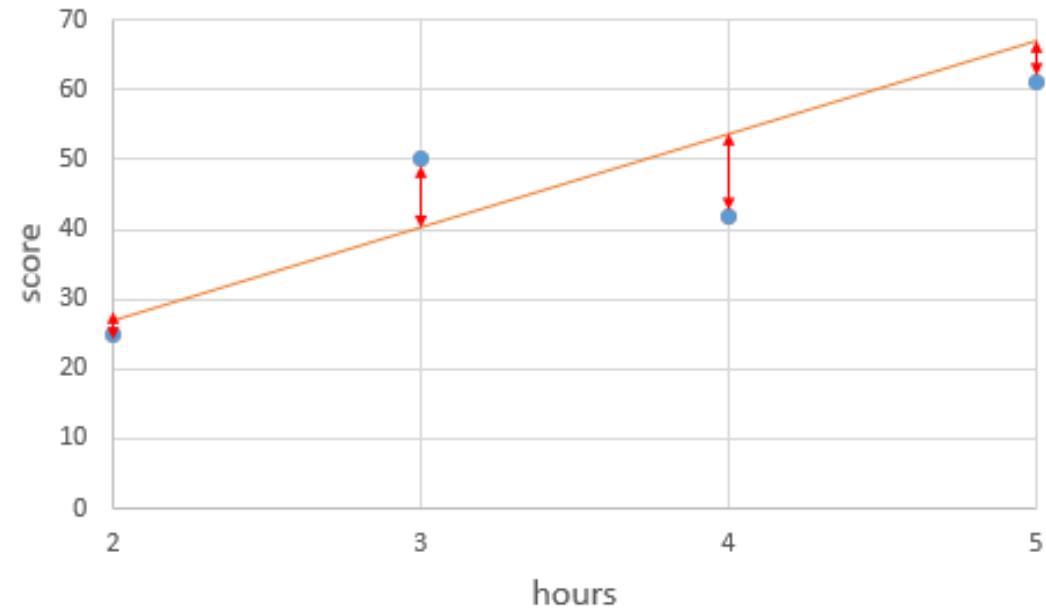


머신러닝 개념입문

선형회귀(다중선형회귀)

- 통계학에서 종속 변수 y 와 한 개 이상의 독립 변수 X 와의 선형 상관 관계를 모델링하는 회귀분석기법
- 한 개의 설명 변수에 기반한 경우 단순 선형 회귀
- 둘 이상의 설명 변수에 기반한 경우가 다중 선형 회귀

- 통계학에서 종속 변수 y 와 한 개 이상의 독립 변수 X 와의 선형 상관 관계를 모델링하는 회귀분석기법
- 다른 언어의 값을 변하게 하는 X 는 독립변수
- X 에 의해 값이 변하는 Y 는 종속변수
- 데이터값이 주어졌을 때 가장 만족시키는 식을 구한다.



