



머신 러닝 소개

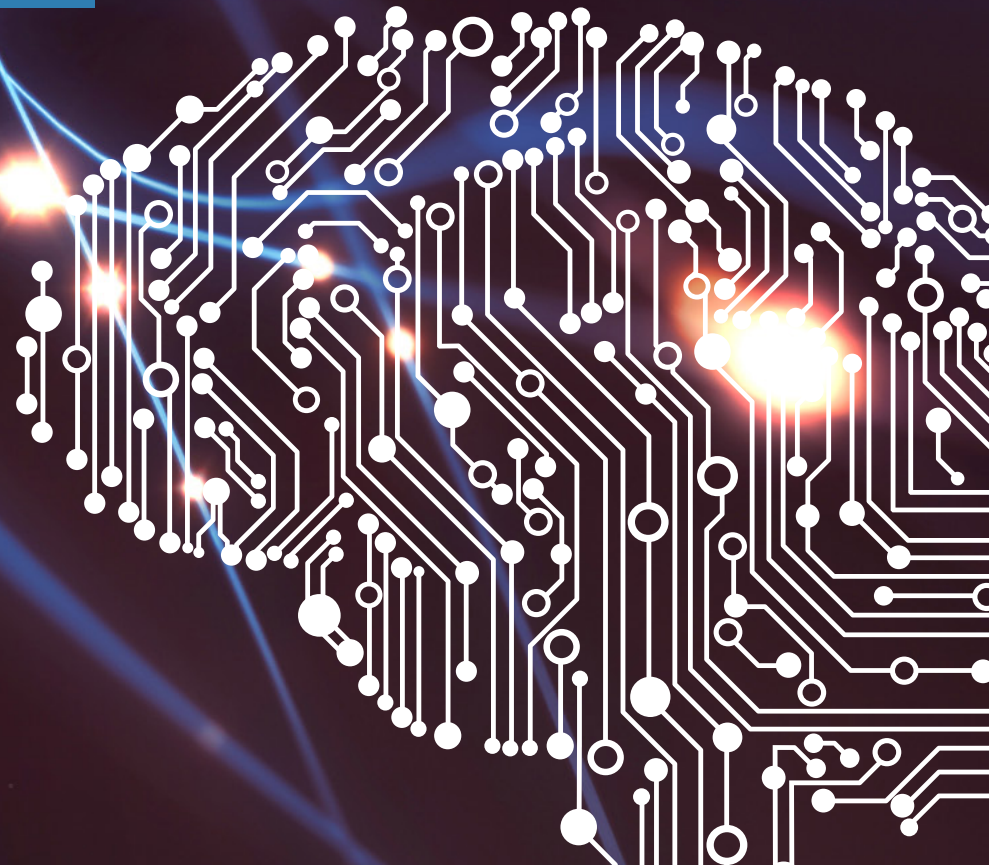
```
%% Generalized Linear Model - Logistic Regression
glm = GeneralizedLinearModel.fit(Xtrain,double(Ytrain),...
    'linear','Distribution','binomial','link','logit');

%% Discriminant Analysis
da = ClassificationDiscriminant.fit(Xtrain,Ytrain,...
    'discrimType','quadratic');
```

```
%% Classification Using Nearest Neighbors
knn = ClassificationKNN.fit(Xtrain,Ytrain,...
    'Distance','seuclidean');

%% Ensemble Learning: TreeBagger
opts = statset('UseParallel',true);

tb = TreeBagger(150,Xtrain,Ytrain,'method','classification',...
    'Options',opts,'OOBVarImp','on','cost',[0 1; 5 0]);
```



머신 러닝이란?

머신 러닝은 사람과 동물에게 자연스러운 일, 즉 경험을 통해 학습하는 것을 컴퓨터가 수행할 수 있도록 가르칩니다. 머신 러닝 알고리즘은 미리 결정된 방정식을 모델로 의존하지 않고 계산 방법을 사용하여 데이터에서 직접 정보를 “학습”합니다. 학습할 수 있는 샘플 수가 증가함에 따라 알고리즘 성능이 적절하게 향상됩니다.

