가상 면접 사례로 배우는

머신러닝 시스템 설계 기초

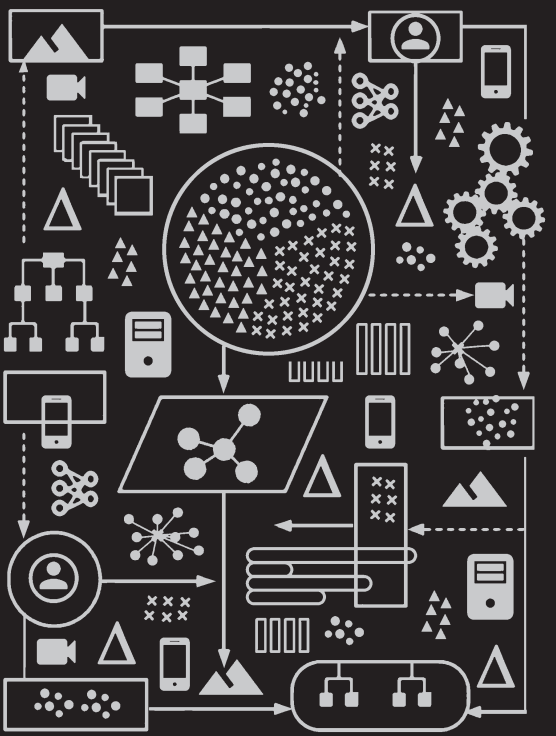
Machine Learning System Design Interview

Machine Learning System Design Interview

가상 면접 사례로 배우는 머신러닝 시스템 설계 기초

전자책 1쇄 발행 2024년 3월 5일 **지은이** 알리 아미니안, 알렉스 쉬 **옮긴이** 최종일 **펴낸이** 한기성 **펴낸곳** (주)도서출판인사이트
편집 백혜영 **등록번호** 제2002-000049호 **등록일자** 2002년 2월 19일 **주소** 서울특별시 마포구 연남로5길 19-5 **전화** 02-322-5143 **팩스** 02-3143-5579 **블로그** https://blog.insightbook.co.kr **이메일** insight@insightbook.co.kr **ISBN** 978-89-6626-439-1

프로그래밍 인사이트



가상 면접 사례로 배우는

머신러닝 시스템 설계 기초

알렉스 쉬 · 알리 아미니안 지음 | 최종일 옮김

인사이트

가장 친한 친구였던 닐루파에게
– 알리 아미니안

줄리아에게
– 알렉스 쉬

차례

옮긴이의 글

사람들은 책을 통해 정보를 습득하고, 현실에 적용하고, 자신의 역량이 발전하길 바란다. 좋은 책은 실용적이고 통찰력 있는 최신 정보를 제공하고 독자들의 성장을 돕는다. 번역하면서 이 책의 세 가지 장점을 느꼈다.

첫 번째로 이 책은 '최신 기술을 잘 설명'한다. 실제 서비스되고 있는 머신러닝 사례, 사례별 개발 사이클의 설명, 저자의 노하우가 담긴 상세한 팁이라는 세 가지 축으로 이야기를 풀어 간다. 실제 사례는 언제나 관심을 끈다. 머신러닝을 공부하는 독자라면 누구나 흥미를 느낄 유튜브 동영상 추천, 구글 스트리트 뷰, 에어비앤비 예약, 페이스북 친구 추천의 핵심 기능을 압축하여 보여 준다. 각 서비스의 개발 과정을 요구사항 명확화, 머신러닝 작업으로 문제를 구조화, 데이터 준비, 모델 개발, 평가, 서빙의 순서로 진행하며 하나의 서비스가 완성되어 가는 모습을 체험하게 한다. 각 축에서는 다양한 분석 기법, 모델을 설명하며 최신 정보를 전달한다. 그리고 머신러닝 유형, 모델, 평가 지표를 선택하는 설명에는 저자의 노하우가 알차게 담겨 있다. 세 가지 정보를 촘촘하게 엮는 방식으로 이 책은 지루하지 않고 입체감 있게 지식을 전달한다. 번역을 하면서 기술 서적이 이렇게 쉽게 읽힌다는 점에 놀랐던 기억이 난다.