단순 선형 회귀

$$Y=Wx+b$$

고정된 W , b 와 x 값에 대해 값이 변하는 y

다중 선형 회귀

$$y=W_1x_1+W_2x_2+...W_nx_n+b$$

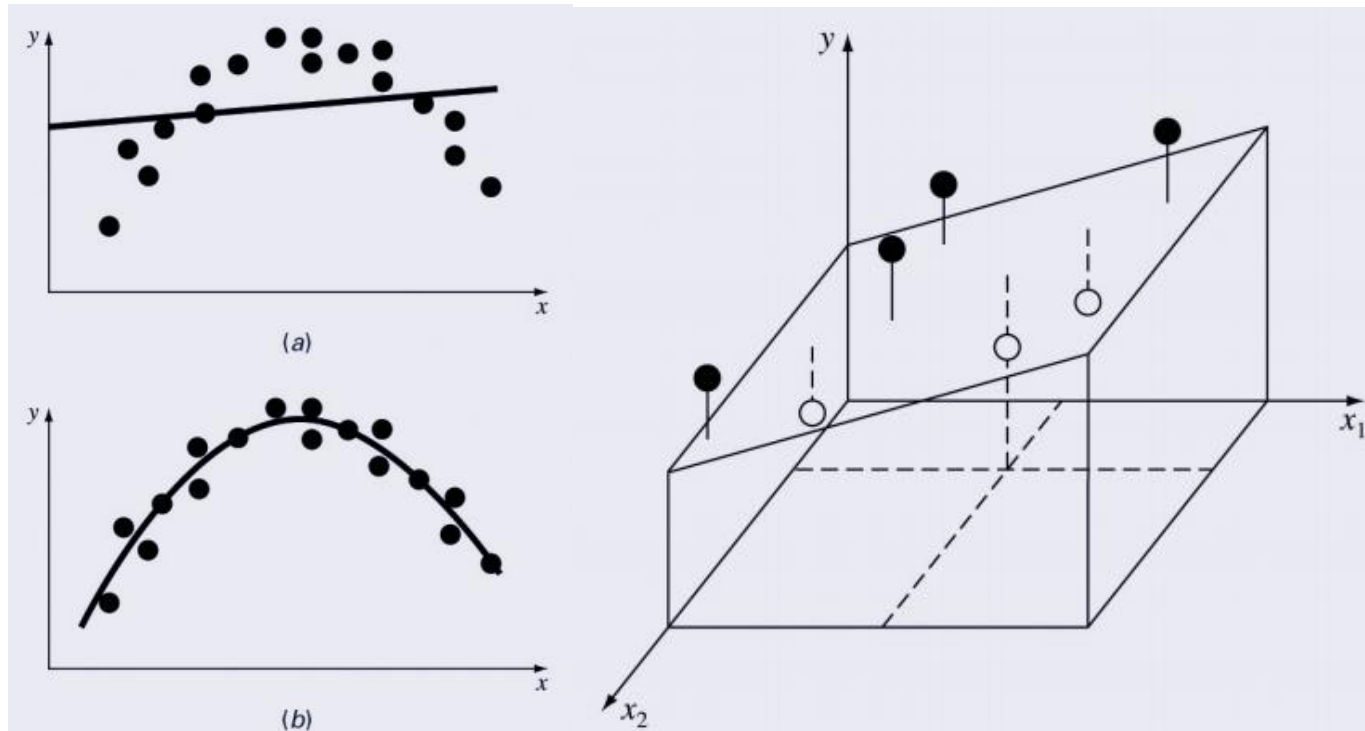
여러 x 값에 대해 값이 변하는 y

4.2 구성

수치해석 책 14, 15장 ->

- 14장 : 최소제곱 회귀분석
- 15장 : 일반적인 선형최소제곱과 비선형회귀분석

다항식인경우와
다중선형인경우 ->



로지스틱회귀

- 독립변수의 선형 결합을 이용 사건 발생 가능성을 예측하는데 사용되는 통계 기법
보통 이항형 문제(즉, 유효한 범주의 개수가 두개인 경우)를 지칭할 때 사용
- 이항형 -> 종속 변수의 결과가 성공/실패처럼 2개의 카테고리로 나뉘는 것
- 다항형 -> 결과가 그 이상의 카테고리로 나뉘는 것

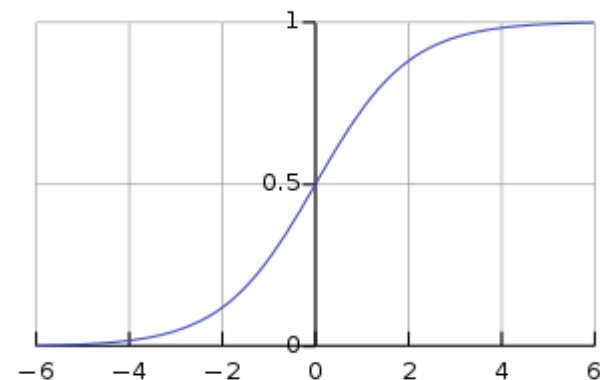
- 선형모델의 특수케이스로 볼 수 있을만큼 유사하지만 가장 큰 차이점 2가지는

