하고 싶은 건 많으나 ‘어떻게’를 모름

아는 만큼 질문, 질문도 할 수 없고.

1. 데이터 사이언스를 시작하기 전 고려해야 할 사항
2. 비지도 학습법
3. 지도 학습법
4. 강화 학습()

비지도 학습, 지도 학습

지도 학습법, 강화학습

무슨 근거로 병원진단을 내리는가?

전문가와 기계의 중간 다리 역할

사이언티스트가 해야할 일: 기계에게 어떻게 알려줄 것인가?

그걸 분석하는 것이 데이터 사이언스가 된다.

데이터 사이언스와 데이터 아날리틱스

IoT와 IoE 수준의 다른 정도

기존에 있던 기술을 빅데이터에 집어넣으면 무언가 나온다

그게 데이터 사이언티스트가 할 일

하지만 데이터 아날리틱스 돈을 잃으면, 손해를 보면 안된다.

손해가 났따면, 그 손해를 메꾸기 위한 공부도 해야 하는 것이 아날리틱스

예: 환자 70%만 선별 가능

30% 감별은 애널리틱스

30%오류는 왜 발생했는가, 줄이기 위해서는 뭘 해야하는가?

사이언티스트든 아날리틱스든 빅데이터가 필요하다.

충분한 데이터. 다운로드, 설문조사든 충분한 데이터를 필요로 한다.

양이 적다면 엑셀도 충분하다

그래프 그리기, 게임도 가능

강화 학습

받아쓰기와 비슷하다

왜 틀렸는지까지, 고칠 수 있게 고려하는

3) 모델을 최적화하기 위한 알고리즘

4) 정확도 평가

실제값과 예측값의 차이가 적다면 평균도 작아지고 정확도도 높아진다.

2--------------

패턴을 찾아야 예측이 가능하기 때문에

비지도 학습이 중요하다

뭉쳐 있는지, 뭉쳐있다면 얼마나 빽빽하게, 어떤 모양을 그리고 있는지 그게 패턴을

찾는 절차

“군집화, 밀도 제각각 공간도 제각각이라서 패턴을 찾을 수 없습니다.”

K-평균 군집화

K는 군집의 수(원의 수를 생각하면 됨)

보통 2개 (보통 차원의 개수를 따라간다)

Means는 면적을 뜻한다

각 원에 들어간 영화의 수가 비슷해야 한다.

고르게 분포. 평균적으로. 면적도 비슷해야 되고, 갯수도 비슷해야 한다.

보통 원형을 많이 사용. 왜? 원을 그리려면 반지름과 중심 정보를 알아야 한다. 원이 가장 단순하다.

하지만 사각형의 경우 시작좌표 끝좌표가 필요하고, 오각형, 육각형 등등.

주성분 분석(PCA)

하나 감소할 때 하나 증가, 하나 증가할 때 하나 감소

연관 규칙

파이썬이 좋은 언어이지만 단점이, 너무 많이 한다.

답, 라이브러리, 방법론이 너무 다양하다.

R은 제단에서 중재적으로 관리를 하기 때문에 겹치는 게 적다.

성능 별로. 결과 나오는 데까지 시간이 걸린다.

개성이 적어서 모양이 거기서 거기다.

사과를 팔았는데 맥주도 팔 수 있었던 것

맥주가 잘 팔려서 사과가 판매되었다는 아니다.

유기적인 관계?

사과 팔린 영수증 X 맥주 영수증

사과와 맥주만 가지고, 함께 팔린 수

많이 팔린 건 크기로, 무조건 같이 사는 관계는 붉은색.

비지도 학습

왜 한국은 미국과 유럽 무기를 많이 사용할까?

인과관계는 모른다

**잊어먹지 마세요 리스트**