두 번째는 '실용성'이다. 책의 내용은 가상 면접 형태로 시작한다. 면접관과 지원자가 등장하는 면접 시나리오는 실제 머신러닝 면접에서 나오는 질문으로 구성되어 있다. 저자의 말대로 머신러닝 면접에서는 데이터 파이프라인, 피처 엔지니어링, 시스템 설계 등의 전체적(엔드투엔드)인 지식을 묻는다. 면접을 준비 중인 개발자들은 반드시 이해해야 할 지식이다. 면접 시나리오로 잘 정리한 질문은 이어지는 본문에서 차근차근 설명된다. 그리고 개발 주기 전체를 다루며 머신러닝 시스템 전체 범위의 지식을 제공하는 점도 이 책의 장점이다.

실제 프로젝트에서 경험하게 되는 선택과 그 선택의 기준을 합리적으로 설명한다.

마지막으로 '통찰력'이다. 최신 기술을 다루는 일반적인 책들이 갖지 못한 이 책만의 강점이라고 생각한다. 안면 인식 소프트웨어는 흑인 여성보다 백인 남성을 더 잘 인식한다. 간혹 우리는 프로젝트 자원의 제약을 핑계로 문제를 못 본 척 넘기기도 한다. 그래서 저자는 각 장에서 독자에게 계속 질문한다. 학습 데이터 자체에 편향이 있다면 어떻게 처리할 건가? 당신은 편견에서 자유로울 수 있나? 검색과 활용의 트레이드 오프에 당신의 기준은 무엇인가? 윤리 준수에 대한 당신의 생각은 무엇인가? 머신러닝이라는 최신 기술 영역에서 아직 명확한 기준이 없는 생각거리를 상기시킨다. 이런 질문들이 머신러닝 커리어의 다면적인 역량 개발에 도움이 되리라 믿는다.

이 책은 머신러닝을 공부하는 모든 이에게 큰 도움이 되는 책이라고 확신한다. 그리고 그런 책을 번역한 것에 살짝 자부심을 가져 본다.

지은이의 글

머신러닝(machine learning) 시스템 설계 면접을 더 잘 준비하기 위한 길에 동행하게 되어 기쁘다. 머신러닝 시스템 설계는 모든 머신러닝 면접에서 가장 까다로운 주제 중 하나이며, 면접을 통과하려면 반드시 준비해야 한다.

머신러닝 시스템 설계 면접이란 무엇인가?

일반적으로 머신러닝 시스템 설계 면접은 머신러닝 시스템 설계 및 구현과 관련된 업무 지원자가 보게 된다. 데이터 엔지니어, 데이터 과학자, 머신러닝 엔지니어 등이 관련 직군에 포함된다.

머신러닝 시스템 설계 면접에서는 지원자가 시각 검색, 동영상 추천, 광고 클릭 예측 등과 같은 머신러닝 시스템을 끝에서 끝까지(end-to-end) 설계할 수 있는지 평가한다. 설계 면접 질문은 명확한 구조가 없기 때문에 까다로울 수 있다. 질문과 주제가 광범위해서 다양한 해석과 접근이 가능하기 때문에 정답이 정해져 있지 않은 경우가 많다.

머신러닝 시스템 설계 면접을 잘 보려면 머신러닝의 기본 개념과 기술을 깊이 이해하고 있어야 할 뿐만 아니라 이를 실제 문제 해결에 적용할 수 있어야 한다. 이를 위해 일반적으로 데이터 파이프라인, 피처 엔지니어링, 효과적인 머신러닝 시스템 설계에 대한 지식을 입증해야 한다. 또한 주어진 문제에 적합한 모델을 선택하고, 매개변수를 조정하고, 성능을 평가할 수 있는 능력을 추가로 보여 줘야 한다. 면접의 목표는 지원자가 머신러닝에 대한 이론을 적용하여 효과적인 시스템을 설계하고 구현하는 능력이 있는지 전반적으로 평가하는 것이다.

머신러닝 시스템 설계 면접이 왜 중요한가?

대부분의 면접 지원자는 머신러닝의 기본 개념은 이해하지만 공통된 가이드라인이 없어 머신러닝 시스템 설계 면접에서 어려움을 겪는 경우가 많다. 그러나 머신러닝 시스템을 설계하는 능력은 엔지니어에게 필수적인 기술이며, 특히 경력이 쌓일수록 더욱 중요해진다. 머신러닝 시스템에 잘못된 아키텍처를 선택하면 많은 시간과 리소스가 낭비될 수 있다.