1. y 의 결과를 $[0, 1]$ 로 제한
2. 종속변수가 이진적으로 조건부 확률의 분포가 정규분포 대신 이항 분포를 따름

가장 대표적인 연결함수

- 로지스틱 모형: $g(x) = \frac{e^x}{1 + e^x}$
- 검벨 모형: $g(x) = e^{-e^x}$

표준함수



선형

- 선형대수학 -> **linear** algebra
- 선형 / 대수학 -> **linear** / algebra
- Linear -> 선형성 -> 직선처럼 똑바른 도형, 또는 그와 비슷한 성질을 갖는 대상
- 선형방정식 = 일차방정식 = 직선

의사결정나무

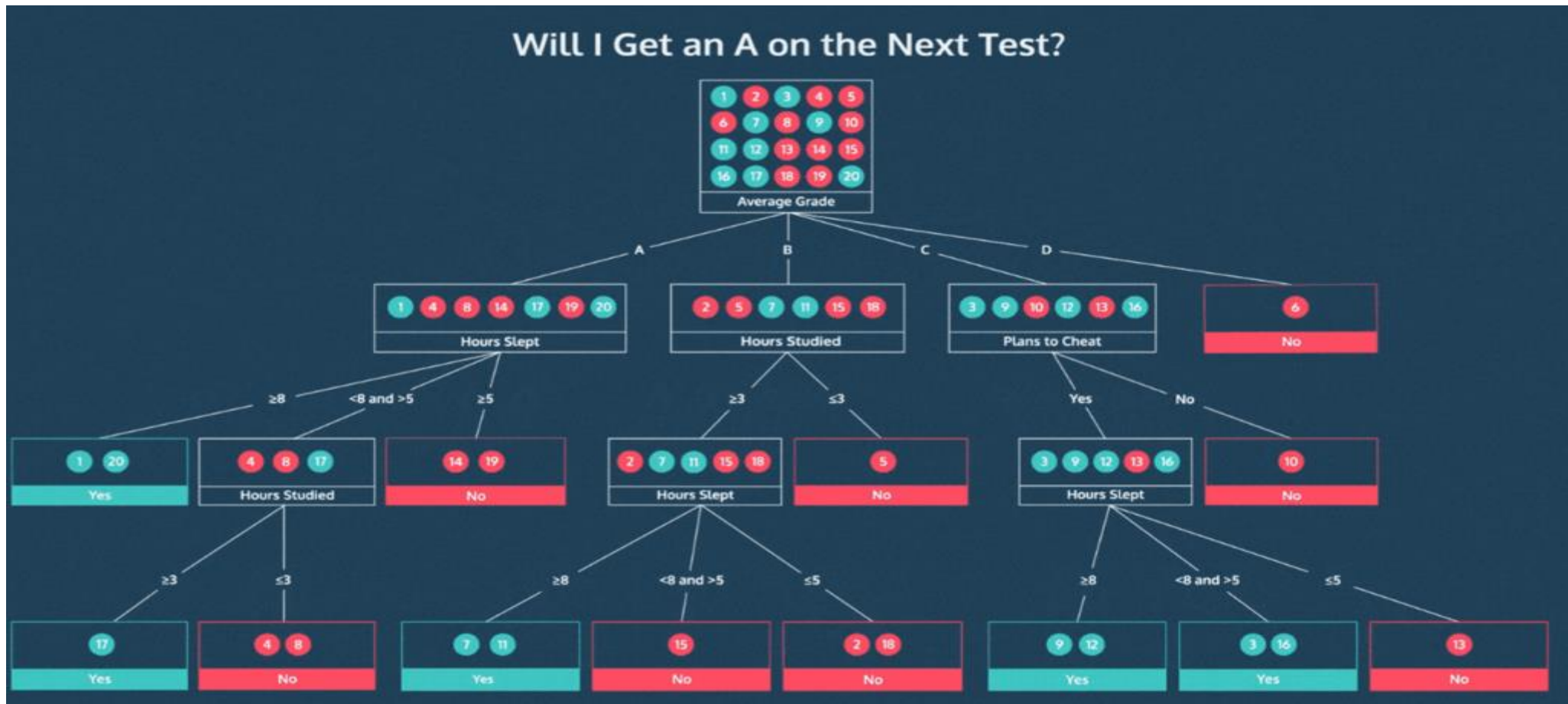
- 의사결정 규칙을 나무(트리)구조로 도표화하여 분류 및 예측을 수행하는 분석방법

