

표준화

인공지능에서 반드시 쓰이는 매우중요한 전처리 기법이다.

표준화를 깜빡해서 데이터가 이상하고 실수가있다면 표준화를 안해서 문제가 생기는 경우이다.

인간은 원칙적으로 알아서 표준화를 거치는데 컴퓨터는 안그렇다. 영화평점이 8/10 이면 상당히 좋고 수능 8/400 이면 매우않좋다 또다른 예시로 150/400 이중에서 영화 8점하고 150중에 어느게 낮다고 판단을해야한다. 이런식으로 어느게 나은지 판단해야한다.이런식으로 데이터를 놓게되면 수능 150 맞은사람이 8점 맞은사람보다 높게 나온다. 무조건 큰수를 기준으로 생각하기 때문이다. 표준화를 통해 어느정도 잘한것인지 어느정도 큰것인지 못건 큰수를 생각하기 때문에 표준화를 통해 어느정도 큰것인지 바꾸어야 한다. 그래서 모든 데이터의 평균을 0으로 표준편차를 1로 변환해야한다. 수식 $z = (a - m) / \sigma$ -1에서 1까지의 수치는 68%이다.

표준편차가 2이면 매우크다. 문제 [1,2,3] $\sigma = \sqrt{1/3 \cdot 2}$

$[-1/\sqrt{6/3}, 0, 1/\sqrt{6/3}]$ 이렇게 된다. 주식데이터에서보면 가격데이터가있고 거래량이있다.

가격데이터는 200,300 거래량은 30000 주가 있으면 표준화를 안하면 거래량만안하면 매우좋다고 보기에 표준화를해서 가격데이터의 비율로 살펴야한다.