머신러닝 시스템 설계 면접은 채용 과정에서 정말 중요한 부분이다. 면접에서 뛰어난 성과를 내면 더 나은 커리어 기회를 얻고 더 높은 연봉을 받을 수 있다.

이 책은 누가 읽어야 하나?

초보자든 숙련된 엔지니어든 머신러닝 시스템 설계에 관심이 있는 모든 사람에게 이 책은 필수 도서이다. 특히 머신러닝 면접을 준비해야 하는 사람들을 위해 집필했다.

이 책에서 다루지 않는 내용

이 책은 머신러닝의 기초를 다루지 않는다. 머신러닝 시스템 설계 면접에 대비하기 위해 추가 자료를 찾는 데이터 과학자, 데이터 엔지니어, 머신러닝 엔지니어를 위해 이 책을 썼다. 주로 기업의 머신러닝 엔지니어를 대상 독자로 하며, 학계나 산업 연구소의 머신러닝 과학자를 대상으로 하지는 않는다.

추가 자료

각 장(chapter) 말미에는 참고 자료가 많이 담겨 있다. 모든 링크는 다음 깃허브(GitHub) 저장소에 있다. *https://bit.ly/ml-bytebytego*

전자책 원서는 *bytebytego.com*에서 이용할 수 있다.

감사의 말

이 책의 모든 설계가 독창적인 내용이라고 할 수 있다면 좋겠다. 하지만 사실 이 책에서 언급된 대부분의 아이디어는 엔지니어링 블로그, 연구 논문, 코드,

유튜브 프레젠테이션 등 다른 곳에서도 찾아볼 수 있다. 우리는 이런 뛰어난 아이디어를 수집하고 검토한 후 개인적인 인사이트와 경험을 더해 알기 쉬운 방식으로 제시하려고 노력했다. 이 책을 완성하기 위해 십여 명의 엔지니어와 관리자가 검토하고 의견을 주었다. 모든 분께 정말 감사드린다!

- B 스리데비(비슈누 공과대학교)
- 다 청(Tiktok)
- 듀왕 술타니아(Adobe)
- 다이알라 에제딘(Tao Media)
- 디미트리스 코사코스(Elastic)
- 젠창 왕(Snapchat)
- 자잉 스(Amazon)
- 저스틴 리(Discord)
- 칼리안 디팍(Flipkart)
- 카우스투브 파드니스(Walmart)
- 리 쉬(TikTok)
- 라비 만들리야(Discord)
- 라비 람찬드란(Walmart Labs)
- 로힛 제인(Twitter)
- 사랑 메트카르(Meta)
- 샤바즈 파텔(One Concern)
- 숴 샹(Parafin)
- 수밤 쿠마르(Amazon)
- 토포조이 비스와스(Walmart)
- 비닛 알루왈리아(Stanford)
- 샤오 주(Databricks)
- 양원준(Twitter)
- 저후이 왕(Amazon)

마지막으로 엘비스 렌, 후아 리, 산 람의 소중한 기여에 특별한 감사를 표한다.

1장

Machine Learning System Design Interview

소개 및 개요

이 책의 목적은 머신러닝(machine learning, ML) 엔지니어와 데이터 과학자가 면접에서 머신러닝 시스템 설계를 잘 설명할 수 있게 돕는 것이다. 또한 이 책은 머신러닝이 실제 세계에 적용되는 방식에 대한 전반적인 정보도 제공한다.

많은 엔지니어가 로지스틱 회귀 또는 신경망과 같은 머신러닝 알고리즘을 머신러닝 시스템의 전부라고 생각한다. 그러나 운영 환경의 머신러닝 시스템에는 단순한 모델 개발보다 훨씬 더 많은 작업이 필요하다. 일반적으로 머신러닝 시스템은 데이터를 관리하기 위한 데이터 스택, 수백만 사용자가 사용할 수 있도록 하는 서비스 인프라, 시스템의 성능을 측정하기 위한 평가 파이프라인, 시간이 지나도 모델의 성능이 저하되지 않도록 하기 위한 모니터링 등 여러 구성요소로 복잡하게 구성된다.