- 형성과정

1. 알고리즘 – 목적과 구조에 따라 적절한 분리기준과 정지규칙을 지정
2. 가지치기 – 오류의 위험이 높거나 부적절한 추론규칙을 제거
3. 타당성평가 – 모형평가도구 또는 평가용 데이터에 의한 교차 타당성을 이용하여 평가
4. 해석 및 예측 – 나무를 해석하고 예측 모형을 구축

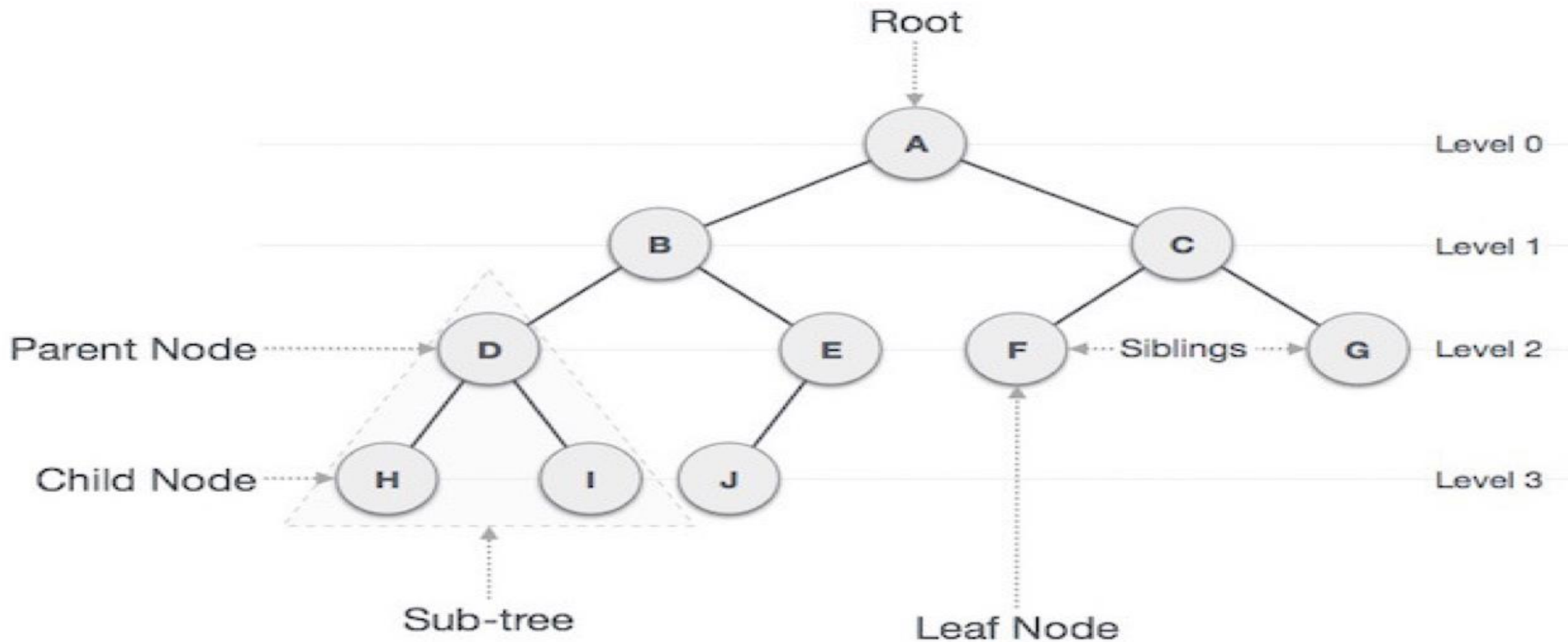
의사결정나무 분리기준

- 분리기준 : 하나의 부모마디마다 자식마디들이 형성될 때, 입력변수의 선택과 범주의 병합이 이루어질 기준

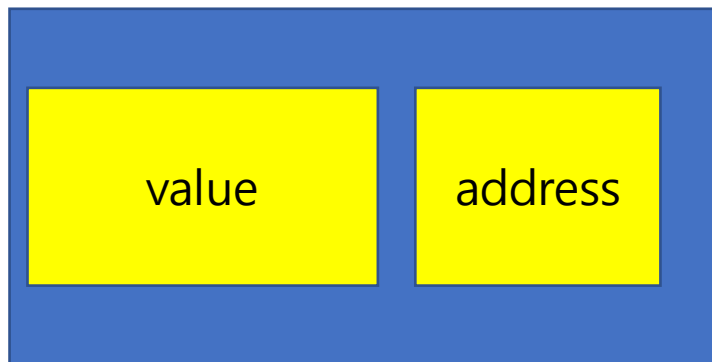


