

[ML] Latent Factor(Variable) 에 대하여

통계 수정 삭제

soyoun9798 · 6분 전

 0

ML latentfactor

ML/DL



▼ 목록 보기

8/8  



1. Latent Variable(잠재 변수) 정의

잠재 변수(latent variable)는 구성개념(Construct)이 직접적으로 관찰되거나 측정이 되지 않는 변수를 의미한다.

사회과학적으로 잠재변수 자체로 전체 통계량을 측정하는 것은 불가능하기 때문에, 관측변수(observed variable)에 의해 간접적으로 통계 측정(수학적 모델링)을 수행해 추론한다.

이 때 latent variable models은

이 모델은 many disciplines 에서 사용되곤 하는데,

심리학, 인구학, 경제학, 공학, 의학, 물리학, 기계 학습/인공지능, 생물 정보학, 화학, 자연 언어 처리, 경제 측정, 관리, 사회 과학등이 포함된다.

2. Latent Variable(잠재 변수) 의 분류

latent variables 은 물리적인 특성을 가지거나 추상적인 개념을 가지는 2가지로 분류할 수 있다.

물리적 측면의 latent variable은 원칙적으로 측정될 수 있지만, 실용/실질적인 이유로 측정되지 않는 경우이다. 이때의 latent variables을 hidden variables 라는 용어로 부른다(변수가 의미 있지만 관측할 수 없다는 사실을 반영)

추상적 개념의 latent variables 은 범주, 행동, 정신 상태 또는 데이터 구조와 같은 것에 해당된다. 이 경우 hypothetical variables 또는 hypothetical constructs 라는 표현을 사용한다.

3. Latent Variable(잠재 변수) 의 기능

잠재 변수를 사용하면 데이터의 차원을 줄일 수 있습니다. 많은 관측 가능한 변수를 모델에 통합하여 기본 개념을 나타낼 수 있으므로 데이터를 더 쉽게 이해할 수 있다.

또한 잠재 변수는 실제 세계에서 관측 가능한 "sub-symbolic" data를 모델링된 세계의 "symbolic" data와 연결한다.

4. 예시

1. 심리학

factor analytic methods(요인 분석 방법)에 의해 생성된 잠재 변수는 일반적으로 "공유" 분산 또는 변수가 함께 "이동"하는 정도를 나타냅니다. 상관 관계가 없는 변수는 혼한(?) factor model을 기반으로 하는 잠재적 구조를 만들 수 없습니다.

▼ 예시들

- "빅 5 성격 특성" factor analysis로 추론된
 - extraversion
 - spatial ability
 - wisdom ("지혜를 평가하는 가장 지배적인 두 가지 방법 중 하나는 지혜와 관련된 성과와 잠재적인 변수 측정이다.")
 - Spearman's g, 또는 사이코메트릭의 일반 지능 인자

2. 경제

경제 분야의 잠재적 변수로는 삶의 질, 기업 신뢰, 사기, 행복, 보수 등이 있다.

그러나 이러한 잠재적 변수를 관측 가능한 다른 변수와 연결시키면 관측 가능한 변수의 측정값으로부터 잠재 변수의 값을 추론할 수 있다.

삶의 질은 직접적으로 측정할 수 없는 잠재 변수이기 때문에 관찰 가능한 변수는 삶의 질을 추론하는 데 사용된다. 삶의 질을 측정하기 위한 관측 가능한 변수로는 **부, 고용, 환경, 신체적 및 정신적 건강, 교육, 레크리에이션 및 여가 시간, 사회적 소속** 등이 있다.

3. 의학

Latent-variable methodology 은 많은 의학 분야에서 사용된다. 잠재 변수 접근법에 자연스럽게 도움이 되는 문제의 클래스는 시간 척도(예: 참가자 연령 또는 연구 기준 이후 시간)가 연구 중인 특성과 동기화되지 않는 **longitudinal studies** 이다.

이러한 연구의 경우, 연구 중인 특성과 동기화된 관찰되지 않은 시간 척도는 잠재 변수를 사용하여 관찰된 시간 척도의 변환으로 모델링될 수 있다.

그 예로는 **disease progression modeling** 과 **modeling of growth**이 있다.

5. 추론 방법

잠재 변수를 사용하고 잠재 변수가 있는 경우 추론을 허용하는 다양한 모델 클래스 및 방법론:

- linear mixed-effects models and nonlinear mixed-effects models
- Hidden Markov models
- Factor analysis

분석론 과 추론 방법론 들:

- **Principal component analysis**
- **Instrumented principal component analysis**
- **Partial least squares regression**
- **Latent semantic analysis and probabilistic latent semantic analysis**
- **EM algorithms**
- **Metropolis–Hastings algorithm**

****Bayesian statistics**** 은 latent variables 를 추론하는데 자주 사용된다

- **Latent Dirichlet allocation**

- **The Chinese restaurant process**

: discrete-time 확률 프로세스로 latent categories 의 객체 할당에 대한 사전 분포를 제공하기 위해 사용된다.

- **The Indian buffet process**

: 행수가 유한하고, 열수가 무한한 sparse binary matrices에 대한 확률 분포를 정의하는 확률 프로세스로, objects 에 대한 latent binary features 에 대한 사전 분포를 제공하기 위해 사용된다.

추론 방법론을 더 자세히 알고 싶은 분들은 아래 wiki의 참조들을 보시면 됩니다!

ref

- https://en.wikipedia.org/wiki/Latent_variable
- <https://medium.com/analytics-vidhya/latent-probabilistic-modeling-part-1-a541722419af>



sosimeow



이전 포스트

[Scikit learn 복습 1] Decision tree 학습, 평가 code & 머신러닝 프로세스 개념

0개의 댓글

댓글을 작성하세요

댓글 작성

