**머신러닝(ML) 공부**

-귀무가설(H0) a null hypothesis : 기존에 믿어온 사실

-대립가설(H1) alternative hypothesis : 실험자가 새로이 주장하려는, 또는 입증하려는 사실



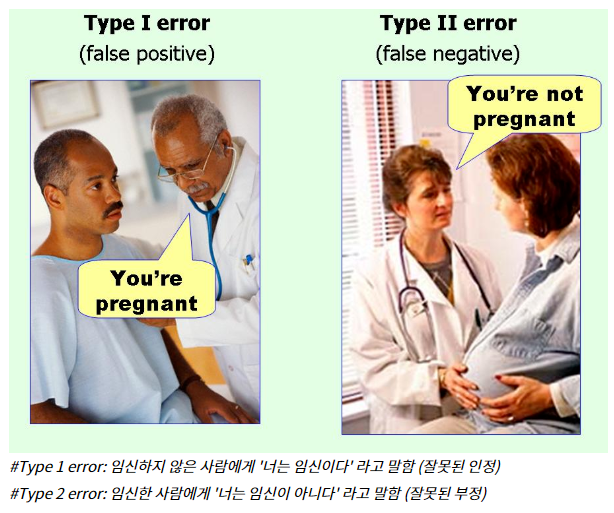
-제 1종 오류(alpha 오류) : 귀무가설이 참인데도 불구하고 귀무가설을 기각 ( 귀무가설이 참인데도 대립가설 채택)

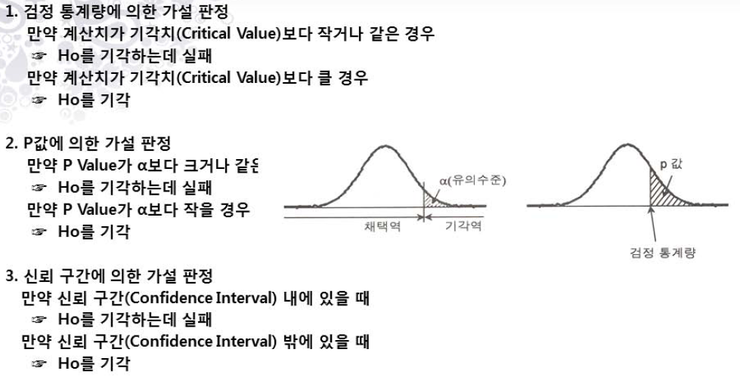
-제 2종 오류(beta 오류) : 귀무가설이 거짓임에도 불구하고 귀무가설을 채택 ( 귀무가설이 거짓인데도 대립가설 기각)

우리가 통계적인 측면에서 가설검정하려는 문제들은 우선적으로 **"귀무가설 위주로 생각"**하자는 것이다. 기존의 귀무가설을 증명하려고, 더 확고히 하려고자 검정한다는것은 무의미할 뿐이다. 따라서 우리가 검정하려고 달려드는 이유는 새로운 대립가설이 옳다 혹은 더 낫다 고 주장하기 위함이다.

경찰이 범하는 제 1종오류는 수상한 자가 범인이 아닌데도 수사 후에 범인으로 구속하는 오류이다.

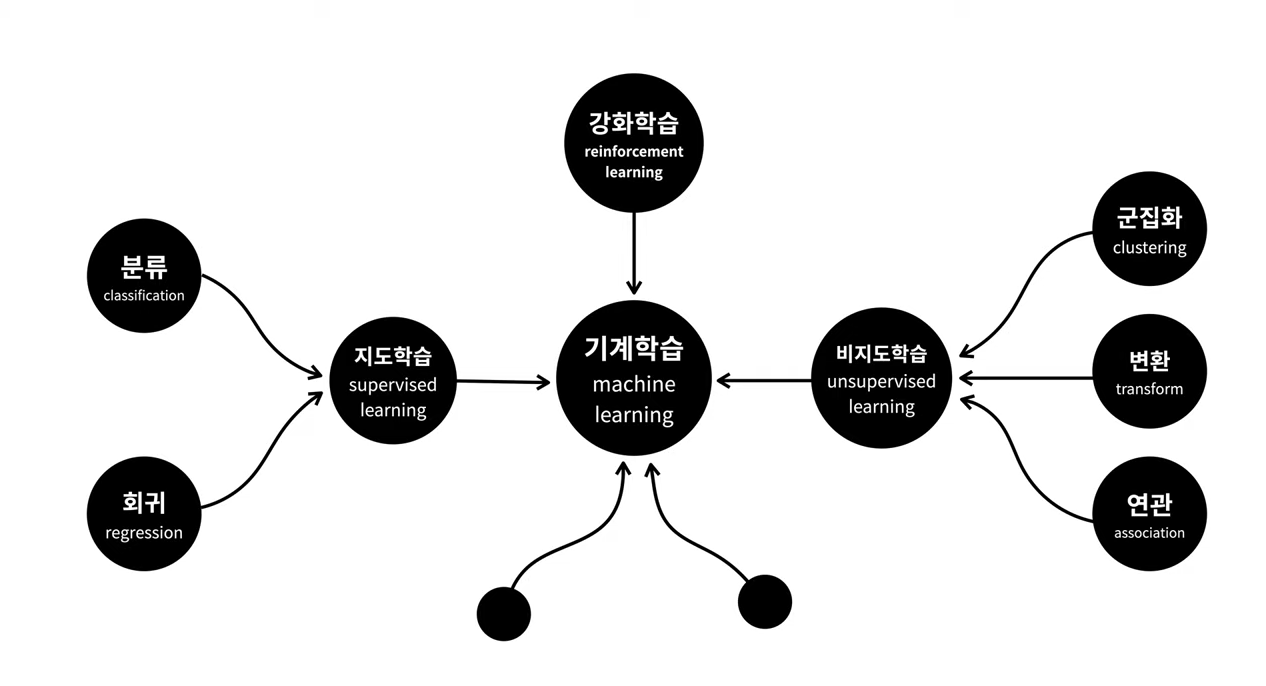
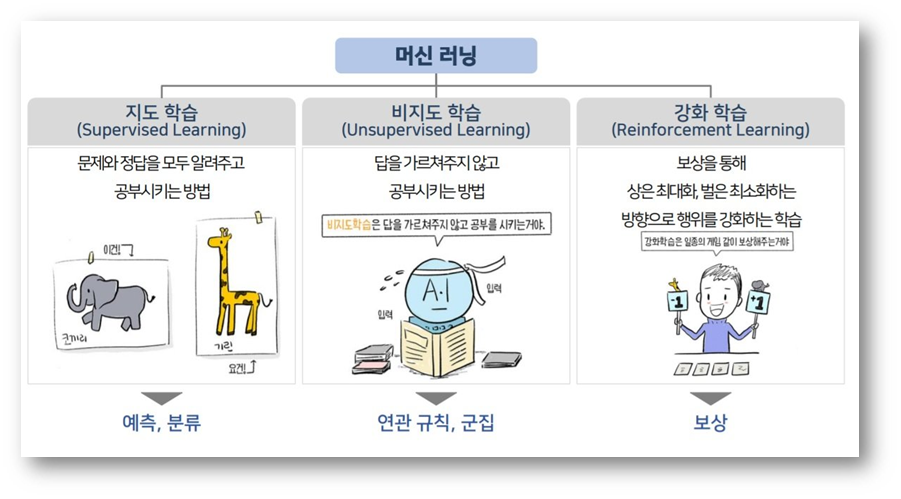
경찰이 범하는 제 2동오류는 수상한 자가 범인임에도 불구하고 증거불충분 등으로 수상한 자를 훈방시키는 오류이다.



