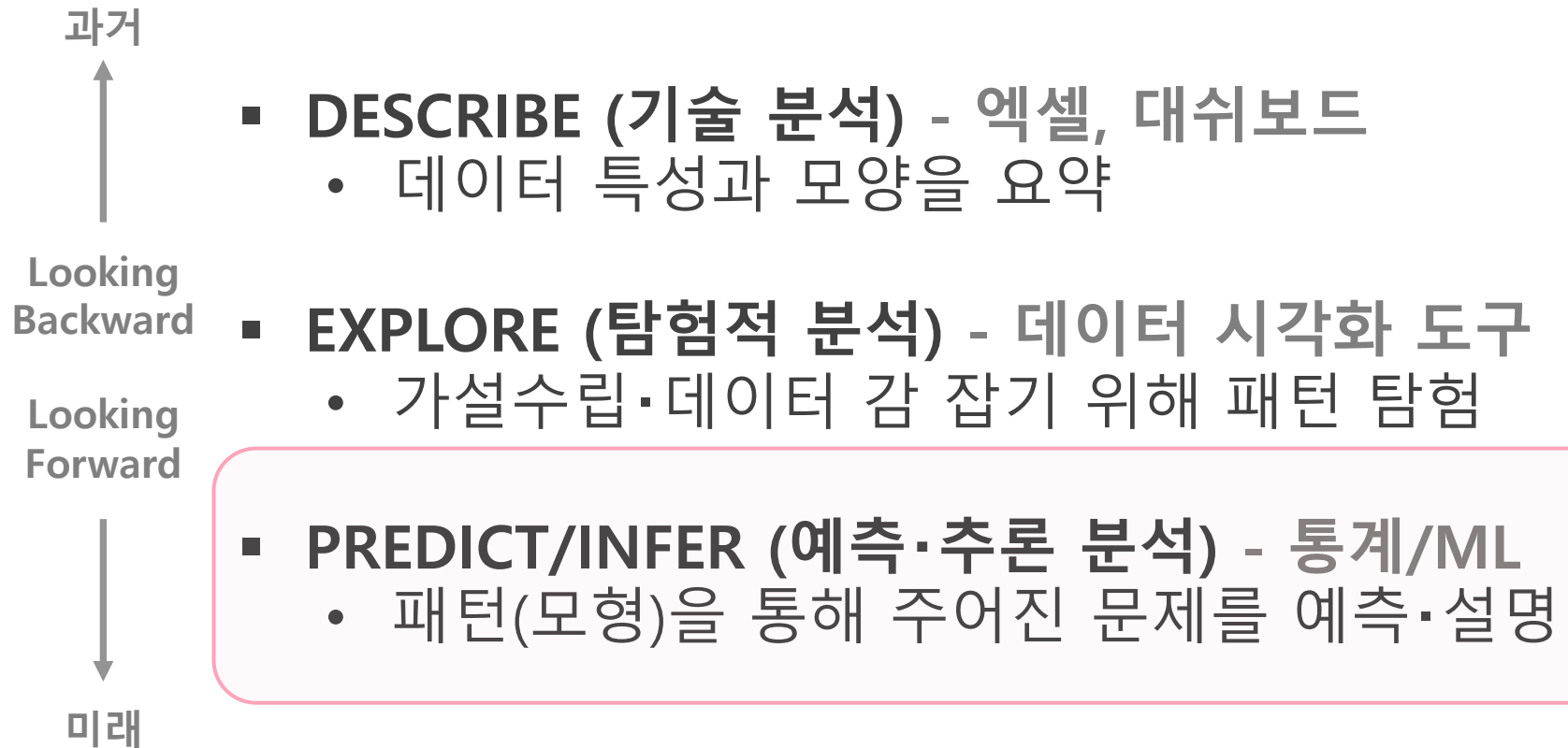


Module III

Machine Learning and Enterprise Decision-Making

아이디케이스퀘어드 양승준 / sidney.yang@idk2.co.kr
<https://www.heartcount.io>

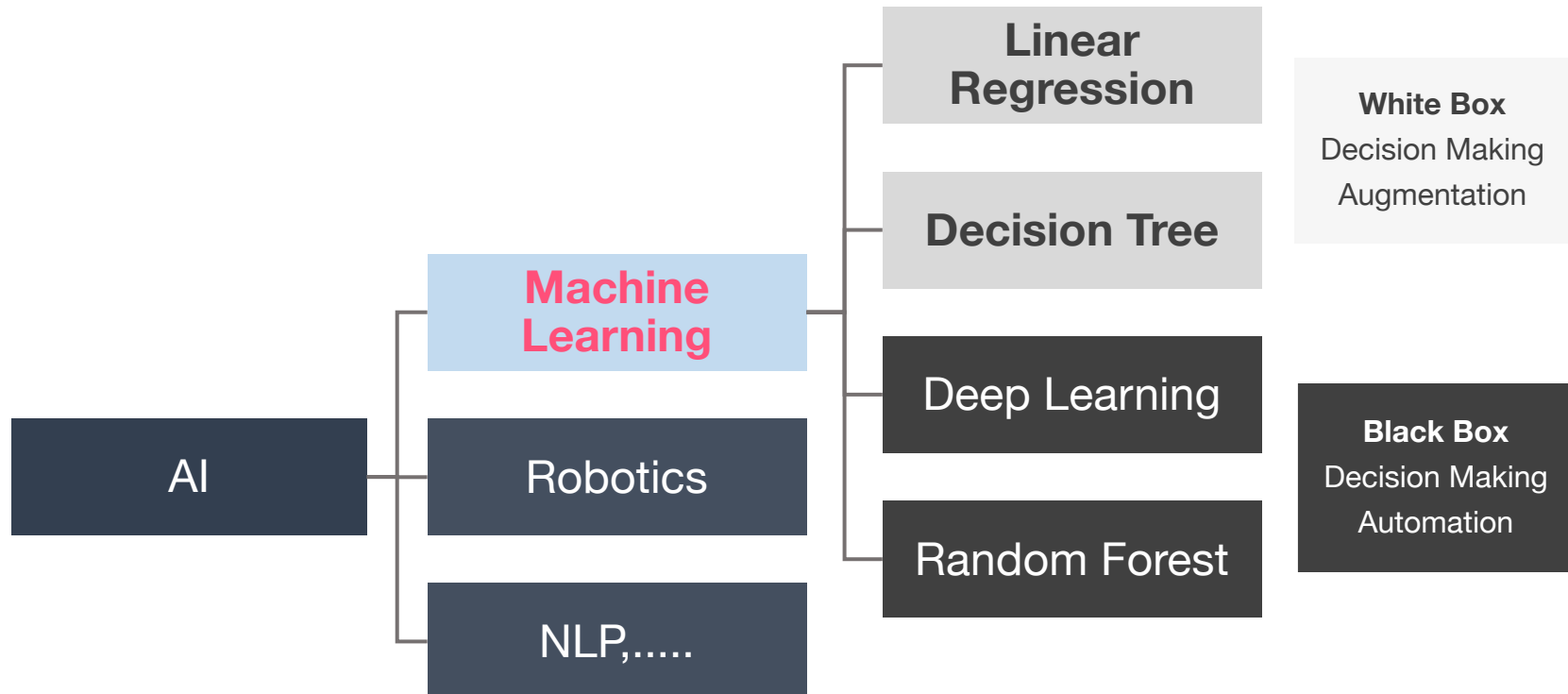
데이터 분석 주요기술



Machine Learning and Decision Making

Machine Learning: Decision Making Technology

- 인공지능의 한 분야로 알고리즘에 아주 **제한된 지시**와 데이터를 제공, 통계적 **학습을 통해** 발견한 패턴을, 의사결정에 활용하는 기술
- **제한된 지시**: 기계에게 언어로 기술(**codify**)하기 힘든 것을 정답(Label)을 알고 있는 과거 데이터를 제공하여 학습하도록 함
- **플라니의 역설**: 할 줄은 아는데 어떻게 하는지 말 못함 (언어의 해상도가 인식의 해상도보다 낮다)