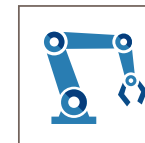
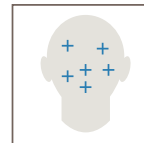
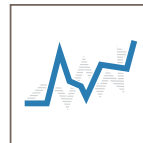
더 많은 데이터, 더 많은 질문, 더 나은 답변

머신 러닝 알고리즘은 통찰력을 생성하고 더 나은 의사결정과 예측을 수행하도록 도와주는 자연 패턴을 데이터에서 찾습니다. 자연 패턴은 의료 진단, 주식 거래, 에너지 부하 예측 등에서 중요한 결정을 내리기 위해 매일 사용됩니다. 미디어 사이트는 머신 러닝에 의존하여 수백만 개의 옵션을 살펴보고 노래 또는 동영상 추천을 제공합니다. 소매업체는 머신 러닝을 통해 고객의 구매 행동에 대한 통찰력을 얻습니다.

실제 응용 분야

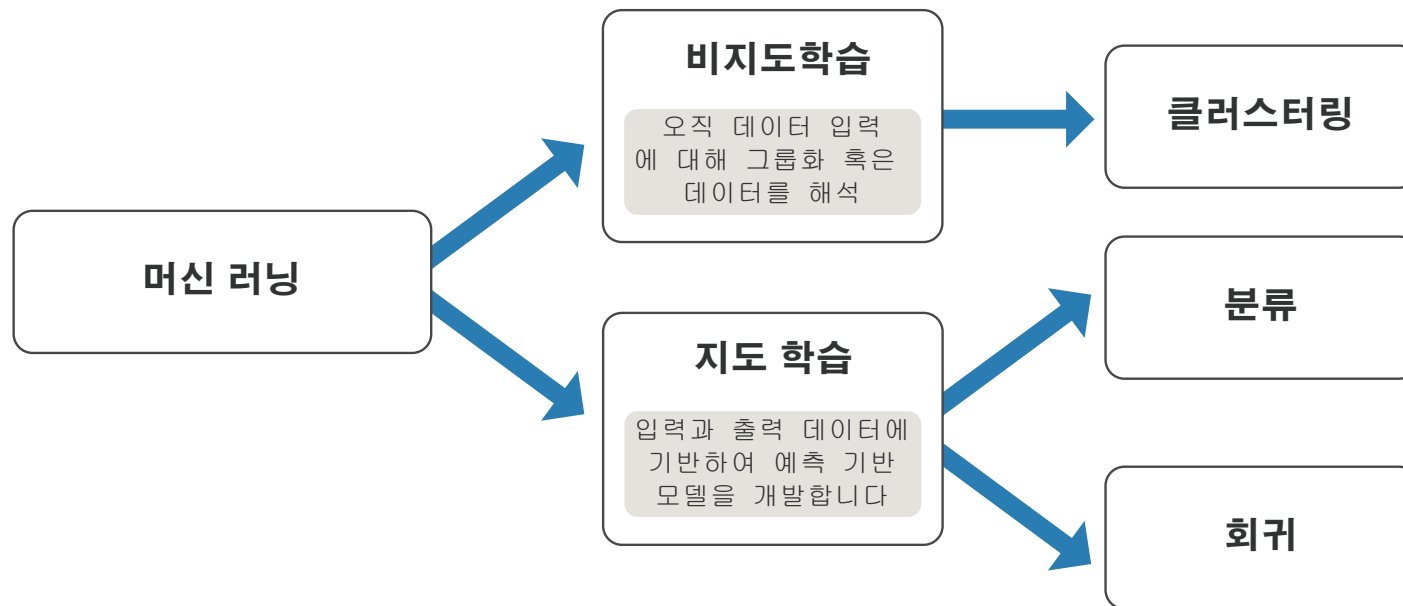
빅데이터가 부상하면서 다음과 같은 영역에서 문제를 해결하는데 머신 러닝이 특히 중요해졌습니다.

