x)$를 극대화시키는 x를 찾는 문제

$$x^* = argmin\, f(x)$$
$$s.t.\, g_j(x) = 0 \quad (j = 1, \dots, M)$$

경제적 최적화

- **라그랑지안 함수와 최적화**
  - N+M개의 연립방정식을 풀어 아래의 미지수를 계산

$$h(x,\lambda) = h(x_1, x_2, \ldots, x_N, \lambda_1, \ldots, \lambda_M) = f(x) + \sum_{j=1}^{M} \lambda_j g_j(x)$$

$$\frac{\partial h}{\partial x_1} = \frac{\partial f}{\partial x_1} + \sum_{j=1}^{M} \lambda_j \frac{\partial g_j}{\partial x_1} = 0$$

$$\frac{\partial h}{\partial x_2} = \frac{\partial f}{\partial x_2} + \sum_{j=1}^{M} \lambda_j \frac{\partial g_j}{\partial x_2} = 0$$

$$\vdots$$

$$\frac{\partial h}{\partial x_N} = \frac{\partial f}{\partial x_N} + \sum_{j=1}^{M} \lambda_j \frac{\partial g_j}{\partial x_N} = 0$$

$$\frac{\partial h}{\partial \lambda_1} = g_1 = 0$$

$$\vdots$$

$$\frac{\partial h}{\partial \lambda_M} = g_M = 0$$

$$x_1, x_2, \ldots, x_N, , \lambda_1, \ldots, \lambda_M$$

머신 러닝: 지도학습

- **지도학습 (감독학습; Supervised learning)**
  - X를 이용해 Y 값을 예측해주는 함수의 학습, 즉, $P(Y|X)$의 분포를 학습하는 것
  - 학습 과정을 안내해주는 응답 변수 Y가 존재한다는 의미에서 지도/감독 학습
    - 분류(Classification): 응답 변수가 유한한 범주/등급인 경우
    - 회귀분석(Regression): 응답 변수가 무한한 연속 변수인 경우
  - 모수 모형 (Parametric model): 데이터의 분포가 몇 개의 모수에 의해 함수의 형태로 정의
    - 예: 정규 분포에 기반을 둔 선형회귀분석
  - 비모수 모형 (Nonparametric model)
    - 예: k-nearest Neighbor (kNN) 알고리즘

머신 러닝: 지도학습

- **최대 가능성 추정 (Maximum Likelihood Estimation; MLE)**
  - 가능성 함수 (Likelihood function)

$$L(\theta) = \Pr(y|X, \theta)$$

    - 모수 θ로 구성된 모형에 의해 정의된 분포에서 학습 데이터가 관찰될 확률(= 가능성)
    - 가능성 함수를 극대화하는 모수의 추정치를 찾는 것이 목표
    - Log 값을 취한 log-likelihood가 계산하기가 더 용이

$$\theta_{ML} = argmax \log[\Pr(y|X, \theta)] = argmin \ -\log[\Pr(y|X, \theta)]$$

머신 러닝: 지도학습

- **선형 모델**
  - 선형회귀분석
  - 로지스틱 회귀분석 또는 이진 분류 (binary classification)
  - 로지스틱 회귀분석 또는 다항 분류 (multi classification)

$$\Pr(Y = c|x) = \frac{\exp(X_c\beta)}{\sum_i \exp(X_i\beta)}$$

머신 러닝: 지도학습

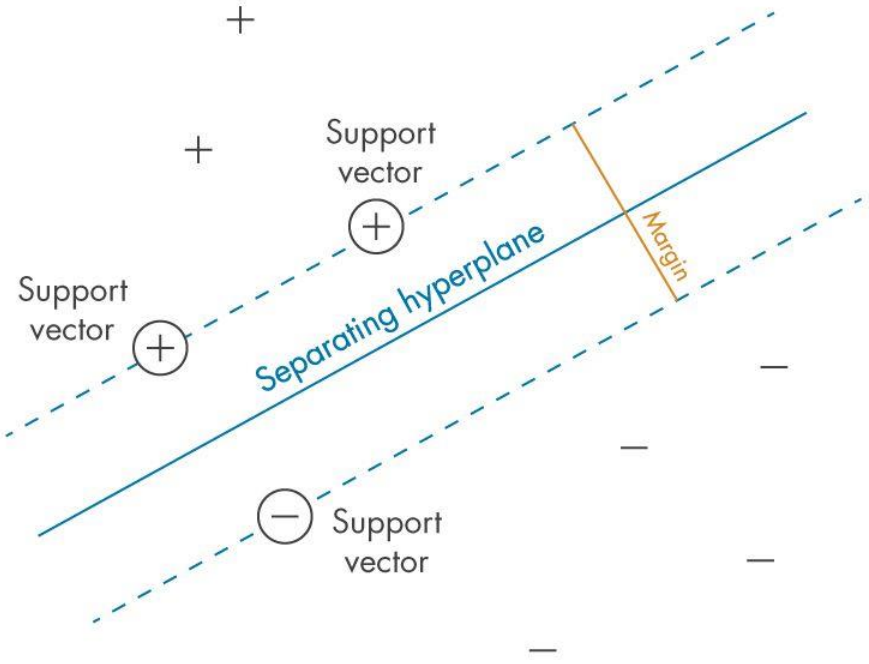
- **나이브 베이즈 분류 (Naïve Bayes Classification)**
  - 텍스트 분류 등에서 유용하게 활용