트리

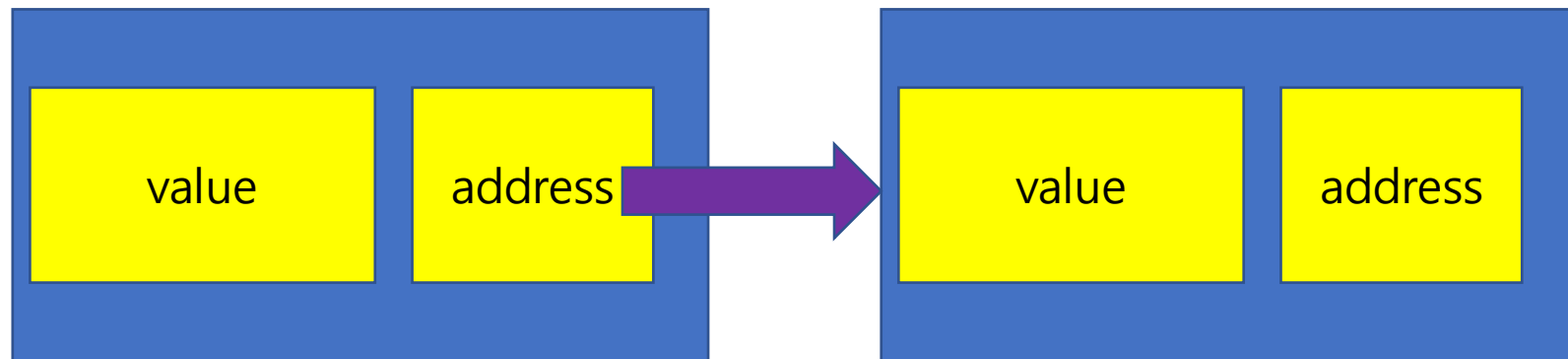
- 노드로 이루어진 자료구조



노드



• 연결리스트



탐색적 데이터 분석

- 수집한 데이터가 들어왔을 때 이를 다양한 각도에서 관찰하고 이해하는 과정

- 지난주차 타이타닉을 예시로

탑승객들의 정보가 주어졌을 때 -> 성별, 나이, 티켓, 좌석 등등과 생존률과의 관계를 알아
보기 위해 그래프를 그리거나 연산을 진행하여 생존확률을 예측하는 것

실제 생존여부와 비교하여 얼마나 정확히 예측했는가를 알 수 있다.