- 계산 금융 - 신용 평가 및 알고리즘 트레이딩
- 이미지 프로세싱 및 컴퓨터 비전 - 얼굴 인식, 얼굴 인식, 동작 인식, 객체 인식
- 계산 생명공학 - 종양 감지, 신약 발견, DNA 염기 서열 (DNA Sequence)
- 에너지 생산 - 가격 및 부하 예측
- 자동차, 항공 및 제조 - 예측 기반 유지 보수 시스템 개발
- 자연어 처리



머신 러닝의 작동 방식

머신 러닝은 두 가지 유형의 기법을 사용합니다. 즉, 지도 학습 기법은 미래 출력을 예측할 수 있도록 알려진 입력 및 출력 데이터를 기반으로 모델을 훈련하고, 비지도 학습 기법은 입력 데이터에서 숨겨진 패턴이나 고유 구조체를 찾습니다.



지도 학습

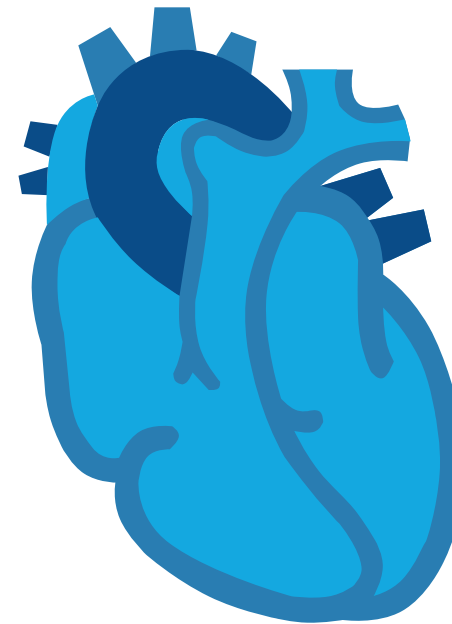
지도 학습의 목적은 불확실성이 있을 때 증거를 기반으로 예측을 하는 모델을 작성하는 것입니다. 지도 학습 알고리즘은 알려진 입력 데이터셋 및 해당 데이터에 대한 알려진 출력을 사용하고 새 데이터에 대한 응답을 위해 합리적인 예측을 생성하도록 모델을 훈련합니다.

지도 학습 분류 및 회귀 기법을 사용하여 예측 모델을 개발합니다.

- **분류 기법**은 이메일이 진짜 또는 스팸인지 여부, 종양이 악성 또는 양성인지 여부 등의 개별 응답을 예측합니다. 분류 모델은 입력 데이터를 범주로 분류합니다. 일반적인 응용 분야에는 의료 이미징, 음성 인식, 신용 평가 등이 있습니다.
- **회귀 기법**은 온도 변화 또는 전력 수요 변동 등의 연속 응답을 예측합니다. 일반적인 응용 분야에는 전기 부하 예측, 알고리즘 트레이딩 등이 있습니다.

지도 학습을 사용하여 심장마비 예측

임상 의사가 누군가 1년 내에 심장마비를 일으킬지 여부를 예측하고자 한다고 가정해 보십시오. 임상 의사는 연령, 체중, 키, 혈압을 비롯하여 이전 환자에게 대한 데이터를 보유하고 있으며, 이전 환자들에게 1년 내에 심장마비가 나타났는지 여부를 알고 있습니다. 따라서 문제는 기존 데이터를 새로운 사람이 1년 내에 심장마비를 일으킬지 예측할 수 있는 모델에 결합하는 것입니다.