머신러닝 시스템 설계 면접에서는 개방형 질문을 많이 한다. 예를 들어 영화 추천 시스템이나 동영상 검색 엔진을 설계하라는 요청을 한다. 정해진 정답은 없으며, 면접관은 지원자의 사고 과정, 다양한 머신러닝 주제에 대한 이해도, 시스템을 끝에서 끝까지(end-to-end) 설계하는 능력, 다양한 선택의 장단점을 반영한 설계 능력을 다각적으로 평가한다.

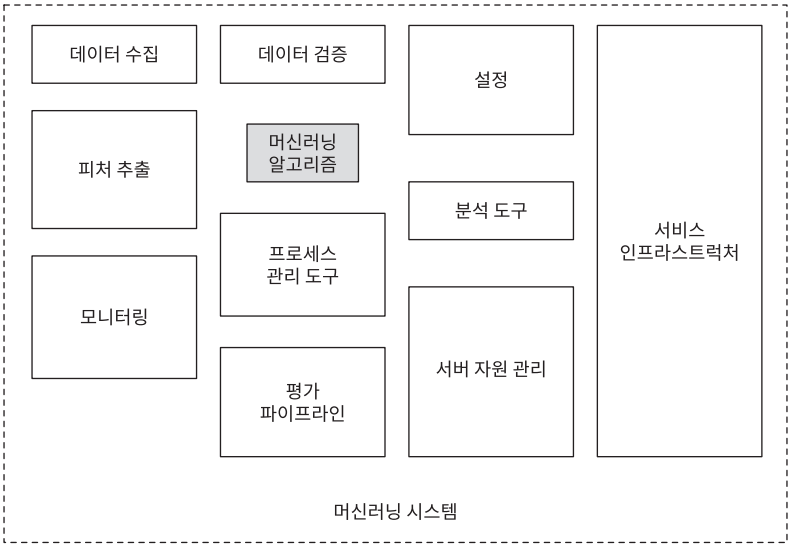


그림 1.1 운영 준비가 완료된 머신러닝 시스템

복잡한 머신러닝 시스템을 성공적으로 설계하려면 프레임워크를 따르는 것이 매우 중요하다. 틀을 갖추지 않은 답변으로는 논리적인 흐름을 만들기 어렵다. 1장에서는 머신러닝 시스템 설계 관련 질문에 효과적으로 대응할 수 있는 프레임워크를 제안한다. 프레임워크는 다음과 같은 단계로 구성된다.

1. 요구사항 명확화
2. 머신러닝 작업으로 문제를 구조화
3. 데이터 준비
4. 모델 개발
5. 평가
6. 배포 및 서비스 제공
7. 모니터링 및 인프라

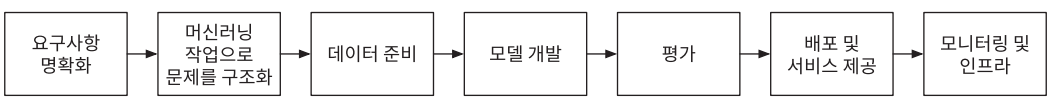


그림 1.2 머신러닝 시스템 설계 절차

개방형 질문이라는 말에서 알 수 있듯이 답변은 상황마다 다를 수 있다. 즉, 단 하나의 정답이 존재하지는 않는다. 이 프레임워크는 생각을 구조화하는 데 도움을 준다. 물론 융통성 있게 상황에 맞춰 적용하길 바란다. 면접관이 특히 모델 개발에 관심이 있다면 이 프레임워크가 효과가 있을 것이다.

프레임워크의 단계별 상세 내용은 다음과 같다.

요구사항 명확화

머신러닝 시스템 설계 관련 질문은 의도적으로 최소한의 정보만 제공하여 모호한 경우가 일반적이다. 예를 들어 면접 질문이 '이벤트 추천 시스템을 설계하시오'일 수 있다. 첫 번째 단계는 질문 내용을 정확하게 파악하는 것이다. 근데 어떻게 파악해야 할까? 요구사항을 명확하게 이해하려면, 다음과 같은 질문을 하는 게 면접을 시작하는 데 도움이 된다.