$$\Pr(Y = c|x) = \frac{\Pr(x|Y = c)\Pr(Y = c)}{\Pr(x)}$$

$$Y^* = argmax \Pr(Y = c|x) = argmax \Pr(x|Y = c)\Pr(Y = c)$$
$$= argmax \Pr(Y = c)\prod_{i=1}^{m}\Pr(x_i|Y = c)$$

머신 러닝: 지도학습

- **커널 기법 추정**
  - 비선형 모형의 하나
  - 특징 공간(feature space)들을 더 높은 차원의 특징 공간으로 변환시키는 방식으로 하나 또는 그 이상의 기존 특징의 비선형 함수로 표현되는 차원을 추가
  - 분류 문제 등에 있어 특징(feature)의 조합에 관하여 상당한 유연성을 제공
  - 예: 서포트 벡터 머신 (Support vector machine; SVM)



머신 러닝: 표현학습

- **비감독 학습(Unsupervised Learning)**
  - 감독학습이 응답변수와 설명변수의 관계를 $\Pr(Y|x)$라는 조건부확률로 모형화는 것이라면,
  - 비감독학습은 응답변수 없이 설명변수의 구조와 패턴을 $\Pr(x)$로 모형화하는 것

  - 데이터의 탐색적 분석에 있어 매우 유용하게 쓰일 수 있음.
    - 예: 고객 데이터의 클러스터링

머신 러닝: 표현학습

- **클러스터링 분석**
  - 기본적으로 비슷한 성질을 가진 개체를 묶어주는 과정
  - 클러스터(cluster) 내 데이터는 높은 유사성을 갖고, 클러스터 간에는 낮은 유사성을 갖도록 분리
  - 클러스터의 수 등을 정하는 데 있어 최적화를 어떻게 할 것인가가 이슈
  - 텍스트 마이닝에서 자주 사용되는 토픽 모형(topic model)도 비지도 학습의 일종으로서 클러스터링 분석으로 해석할 수 있음.

  - 응용의 예: 고객 세그멘테이션(segmentation)
    - 인구통계학적 특성, 고객 행동, 구매 행동, 추구 편익 등에 따라 세분화 가능

머신 러닝: 표현학습

- **클러스터링 분석**
  - K-평균 알고리즘 (K-means algorithm)
    - 주어진 데이터를 k개의 클러스터로 묶는 알고리즘
    - K: 데이터셋으로부터 찾을 것으로 예상되는 그룹(클러스터)의 수
    - Means: 각 데이터 포인트와 그 데이터가 속한 클러스터의 중심까지의 평균 거리
      - 기본적으로 평균 거리를 최소화하는 것이 목표
    - 목표함수의 예:

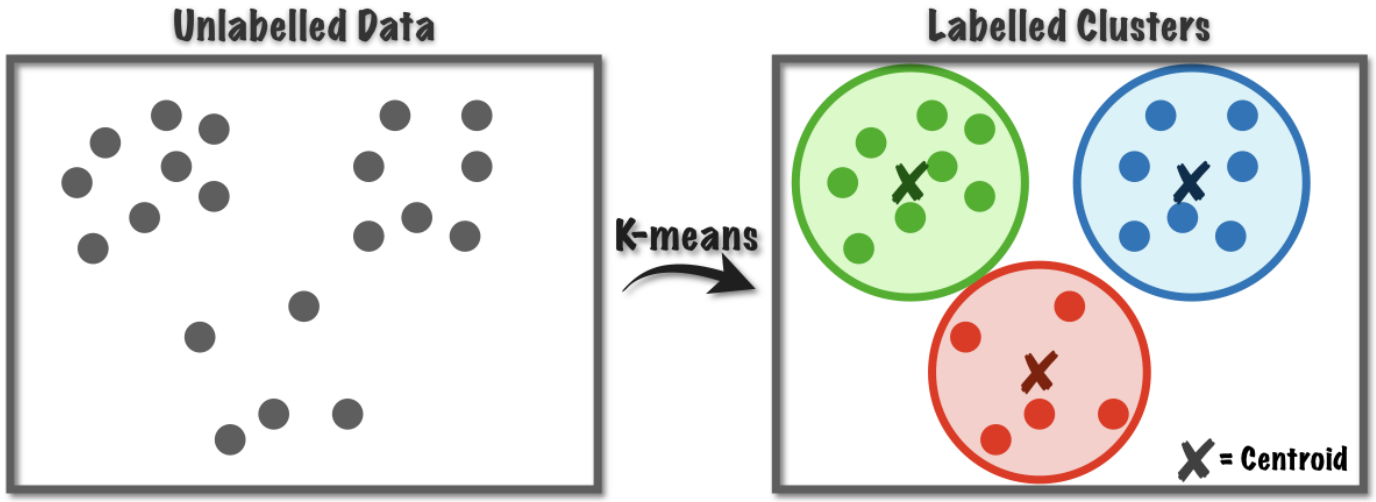
$$V = \sum_{i=1}^{k} \sum | x_j - \mu_i |^2$$

머신 러닝: 표현학습

- **클러스터링 분석**
  - K-평균 알고리즘 (K-means algorithm)
    - 알고리즘 업데이트 과정 (iterative method)
    - 1. 연구자가 부여한 k 값에 따라 k개의 임의의 중심점(centroid)을 선정
    - 2. 각 데이터들을 가장 가까운 중심점으로 할당 (초기 클러스터의 형성)
    - 3. 클러스터별 지정된 데이터를 기준으로 목표 함수를 계산하고 중심점을 업데이트
    - 4. 수렴이 될 때까지, 즉, 더 이상 중심점이 업데이트되지 않을 때까지 반복

머신 러닝: 표현학습

- **클러스터링 분석**
  - K-평균 알고리즘 (K-means algorithm)



머신 러닝: 표현학습

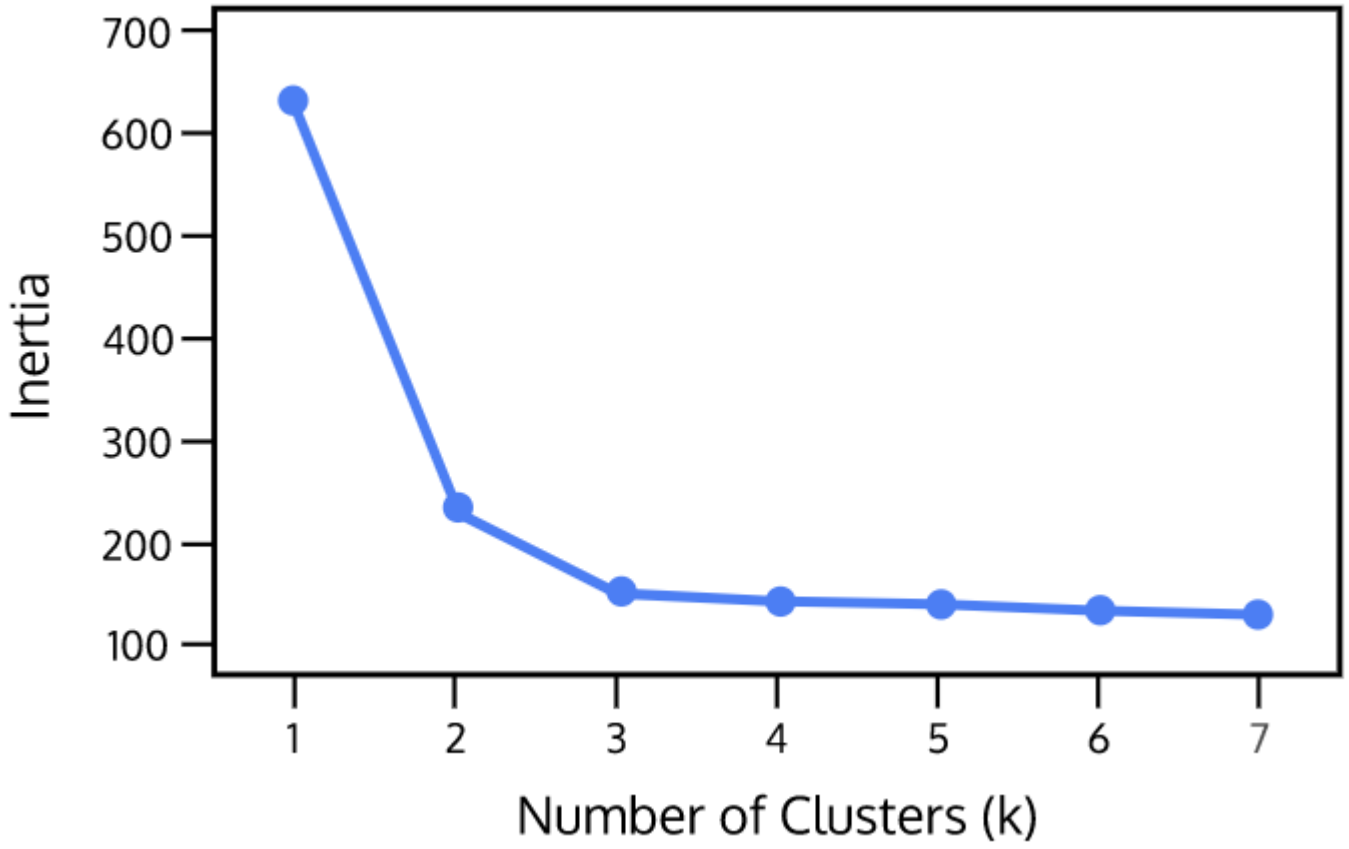
- **클러스터링 분석**
  - K-평균 알고리즘 (K-means algorithm)
    - 한계 1. 연구자가 임의로 k를 지정해주어야 한다.
    - 한계 2. 초기 임의의 중심점(centroid)에 결과가 의존할 가능성이 있다.
    - 한계 3. 이상치(outlier)에 민감한 편이다.