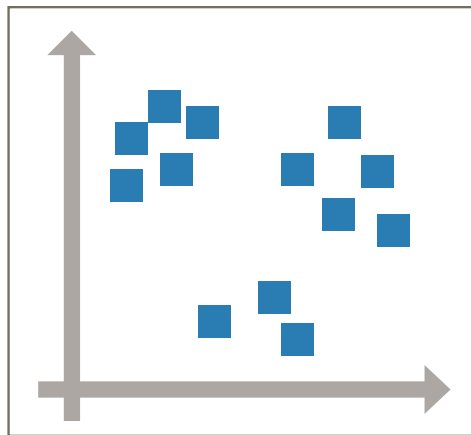


비지도 학습

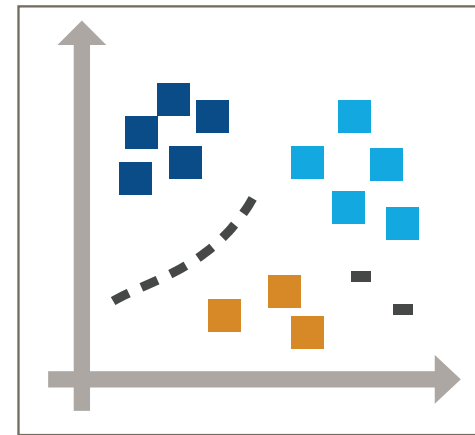
비지도학습은 데이터에서 숨겨진 패턴이나 고유 구조체를 찾습니다. 그러한 패턴이나 구조체는 분류된 응답 없이 입력 데이터로 구성된 데이터셋에서 추론하는 데 사용됩니다.

클러스터링은 가장 일반적인 비지도학습 기법입니다. 이 기법은 탐색적 데이터 분석을 통해 데이터에서 숨겨진 패턴이나 그룹을 찾는 데 사용됩니다.

클러스터링 응용 분야에는 유전자 서열 분석, 시장 조사, 객체 인식 등이 있습니다.



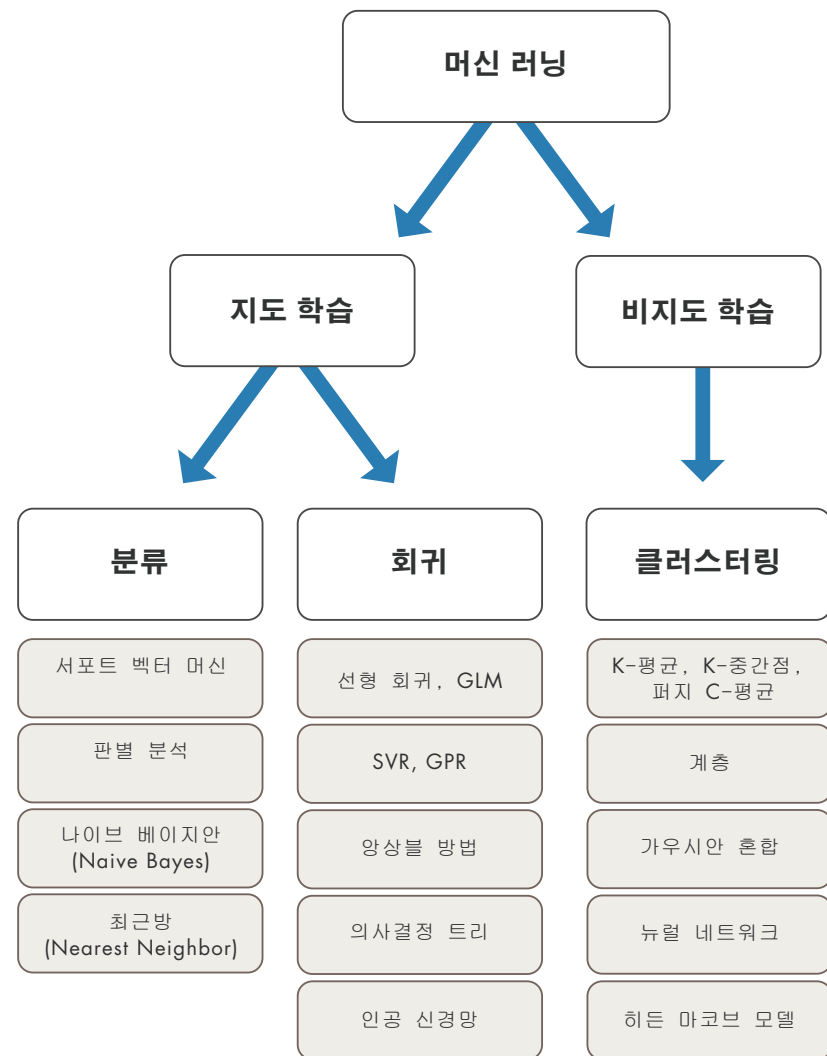
데이터의 패턴
클러스터링



어떤 알고리즘을 사용할지 어떻게 결정합니까?