- **비즈니스 목표.** 휴가철 숙소 추천 시스템을 만들라는 요청을 받은 경우, 가능한 두 가지 비즈니스 목표는 예약 수와 수익을 늘리는 것이다.
- **시스템이 제공해야 하는 피처**[1]**.** 머신러닝 시스템 설계에 영향을 줄 수 있는 피처는 무엇인가? 예를 들어 동영상 추천 시스템을 설계하라는 요청을 받았다고 가정해 보자. 사용자가 추천 영상에 '좋아요' 또는 '싫어요'를 선택할 수 있는 기능이 있는지 확인하면 좋을 텐데, 이러한 상호작용은 훈련 데이터에 라벨을 지정하는 데 유용하게 사용할 수 있기 때문이다.
- **데이터.** 데이터 소스는 무엇인가? 데이터세트는 얼마나 큰가? 데이터에 라벨이 지정되어 있나?
- **제약.** 사용할 수 있는 컴퓨팅 자원은 얼마나 되나? 클라우드 기반 시스템인가, 아니면 단말에서 구동하는 시스템인가? 시간이 지남에 따라 모델이 자동으로 개선되길 바라나?

1 (옮긴이) 피처(feature)는 머신러닝과 패턴 인식 용어이다. 데이터의 특징이나 속성을 말한다.

- **시스템 규모.** 사용자 수는 어느 정도인가? 동영상과 같은 콘텐츠가 얼마나 많이 제공되나? 사용자 수, 콘텐츠 수 등의 증가율은 어느 정도인가?
- **성능.** 기대 성능은? 실시간 솔루션이 필요한가? 정확도와 대기 시간 중 무엇이 더 중요한가?

이 목록이 전부는 아니지만, 시작점이 될 수 있다. 개인정보 및 윤리와 같은 주제들도 마찬가지로 중요할 수 있다. 이 단계가 끝나면, 시스템의 구축 범위와 요구사항에 대해 면접관과 합의하게 된다. 합의된 요구사항 및 제약 조건을 잘 적어 두는 것이 좋다. 이렇게 하면 모두가 같은 내용에 동의했다고 생각할 수 있다.

머신러닝 작업으로 문제를 구조화

머신러닝 문제를 해결하기 위해서는 문제의 구조를 잘 정의하는 것이 매우 중요하다. 면접관이 동영상 스트리밍 플랫폼에 대한 사용자 참여도를 높여 달라는 요청을 한다고 해보자. 사용자 참여를 늘리는 것은 비즈니스 관점의 문제이고 아직 머신러닝이 작업할 수 있는 영역이 아니다. 따라서 이를 해결하기 위해 이 문제를 머신러닝 작업으로 구조화하자.

현실 작업에서는 주어진 문제를 해결하기 위해 머신러닝이 필요한지부터 먼저 판단해야 한다. 면접에서는 머신러닝이 도움이 된다고 당연하게 가정할 수 있으므로, 문제 해결을 위해 다음과 같이 머신러닝 작업의 틀을 잡는다.

- 머신러닝 목표 정의
- 시스템의 입력 및 출력 지정
- 적합한 머신러닝 유형 선택

머신러닝 목표 정의

비즈니스 목표는 매출을 20% 늘리거나 사용자 유지율을 높이는 것으로 잡았다. 그러나 목표를 단순히 '매출 20% 증가'라고 설정해서는 모델을 훈련할 수

없다. 머신러닝 시스템이 작업을 수행하려면 비즈니스 목표를 잘 정의된 머신러닝 목표로 변환해야 한다. 가장 좋은 방법은 머신러닝 모델이 해결할 수 있는 목표를 잡는 것이다. 표 1.1에 나와 있는 몇 가지 예를 살펴보자. 이후 장에서 더 많은 예시를 볼 것이다.

애플리케이션	비즈니스 목표	머신러닝 목표
이벤트 티켓 판매 앱	티켓 판매량 증가	이벤트 등록 수를 극대화
비디오 스트리밍 앱	사용자 참여 증가	사용자 시청 시간 극대화
광고 클릭 예측 시스템	사용자 클릭 수 증가	클릭률(click-through rate) 극대화
소셜 미디어의 유해 콘텐츠 감지	플랫폼 안정성 증대	해당 콘텐츠의 유해성 정확히 예측
친구 추천 시스템	사용자 네트워크 확장	형성된 연결 수 극대화

표 1.1 비즈니스 목표를 머신러닝 목표로 변환

시스템의 입력 및 출력 지정

머신러닝 목표를 결정한 후에는 시스템의 입력과 출력을 정의해야 한다. 예를 들어 소셜 미디어 플랫폼의 유해 콘텐츠 감지 시스템의 경우 입력은 게시물이고 출력은 이 게시물의 유해성 여부이다.