머신 러닝: 표현학습

- **클러스터링 분석**
  - K-평균 알고리즘 (K-means algorithm)
    - K 값을 잘 선택할 수 있는 방법이 없을까?
      - 1. Visualization
      - 2. Rule of thumb: $k = \sqrt{\frac{n}{2}}$
      - 3. Elbow method: k 값을 하나씩 늘려가면서 클러스터링 결과를 모니터링
        - 군집분석 결과가 사람이 해석하기에 의미 있게 나타나는가?
        - 클러스터링 모형을 가능도(likelihood)로 나타날 수 있다면 AIC, BIC 등을 참고

머신 러닝: 표현학습

- **클러스터링 분석**
  - K-평균 알고리즘 (K-means algorithm)
    - 2. Elbow method



머신 러닝: 표현학습

- **중요 요석 분석(Principal Component Analysis; PCA)**
  - 데이터의 압축되고 독립된 표현 방식을 찾는 방법
  - 특정 성질을 가지도록 데이터를 변형시켜 데이터의 구조를 설명하는 방법

  - 각 변수가 정규 분포를 따르고, “무상관관계”가 통계적 독립을 의미한다는 가정 하
  - 설명변수로 구성된 디자인 행렬(Design matrix; X)의 선형 변환을 통해 서로 독립인 요소로 구성된 대각 행렬(Diagonal matrix)을 도출
    - 예: 특이값 분해 (단일 가치 분해; Singular value decomposition)

$$X = U\Sigma V^T$$

머신 러닝: 표현학습

- **특이값 분해 (Singular Value Decomposition; SVD)**
  - 전치 행렬(transposed matrix)

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

  - 단위 행렬(identity matrix)

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad I^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad A \times I = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = A$$

  - 역행렬(Inverse matrix)

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} \qquad A \times A^{-1} = I$$

머신 러닝: 표현학습

- **특이값 분해 (Singular Value Decomposition; SVD)**
  - 직교 행렬(orthogonal matrix)

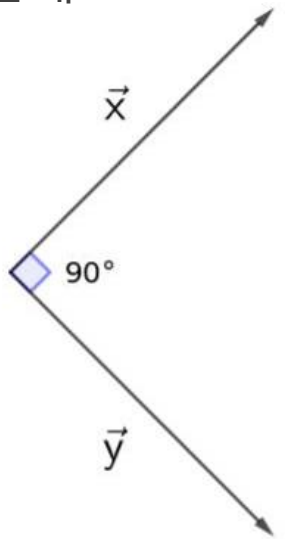
$$A \times A^{-1} = A^{-1} \times A = I \qquad A^{-1} = A^{T}$$