Decision Tree vs. Random Forest

Trade Off Between Interpretability and Accuracy
모형의 설명력과 해석력 사이에 길항관계가 존재

Interpretability
White Box

Decision Tree

input data



tree



output: "yes"

Accuracy
Black Box

Random Forest

input data



tree 1



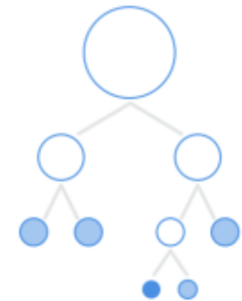
output: "yes"

tree 2



output: "yes"

tree 3



output: "no"

final vote = 'yes'

Machine Learning and Enterprise Decision Making

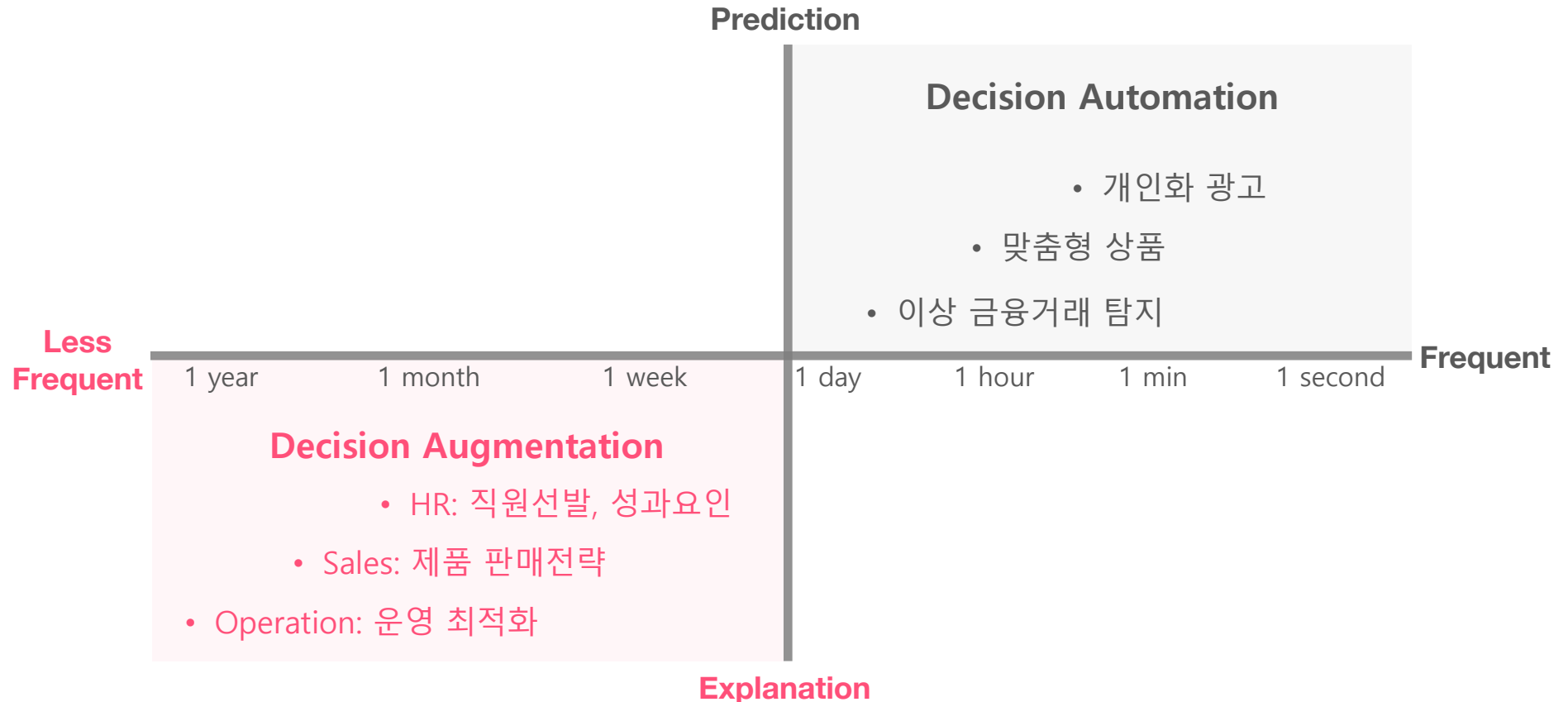
Are You in Prediction Business or **Explanation Business**?

