

< Supervised Learning >

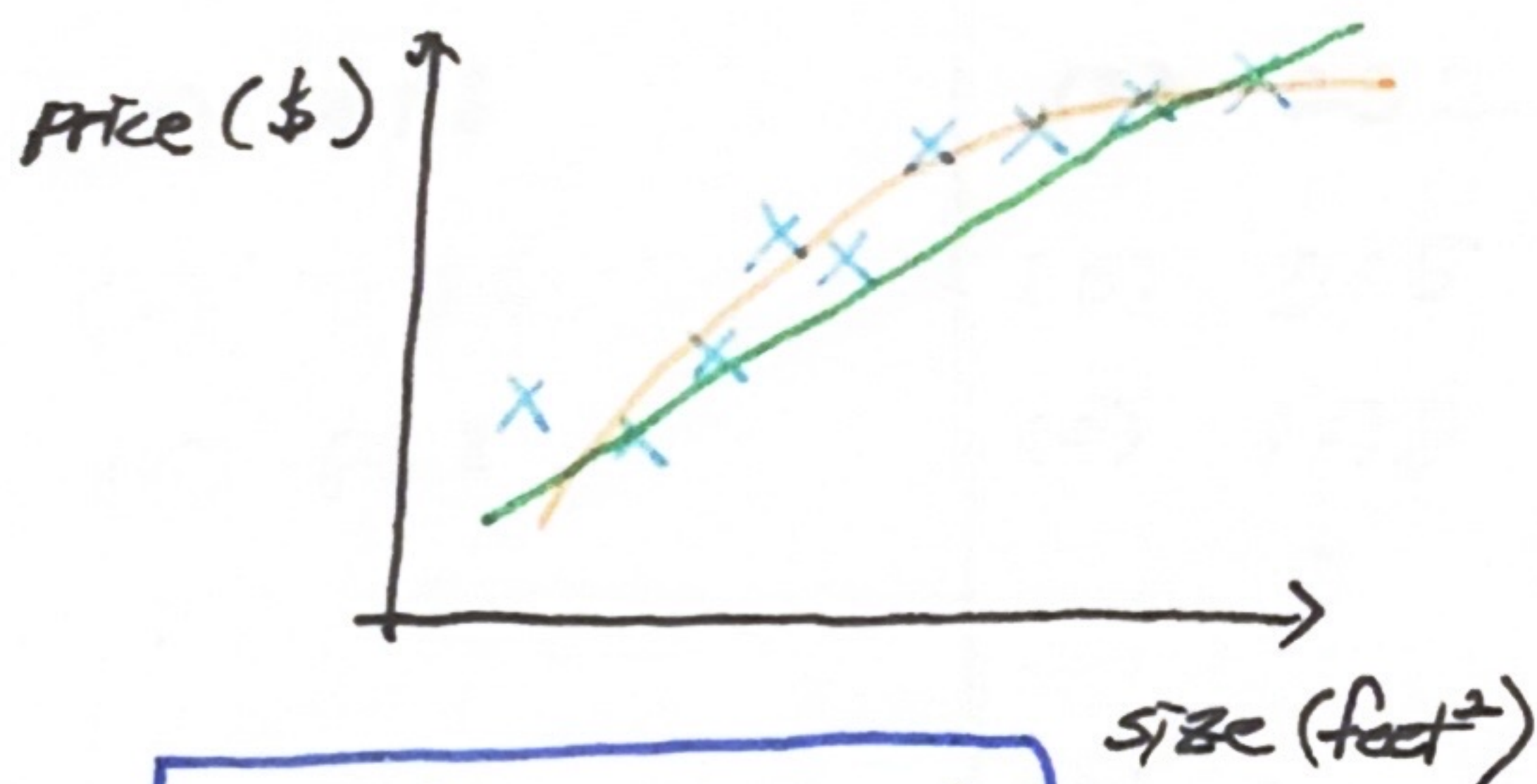
- 특정 input에 대해 "정답 (label)" output이 있는 데이터셋이 주어지는 경우
- 프로그램은 이 정보로부터 input과 output의 관계를 유추
- regression, classification

① Regression (회귀) : 독립변인이 종속변인에 영향을 미치는지 알아보거나 한때 신사라는 방법

- 단순 선형 회귀분석은 독립변수 X 에 대하여 종속변수 Y 들 사이의 관계는 회귀 모형을 이용하여 규명하는 것
- 규명된 함수를 이용하여 독립변수들의 변화에 따른 종속변수의 변화를 예측하는 분석
- 즉, 주어진 input 변수를 output 변수에 대응시키는 어떤 연속함수를 찾는 과정