  - 대각 행렬(diagonal matrix)

$$\Sigma = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix}$$

머신 러닝: 표현학습

- **특이값 분해 (Singular Value Decomposition; SVD)**
  - 목적: 직교하는 벡터 집합에 대해 선형 변환 후에도 여전히 직교하도록 유지
  - PCA 맥락에서의 목적: 임의의 행렬을 정보량에 따라 여러 개의 층(layer)으로 분해
    - 벡터 간 직교(수직) 관계는 선형 독립(linear independence)을 의미



머신 러닝: 표현학습

- **특이값 분해 (Singular Value Decomposition; SVD)**
  - 정의: 임의의 m×n 차원의 행렬 A에 대하여 다음과 같이 행렬을 분해(decomposition) 하는 방법

$$A = U\Sigma V^T$$

  - A: m×n 임의의 행렬(rectangular matrix)
  - U: m×m 직교 행렬(orthogonal matrix)
  - Σ: m×n 대각 행렬(diagonal matrix)
  - V: n×n 직교 행렬(orthogonal matrix)

머신 러닝: 표현학습

- **특이값 분해 (Singular Value Decomposition; SVD)**

$$A = U\Sigma V^T$$

  - Σ: m×n 대각 행렬(diagonal matrix)

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

    - σ_i: 대각 행렬의 각 원소, 즉 대각 원소가 행렬 A의 특이값(singular value)
    - 특이값은 해당 차원의 정보량으로 해석할 수 있음.
    - 특이값은 큰 값 순서대로 내림차순으로 정렬됨.

머신 러닝: 표현학습

- **절단된 특이값 분해 (Truncated SVD)**

$$A = U\Sigma V^T$$

  - 지금까지 설명한 특이값 분해는 full SVD로서 행렬의 모든 정보를 이용
  - 실제 LSA에서는 full SVD가 아닌 truncated SVD를 사용
  - 즉, SVD 결과 도출된 행렬 Σ의 원소들 중 상위값 일부(t개)만 남겨 분석을 진행
    - 효과 1: 의미적으로 중요한 또는 설명력이 높은 요소만 고려 (중요하지 않은 정보 삭제)
    - 효과 2: 계산 비용을 줄이기 위함.

Singular Value Decomposition

- **Example**
  - Calculate the SVD of A, $U\Sigma V^T$, where $\mathrm{A} = \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix}$.
  - **1. We need to compute the singular values by finding eigenvalues of $A^T A$.**

$$A^T A = \begin{pmatrix} 4 & 3 \\ 0 & -5 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix} = \begin{pmatrix} 25 & -15 \\ -15 & 25 \end{pmatrix}$$

Singular Value Decomposition

- **Example**

$$Av = \lambda v$$

  - What is eigenvalue?
    - 고유 벡터(eigenvector): 행렬 A의 선형 변환 결과가 자기 자신의 상수(constant) 배가 되는 0이 아닌 벡터 (v)
    - 고유값(eigenvalue): 그 때의 상수배의 값 (λ)
  - How do we get eigenvalues?
    - $(A - \lambda I)v = 0 \Rightarrow (A - \lambda I)$의 역행렬이 존재하면 $v = 0$
    - $det(A - \lambda I) = 0 \Rightarrow$ 역행렬이 존재하지 않을 조건
    - 참고: 정방 행렬일 경우에만 고유값과 고유 벡터를 계산할 수 있음.

Singular Value Decomposition

- **Example**

$$A^T A = \begin{pmatrix} 4 & 3 \\ 0 & -5 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix} = \begin{pmatrix} 25 & -15 \\ -15 & 25 \end{pmatrix}$$

$$\det(A^T A - \lambda I) = \lambda^2 - 50\lambda + 400 = (\lambda - 10)(\lambda - 40) = 0$$

  - The eigenvalues of $A^T A$: 40, 10
  - Singular values: $\sigma_1 = \sqrt{40}\ and\ \sigma_2 = \sqrt{10}$

Singular Value Decomposition

- **Example**
  - **2. We need to find the right singular vectors (the columns of V in $U\Sigma V^T$) by finding eigenvectors of $A^T A$.**
    - We can proceed by finding the left singular vectors (the columns of U) instead.
    - Since $A^T A$ is symmetric, the eigenvectors will be orthogonal.
    - For $\lambda = 40$,

$$A^T A - 40I = \begin{pmatrix} -15 & -15 \\ -15 & -15 \end{pmatrix}$$

    - In order to find the eigenvector, we need to find the null space of a matrix where

$$\begin{pmatrix} -15 & -15 \\ -15 & -15 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

Singular Value Decomposition

- **Example**
  - **2. We need to find the right singular vectors (the columns of V in $U\Sigma V^T$) by finding eigenvectors of A^TA.**
    - $A^TA - 40I$ can row-reduces to $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$.
    - If we solve

    $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$, then $v_1 = -v_2$.