수십 개의 지도 학습 및 비지도 학습 알고리즘 알고리즘이 있고 각 알고리즘에는 다양한 학습 접근법이 사용되기 때문에 적합한 알고리즘을 선택하는 일은 매우 어려운 일처럼 보일 수 있습니다.

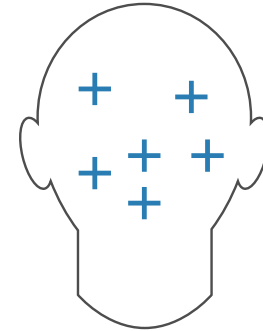
최상의 방법이나 모든 상황에 맞는 알고리즘은 없습니다. 적합한 알고리즘을 찾는 것은 어느 정도는 시행착오 과정이라 할 수 있습니다. 경험이 많은 데이터 과학자조차도 시도해보지 않으면 알고리즘이 적합한지 여부를 알 수가 없습니다. 하지만 알고리즘 선택은 작업 중인 데이터의 크기와 유형, 데이터에서 얻으려는 통찰력, 이 통찰력을 사용하는 방식에 따라서도 달라집니다.



머신 러닝을 언제 사용해야 할까요?

대용량 데이터와 많은 변수가 관련되어 있지만 기존 공식이나 방정식이 없는 복잡한 작업이나 문제가 있을 경우 머신 러닝을 사용해 보십시오. 예를 들어 다음과 같은 상황을 처리해야 할 경우 머신 러닝이 좋은 옵션입니다.

- 얼굴 인식, 음성 인식의 경우처럼 직접 작성하는 규칙과 방정식이 너무 복잡한 상황
- 거래 기록에서 사기를 감지하는 경우처럼 작업 규칙이 지속적으로 바뀌는 상황
- 자동화된 트레이딩, 에너지 수요 예측, 쇼핑 추세 예측의 경우처럼 데이터 특징이 계속 바뀌고 프로그램을 조정해야 하는 상황



고객 활용 사례

예술 작품을 분석할 수 있는 알고리즘 작성

러트거스 대학교의 예술 및 인공 지능 실험실 연구원들은 컴퓨터 알고리즘이 사람처럼 쉽게 그림을 스타일, 장르, 예술가별로 분류할 수 있는지 확인하고자 했습니다. 연구원들은 먼저 그림 스타일을 분류하기 위해 시각적 특징을 식별했습니다. 개발된 알고리즘은 비전문가인 일반인을 능가하여 60%의 정확도로 데이터베이스에서 그림 스타일을 분류했습니다.

연구원들은 스타일 분류에 유용한 시각적 특징(지도학습)이 예술적 영향(비지도학습)을 확인하는 데 사용될 수도 있다는 가설을 세웠습니다.

그들은 Google 이미지를 기반으로 훈련된 분류 알고리즘을 사용하여 특정 객체를 식별했으며, 550년의 기간에 걸쳐 서로 다른 예술가 66명이 그린 그림 1,700점 이상을 대상으로 알고리즘을 테스트했습니다. 이 알고리즘은 디에고 벨라스케스의 “교황 이노센트 10세의 초상”이 프랜시스 베이컨의 “교황 이노켄티우스 10세의 초상화 연구”에 미치는 영향을 비롯하여 관련 작품을 쉽게 식별했습니다.



고객 활용 사례

대형 건물의 HVAC 에너지 사용 최적화

사무실 건물, 병원, 기타 대형 상업 건물의 HVAC(난방, 환기 및 냉방) 시스템은 변화하는 날씨 패턴, 가변적인 에너지 비용 또는 건물의 열 특성을 고려하지 않기 때문에 비효율적인 경우가 많습니다.