(output은 continuous 값)

ex) 집 넓이에 따른 적절한 집 값 추정 (housing price prediction)



ex) Predicting final exam score based on time spent

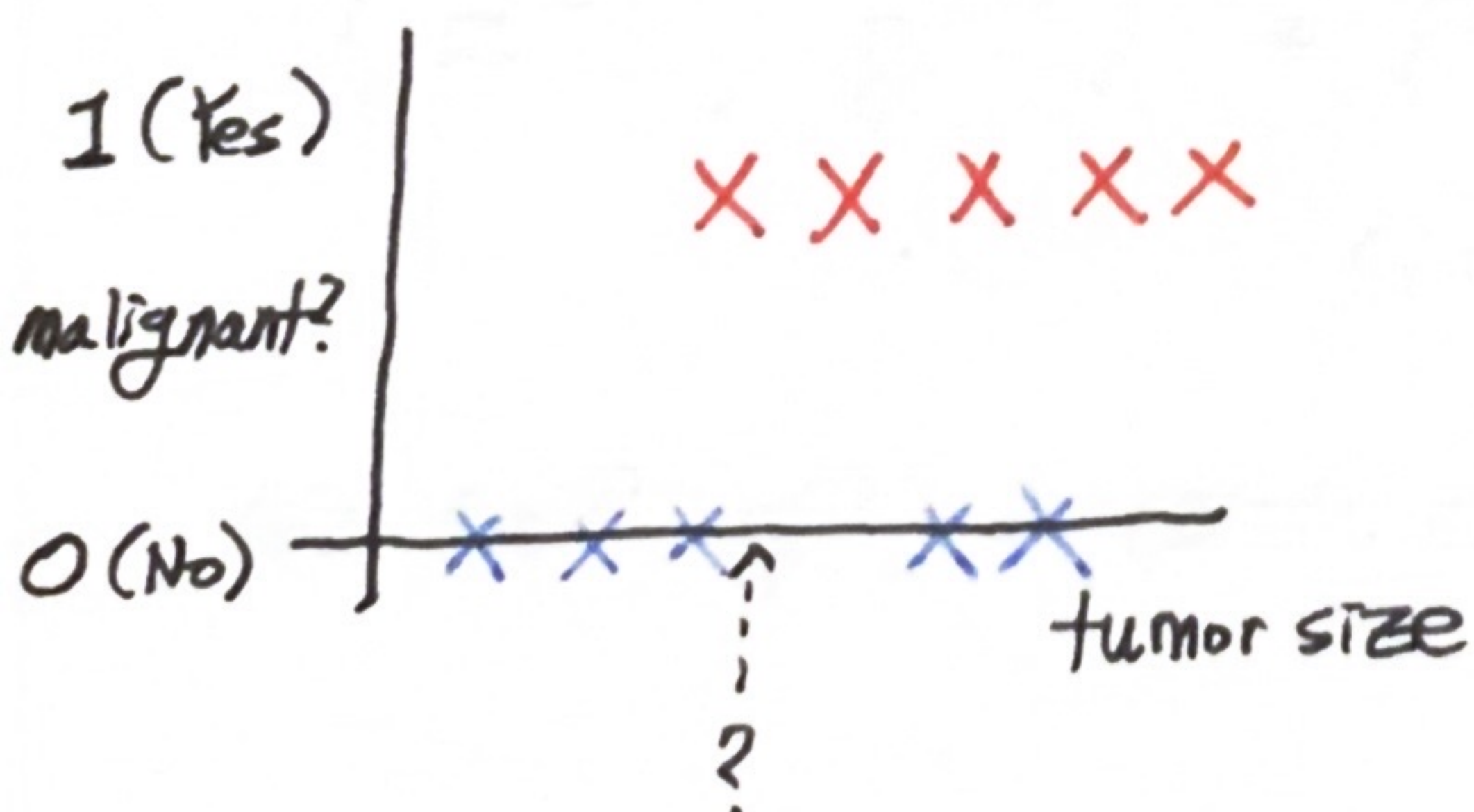
X (hours)	Y (Score)
10	90
9	80
3	50
2	30

0 ~ 100
(continuous)

② Classification (분류)

- 주어진 input 변수가 어느 discrete category에 속하는지 찾아내는 것 (output은 discrete)

ex) 종양이 악성인지 양성인지 진단



ex) Pass / Fail based on time spent

⇒ binary classification

ex) Letter grade (A, B, C, D, F) based on time spent

⇒ multi-label classification