    - For $v_2 = 1$, $v_1 = -1$. Then, we can get an eigenvector of eigenvalue 40 by converting it to unit vector $\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

Singular Value Decomposition

- **Example**
  - Now we know that

$$A = U\Sigma V^T = U\begin{pmatrix}\sqrt{40} & 0\\ 0 & \sqrt{10}\end{pmatrix}\begin{pmatrix}-1/\sqrt{2} & 1/\sqrt{2}\\ 1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}$$

  - Principal components analysis (PCA)의 관점에서 보면 데이터의 분산을 가장 잘 설명하는 상위 요소들이 SVD 계산 결과에서 상위에 위치한 특이값들에 해당

Singular Value Decomposition

- **Example**
  - **3. We can compute U.**

$$U = AV\Sigma^T = \begin{pmatrix} -1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix}$$

  - Finally, the result of SVD is as follows:

$$A = \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix} = U\Sigma V^T = \begin{pmatrix} -1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix} \begin{pmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{pmatrix} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Singular Value Decomposition

- **Example**
  - Finally, we can get the result of SVD of A.

$$A = \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix} = U\Sigma V^T = \begin{pmatrix} -1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix} \begin{pmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{pmatrix} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

  - LSA는 DTM 또는 TF-IDF 행렬을 truncated SVD로 분석함으로써 차원을 축소시키는 한편, 문서 내 단어들에 내재된 잠재적인 의미를 추출
  - LSA 관점에서의 해석
    - $U \times \Sigma$: 문서와 토픽 간의 행렬 ⇒ 문서 간 유사성 계산에 활용
    - $\Sigma \times V^T$: 토픽과 단어 간의 행렬 ⇒ 단어 간 유사성 계산에 활용

머신 러닝: 표현학습

- **중요 요석 분석(Principal Component Analysis; PCA)**
  - PCA 결과 도출된 중요 벡터가 데이터 분산의 크기에 따라 정렬
    - 고분산 차원이 저분산 차원보다 좀 더 많은 정보를 담고 있다는 점에서 의미
    - 이를 데이터의 차원 축소(dimension reduction)에 응용할 수 있음.
      - 예: 임베딩 후 지각도(perceptual map)를 시각화하는 경우
    - 마케팅 및 사회과학 데이터는 기본적으로 sparsity나 noise의 문제가 더 많은 편인데, 따라서 PCA 등을 이용한 데이터 축약 표현 기법이 매우 유용하게 쓰일 수 있음.

선택 모형(choice model)

- **고객 선택 이론의 기초**
  - 고객 선택의 이해와 예측은 마케팅과 미시경제학의 가장 중요한 문제 중 하나
    - 앞서 배운 로지스틱 회귀분석 또는 다항 로지스틱 회귀분석이 대표적
    - 연구의 목적, 데이터의 특성 등에 따라 다양한 계량경제학 모형이 존재

  - 소비자(의사결정자)가 일관된 방법으로 최상의 옵션을 선호한다고 가정
    - 예: Consumer n will choose Y_{nj} if $u(Y_{nj}) > u(Y_{nk})$ for all k

선택 모형(choice model)

- **고객 선택 이론의 기초**
  - 소비자에게 주어진 옵션의 효용에 비례하는 가상의 수리적 지표를 구성
    - 예: $Y_{nj} = Y(x_{nj}, h_{nj})$
      - x_{nj}: 알려진 요소, 혹은 데이터로 수집할 수 있는 요소 / h_{nj}: 미지의 요소
      - 대안: $Y_{nj} = Y(x_{nj}) + \varepsilon_{nj}$
  - 랜덤 효용 모형(Random utility model)으로서 확률적으로 표현 가능
    - $P_{ni} = \Pr(Y_{ni} > Y_{nj}, \forall j \neq i) = \Pr(Y(x_{ni}) + \varepsilon_{ni} > Y(x_{nj}) + \varepsilon_{nj}, \forall j \neq i)$