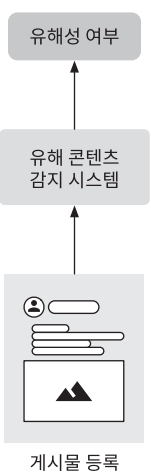


그림 1.3 유해성 감지 시스템의 입력-출력

때에 따라 시스템을 둘 이상의 머신러닝 모델로 구성할 수 있다. 이 경우 머신러닝 모델의 입력과 출력을 각각 지정해야 한다. 예를 들어, 유해한 콘텐츠 감지를 위해 폭력을 예측하는 하나의 모델과 과도한 노출을 예측하는 또 다른 모델을 적용할 수 있다. 시스템은 게시물이 유해한지 여부를 판단하기 위해 이 두 가지 모델을 사용한다.

각 모델의 입력-출력을 지정하는 다양한 방법이 있을 수 있다는 것도 또 다른 중요한 고려 사항이다. 그림 1.4를 참고하자.

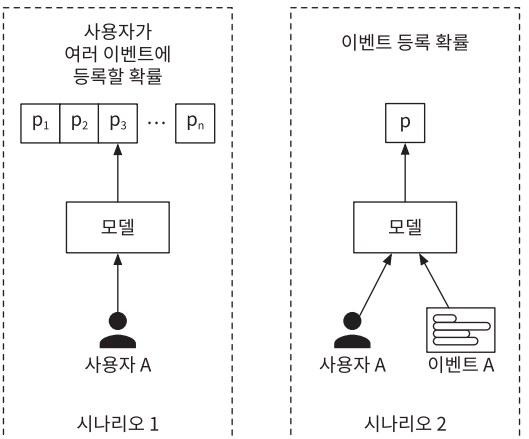


그림 1.4 모델의 입력-출력을 지정하는 다양한 방법

적합한 머신러닝 유형 선택

문제를 머신러닝 작업으로 구성하는 방법은 여러 가지이다. 대부분의 문제는 그림 1.5에 표시된 머신러닝 유형(리프 노드) 중 하나로 구성될 수 있다. 대부분의 독자가 이미 익숙할 테니 여기서는 개요만 설명한다.

- **지도 학습.** 지도 학습(supervised learning) 모델은 훈련 데이터세트를 사용하여 작업을 학습한다. 실제로 많은 문제가 이 유형에 속한다. 일반적으로 라벨이 지정된 데이터세트에서 학습하면 더 나은 결과를 얻을 수 있다.
- **비지도 학습.** 비지도 학습(unsupervised learning) 모델은 정답이 없는 데이터를 처리하여 예측한다. 비지도 학습 모델의 목표는 데이터 사이에서 의미

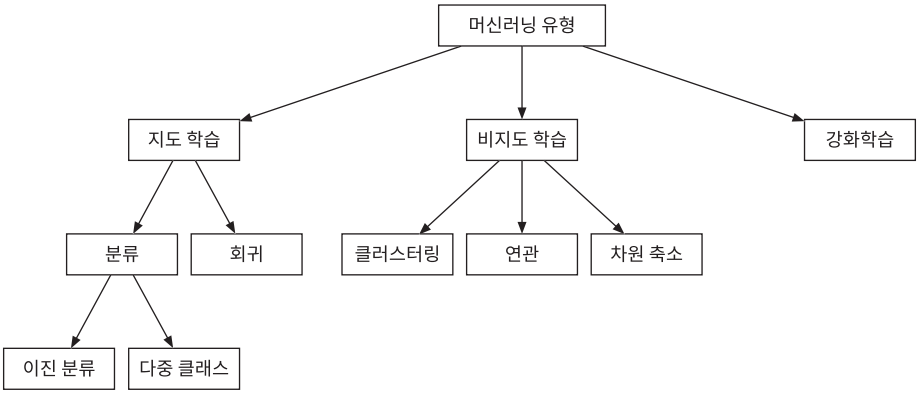


그림 1.5 일반적인 머신러닝 유형

있는 패턴을 식별하는 것이다. 일반적으로 사용되는 비지도 학습 알고리즘은 클러스터링, 연관 그리고 차원 축소이다.