Building IQ의 클라우드 기반 소프트웨어 플랫폼은 이러한 문제를 해결합니다. 이 플랫폼은 고급 알고리즘과 머신 러닝 방법을 사용하여 전력계, 온도계, HVAC 압력 센서에서 수집된 수 기가바이트의 정보와 날씨 및 에너지 비용을 지속적으로 처리합니다. 특히, 머신 러닝을 사용하여 데이터를 세분화하고 난방 및 냉방 프로세스에 대한 가스, 전기, 증기 및 태양열의 상대적 기여도를 결정합니다. Building IQ 플랫폼은 정상 운영 시 대형 상업 건물의 HVAC 에너지 소비량을 10%~25% 절감합니다.

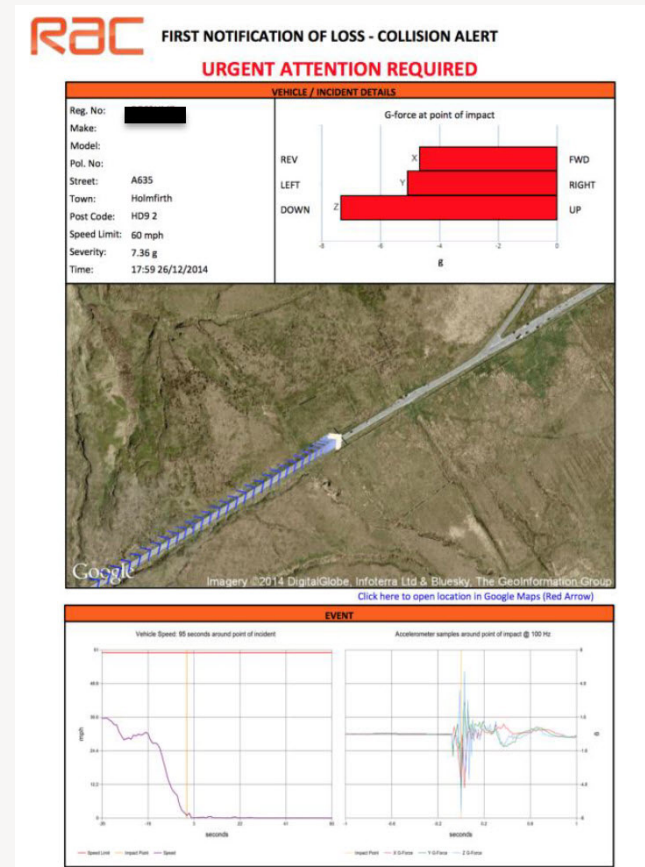


고객 활용 사례

저속 차량 충돌 감지

8백 만 이상의 회원을 보유한 RAC는 영국의 가장 큰 자동차 단체 중 하나로서, 개인/비즈니스 운전자에게 긴급출동 서비스, 보험 및 기타 서비스를 제공합니다.

도로 사고에 신속히 대응하고, 추돌 사고를 줄이고, 보험비를 낮추기 위해 RAC는 고급 머신 러닝 알고리즘을 사용하여 저속 충돌을 감지하고 과속 방지턱이나 포트홀 운행과 같이 더 일반적인 운전 이벤트와 이러한 이벤트를 구분하는 온보드 충돌 감지 시스템을 개발했습니다. 독립적인 테스트에서 RAC 시스템은 테스트 충돌 감지 시 92% 정확도를 나타냈습니다.



추가 정보

자세히 살펴볼 준비가 되셨습니까? 다음 리소스를 통해 머신 러닝 방법, 예제, 도구를 자세히 알아보십시오.

▶ 시청 자료

[머신 러닝으로 간편하게](#) 34:34

[센서 데이터 분석을 위한 신호 처리 및 머신 러닝 기법](#) 42:45

📄 읽기 자료

[머신 러닝 블로그 게시물: 소셜 네트워크 분석, 텍스트 마이닝, 베이지안 추론 등](#)

[Netflix Prize 및 프로덕션 머신 러닝 시스템: 내부자 관점](#)

[머신 러닝 과제: 최상의 모델 선택과 과적합 방지](#)

🔍 살펴보기

[MATLAB 머신 러닝 예제](#)

[머신 러닝 솔루션](#)

[Classification Learner 앱을 사용한 데이터 분류](#)