생존 분석

- **생존 분석(Survival Analysis)**
  - 특정 사건이 일어날 때까지 걸리는 시간을 추정하는 방법론
  - 마케팅 및 미시경제학 문제에서는 고객의 선택과 선택확률을 추정하는 것도 중요하지만, 고객의 구매나 이탈과 같은 특정 사건이 일어날 때까지 걸리는 시간을 알아내는 것 또한 중요한 문제
  - 원래 의학 및 생물학 분야에서 먼저 발달한 방법론
    - 예: 의료 행위와 특정 이벤트(예: 죽음) 간의 시간을 예측
  - 마케팅 문제에서는 프로모션(예: 광고)의 효과 등을 측정하는 데 유용

생존 분석

- **생존 분석(Survival Analysis)**
  - 생존 함수: 시작 시간부터 시간 t까지 생존할 확률
    - 생존 함수의 값이 급격히 감소한다는 것은 대부분의 고객이 특정 이벤트를 곧 경험한다는 것
  - 고객 생존시간이 T일 때의 확률 분포

$$F(t) = \Pr(T \leq t) = \int_0^t f(\tau) d\tau$$

  - 확률 분포 기반의 생존 함수

$$S(t) = \Pr(T > t) = 1 - F(t)$$

생존 분석

- **생존 분석(Survival Analysis)**
  - 위험 함수(Hazard function): 생존 함수가 생존 확률(이벤트가 일어나지 않을 확률)을 다룬다면, 위험 함수는 이벤트가 발생할 위험을 모형화

$$h(t) = \lim_{dt \to 0} \frac{\Pr(t < T \leq t + dt \mid T > t)}{dt}$$

    - 시간 t까지 이벤트가 발생하지 않았다는 가정 하에, t와 t+dt 사이의 짧은 시간에 이벤트가 발생할 확률

생존 분석

- **생존 분석 (Survival Analysis)**
  - 위험 함수 (Hazard function)
    - 식을 정리하여 생존함수 S(t)에 대응시키면 다음과 같이 정리할 수 있다.

$$h(\mathrm{t}) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt}\log\big(1 - F(t)\big) = -\frac{d}{dt}\log(S(t))$$

$$S(t) = \exp\big(-H(t)\big)$$

생존 분석

- **생존 분석 (Survival Analysis)**
  - 콕스 비례 위험 모형(Cox proportional hazard model): 위험 함수를 이용한 기본적인 회귀분석 방법
    - 기본적으로 설명변수 x가 위험 함수(생존 함수)에 어떤 영향을 미치는지 추정

$$h(t|w,x) = h_0(t)\exp(w'x)$$

  - 기본 위험(baseline hazard; $h_0(t)$)
    - 확률 분포를 부여하고 모수적으로 추정하거나, 아예 비모수적으로 추정하는 것도 가능
    - 기본 위험으로부터 $\exp(w'x)$만큼 비례하여 증가하는 구조

생존 분석

- **생존 분석 (Survival Analysis)**
  - Seetharaman, P. B., & Chintagunta, P. K. (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. Journal of Business & Economic Statistics, 21(3), 368-382.
    - 1. Continuous-time vs. Discrete-time
    - 2. Parametric specifications of baseline hazard: exponential, Weibull, Erlang-2, log-logistic, expo-power
    - 3. Single-risk formation vs. Competing-risks formation
    - 4. Parametric vs. Nonparametric (e.g., step function)
    - 5. Homogeneous vs. Heterogeneous

Data Ownership

- 오늘날 대부분의 플랫폼에서는 data를 수집하는 문제에 대해 관련 규정을 따로 두고 있음.

- Naver 예: https://policy.naver.com/policy/service_en.html
  - "Users should not use Naver Services in ways which do not correspond with Naver's purpose of providing its services, which is based on the premise that it is the certain user (person) himself or herself that actually uses Naver's Services, such as joining or attempting to join as a Member, logging in or attempting to log in to Naver service, posting or attempting to post on Naver Service, engaging in communication through Naver service (e.g., e-mail, messages), collecting ID or posts on Naver Services, searching with certain query terms on Naver search service or selecting (i.e., click) certain item among the search results by using automated means (e.g., macro program, robot (bot), spider, scraper, etc.) without the prior permission from Naver, or attempt to disable Naver's technical measures to block such abuse of Naver Services (e.g., accessing by constantly changing IPs, making a detour or disabling Captcha through external solution, etc.)."

Data Ownership

- 오늘날 대부분의 플랫폼에서는 data를 수집하는 문제에 대해 관련 규정을 따로 두고 있음.