- **강화학습.** 강화학습에서 에이전트(agent)[2]는 환경과의 상호작용에서 시행착오를 반복하여 수행하는 방법을 찾아낸다. 예를 들어 로봇이 방을 돌아다니고, 알파고가 바둑을 둘 수 있는 것이 이 강화학습을 통해 가능해진다.

지도 학습에 비해 비지도 학습과 강화학습은 많이 사용되지 않는데, 이는 머신러닝 모델이 일반적으로 학습 데이터를 사용할 때 특정 작업을 더 잘 학습하기 때문이다. 결과적으로 이 책에서 다루는 대부분의 문제는 지도 학습을 사용한다. 지도 학습의 다양한 유형에 대해 자세히 살펴보자.

- **회귀 모델.** 회귀는 연속적인 숫자 값을 예측하는 작업이다. 예를 들어, 집의 기대 가치를 예측하는 모델은 회귀 모델이다.
- **분류 모델.** 분류는 불연속 클래스 라벨을 예측하는 작업이다. 예를 들어 입력 이미지를 '개', '고양이' 또는 '토끼' 중 어느 쪽으로 분류할지 판단하는 것이다. 분류 모델은 두 그룹으로 나눌 수 있다.
  - **이진 분류 모델**은 이진 결과를 예측한다. 예를 들어 이미지에 개가 포함되어 있는지 예측한다.

2 (옮긴이) 강화학습의 대상이 되는 프로그램을 의미한다.

- 다중 클래스 분류 모델은 입력을 둘 이상의 클래스로 분류한다. 예를 들어 이미지를 개, 고양이 또는 토끼로 분류할 수 있다.

이 단계에서는 적합한 머신러닝 유형을 선택해야 한다. 이후 장에서는 면접 중에 올바른 유형을 선택하는 방법에 대한 예를 제공하겠다.

논의 주제

다음은 면접에서 언급하기 좋은 주제들이다.

- 좋은 머신러닝 목표는 무엇인가? 다른 머신러닝 목표와는 어떻게 비교하나? 장단점은 무엇인가?
- 주어진 머신러닝 목표에서 시스템의 입력 및 출력은 무엇인가?
- 머신러닝 시스템에 둘 이상의 모델이 적용된 경우 모델별 입력 및 출력은 무엇인가?
- 지도 또는 비지도 학습 중 어느 것을 사용해야 하나?
- 회귀 또는 분류 모델 중 어느 것을 사용하는 것이 좋을까? 분류의 경우 이진 분류 모델인가 다중 클래스 분류 모델인가? 회귀 모델이라면 출력 범위는?

데이터 준비

머신러닝 모델은 데이터를 통해 직접 학습한다. 그렇기 때문에 예측력 있는 데이터는 머신러닝 모델 훈련에 필수적이다. 이 절에서는 데이터 엔지니어링과 피처 엔지니어링이라는 두 가지 필수 프로세스를 사용하여 어떻게 고품질의 데이터를 모델에 입력할 것인지 살펴볼 예정이다. 각 프로세스의 중요한 측면을 다룰 것이다.

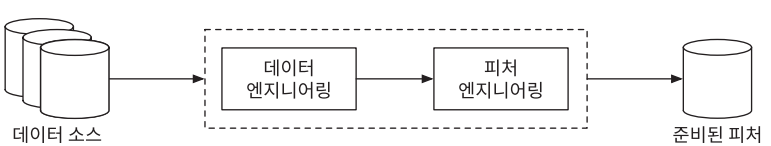


그림 1.6 데이터 준비 프로세스

데이터 엔지니어링

데이터 엔지니어링은 데이터의 수집, 저장, 검색 및 처리를 위한 파이프라인을 설계하고 구축하는 작업이다. 핵심 구성요소를 이해하기 위해 데이터 엔지니어링 기본 사항을 간략하게 살펴보자.

데이터 소스

머신러닝 시스템은 다양한 소스의 데이터를 사용한다. 다음과 같은 질문을 통해 데이터 소스를 잘 이해할 수 있다. 데이터 수집은 누가 하나? 데이터가 얼마나 깨끗한가? 데이터 소스를 신뢰할 수 있나? 사용자가 생성한 데이터인가 아니면 시스템이 생성한 것인가?