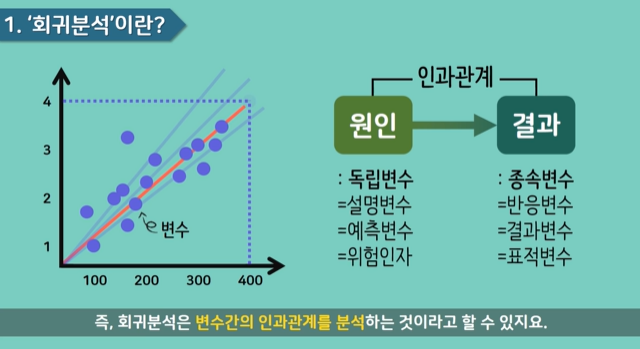


유의수준(level of significance)

신뢰구간(confidence interval)



회귀분석(regression analysis)은 매개변수 모델(parametric model)을 이용하여 통계적으로 변수들 사이의 관계를 추정하는 분석방법이다. 주로 독립변수(independent variable)가 종속변수(dependent variable)에 미치는 영향을 확인하고자 사용하는 분석방법이다. 회귀분석은 다른 독립변수들을 고정시키고 한 가지 독립변수만을 변화시킬 때 종속변수가 어떻게 변화하는지를 확인한다. 종속변수와 관련이 있는 독립변수를 찾을 때, 또 독립변수들 간의 관계를 이해하고자 할 때 사용한다. 하나의 종속변수와 하나의 독립변수 사이의 관계를 분석할 때 단순회귀분석(simple regression analysis)이라 하고, 하나의 종속변수와 여러 독립변수 사이의 관계를 규명하고자 할 때 다중회귀분석(multiple regression analysis)이라 한다.



**비지도 학습**

# PCA (Principal component analysis)  
주성분 분석(PCA)은 특성들이 통계적으로 상관관계가 없도록 데이터셋을 회전시키는 기술입니다. 먼저 성분1 이라고 쓰여 있는 분산이 가장 큰 방향을 찾습니다. 이 방향(또는 벡터)이 데이터에서 가장 많은 정보를 담고 있는 방향입니다.

다음으로 첫 번째 방향과 직각인 방향 중에서 가장 많은 정보를 담은 방향을 찾습니다. 2차원에서는 가능한 직각 방향이 하나뿐이지만 고차원에서는 많은 직각 방향이 있을수 있습니다.

# K-means  
분류 알고리즘은 데이터셋을 클러스터 라는 그룹으로 나누는 작업입니다. 한 클러스터 안의 데이터 포인트끼리는 매우 비슷하고, 다른 클러스터의 데이터 포인트와는 구분되도록 데이터를 나누는 것이 목표입니다.

K-평균 알고리즘은 먼저 데이터의 어떤 영역을 대표하는 클러스터 중심을 찾습니다. 그리고 나서 알고리즘은 다음 두 단계를 반복합니다.  
1. 데이터 포인트를 가장 가까운 클러스터 중심에 할당  
2. 클러스터에 할당된 데이터 포인트의 평균으로 클러스터 중심을 지정  
클러스터에 할당되는 데이터 포인트에 변화가 없을때 알고리즘이 종료됩니다.