- Instagram 예: https://help.instagram.com/581066165581870
  - “You can't attempt to create accounts or access or collect information in unauthorized ways. This includes creating accounts or collecting information in an automated way without our express permission.”

Data Ownership

- **뉴스 기사 저작권 문제**
  - **2020 뉴스 저작권 이용가이드북, 한국언론진흥재단**

  - **저작권:** 시, 소설, 뉴스, 음악, 미술, 영화 등과 같은 저작물에 대하여 창작자가 가지는 권리
  - 예를 들어 소설가가 소설 작품을 창작한 경우, 원고 그대로 출판·배포할 수 있는 복제·배포권뿐만 아니라 그 소설을 영화나 번역물 등과 같이 다른 형태로 제작할 수 있는 2차적 저작물 작성권, 연극 등으로 공연할 수 있는 공연권, 방송물로 만들어 방송할 수 있는 방송권 등 여러 가지 권리를 가지게 됨.
  - **뉴스 저작물:** 시사보도, 여론형성, 정보전파 등을 목적으로 발행되는 정기간행물, 방송 또는 인터넷 매체 등에 수록된 저작물을 의미하며, 뉴스는 저작권법 제4조 1항 1호에 명시된 어문저작물에 해당하며 법에 의해 보호를 받음.

Data Ownership

- **뉴스 기사 저작권 침해 사례**
  - 1: 뉴스 기사의 출처를 밝히더라도 언론사의 허락 없이 기사를 온라인, SNS 등에 게시하는 것은 무단전재로서 불법 이용에 해당
  - 2: 외부인들이 볼 수 없는 사내 게시판이라 하더라도 임의로 사내 뉴스 데이터베이스를 구축하여 게재, 배포하는 것은 저작권법 위반 행위
  - 3: 업무상 목적으로 뉴스를 스크랩하여 다수의 사람들에게 배포하는 것은 뉴스저작권 침해
  - 4: 개인 블로그나 인터넷 카페와 같은 공중이 볼 수 있는 환경에 뉴스저작물을 무단으로 복사하여 올리거나 이를 재배포하는 것은 저작권법 위반
  - 5: 영리적 목적이 아닌 공리, 비영리 목적이라도 저작권자의 허락 없이 무단으로 뉴스 저작물을 출판 및 인쇄하는 행위는 뉴스저작권 침해에 해당

Data Ownership

- **올바른 뉴스저작물 이용 방법**
  - 1: 일정 요건을 충족할 경우, 뉴스기사의 일부분을 인용해 새로운 창작물 작성 가능
    - 보도, 비평, 교육, 연구 등을 위한 인용일 것
    - 인용 저작물과 피인용 저작물이 양적, 질적으로 주종관계가 성립하여 분명하게 구별될 것
    - 저작물 이용의 목적과 방법이 건전한 사회통념에 비추어 판단할 때 공정한 관행에 합치되며, 출처를 표시할 것
  - 2: 홈페이지, 블로그 등에 게재할 경우, 해당 언론사 홈페이지로 연결되는 '단순링크' 방식을 사용
  - 3: 한국언론진흥재단 등을 통해 이용계약을 체결하여 사용료 지불

개인 프로젝트 안내

- **과제 목적**
  - 석사 과정: 수업 내용과 추가적인 문헌 연구를 바탕으로 구체적인 연구 주제 제안
  - 박사 과정: (석사 과정 목적) + 데이터를 통한 분석 결과 제시

- **연구 주제**: 본인이 관심 있는 분야

  - 단, [알고리즘 마케팅]의 수업 내용이 어느 단계에서든 반영되어야 할 것
- **데이터**: 자유
- **방법론**: 자유

개인 프로젝트 안내

- **무엇이 흥미로운 연구 주제인가?**
  - Theoretical implication
    - 산업공학, 경영학 등 관련 문헌에 이론적으로 기여할 수 있는가?
    - 해당 주제가 최근 학계로부터 관심을 받고 있는가?
    - **Key literature의 제시 + Research gap의 제시**
  - Practical implication
    - 기업이 실질적으로 수익을 높이는 데 기여할 수 있는가?
    - 기업 등이 실무적으로 활용할 가치가 높은가?

개인 프로젝트 안내

- **프로젝트 주제 발표**
  - 일시: **4월 6일(목)**
  - 내용

    - 석사 과정: 문헌 연구 요약 발표 (예: 본인이 follow할 key paper의 요약 발표)

    - 박사 과정: 주제 개요 + (예상되는) 시사점 + 주요 선행 연구 + 데이터 + 방법론
  - 분량: 발표 시간 10~15분