데이터 저장소

데이터를 상시로 저장하고 관리하기 위한 데이터베이스이다. 사용 사례별로 서로 다른 데이터베이스가 구축되므로 각 데이터베이스가 작동하는 방식을 높은 수준에서 이해하는 것이 중요하다. 일반적으로 머신러닝 시스템 설계 면접 중에 데이터베이스 내부에 대한 상세 질문은 잘 나오지 않는다.

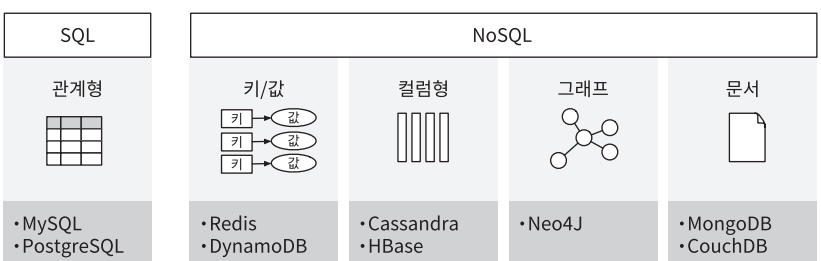


그림 1.7 다양한 종류의 데이터베이스

ETL(추출, 변환 및 적재)

ETL(extract, transform, load)은 세 단계로 구성된다.

- **추출**. 다양한 데이터 소스에서 데이터를 추출한다.
- **변환**. 이 단계에서 요구사항에 맞게 데이터 정제, 매핑 및 특정 형식으로 변환한다.

- **적재**. 변환된 데이터를 파일, 데이터베이스 또는 데이터 웨어하우스[1]에 적재한다.

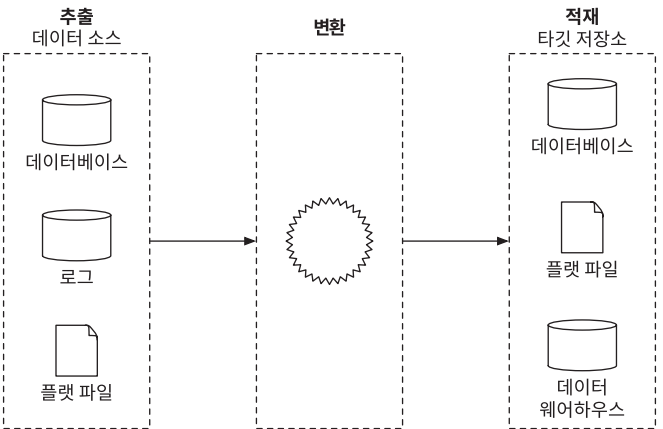


그림 1.8 ETL 프로세스 개요

데이터 유형

머신러닝의 데이터 유형은 프로그래밍 언어의 데이터 유형(int, float, string 등)과는 다르다. 크게 정형 데이터와 비정형 데이터 두 가지로 나뉜다(그림 1.9 참고).

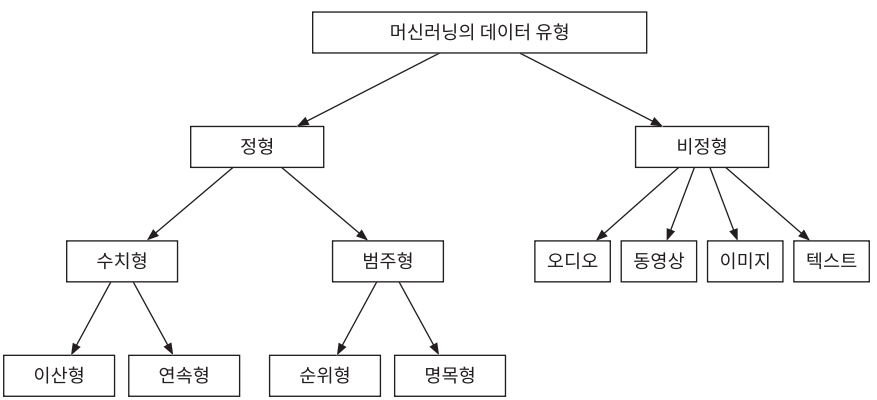


그림 1.9 머신러닝의 데이터 유형