빈번한 덜 중요한 결정

데이터 → 기계학습 → 예측 통한 의사결정 자동화

덜 빈번한 중요한 결정

데이터 → 기계학습 → 사실·증거 기반 더 좋은 결정



Supervised Machine Learning (지도 학습)

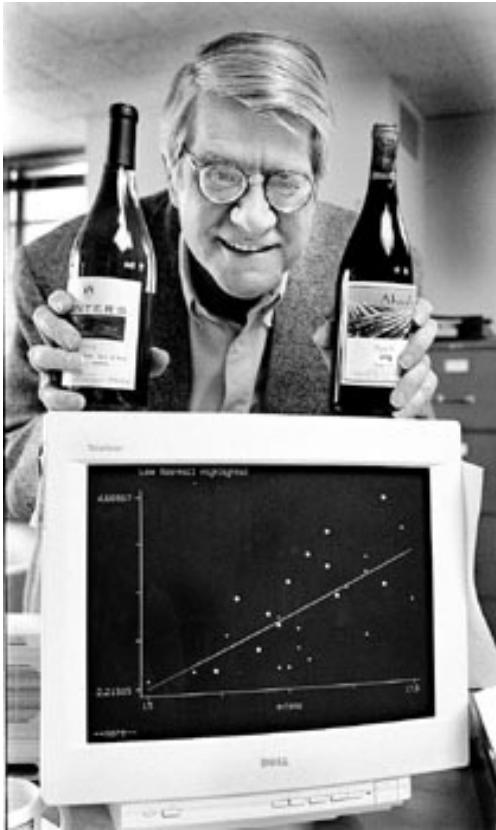
설명/예측하고 싶은 문제에 관련하여
정답을 알고 있는 데이터가 충분히 있는 경우
X로부터 Y를 유추하는 패턴을 찾는 것

X (Input; Feature)	Y (Output; Label)	APPLICATION
Voice recording	Transcript	Speech recognition
Transaction records	Fraudulent (yes/no)	Fraud detection
Emails	Spam (yes/no)	Spam filtering
Ad + User Profile	Click (yes/no)	Personalized AD
Faces	Names	Face recognition
Korean	English	Language Translation

Supervised Machine Learning

Supervised Machine Learning (지도 학습)

모형이 투명한 경우(White Box), 설명·예측 둘 다 활용 가능



선형회귀 알고리즘

$$Y = a + b X_1 + c X_2 + d X_3$$

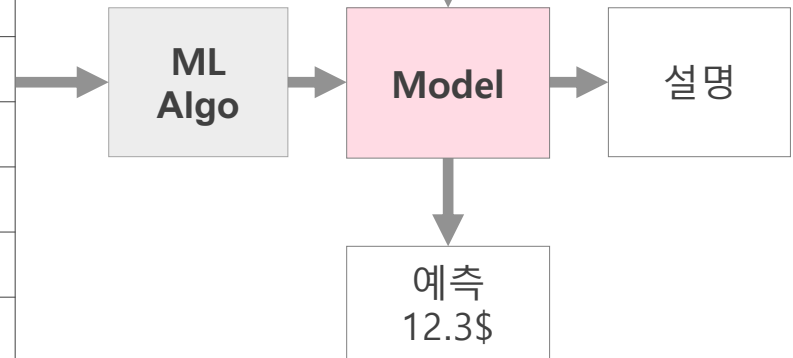
Ashenfelter's Wine Formula

$$\text{Price} = 12 + 0.001 \text{ 겨울강수량} + 0.06 \text{ 평균온도} - 0.004 \text{ 수확철강수량}$$

Training Data Set

X_1	X_2	X_3	Y
겨울강수량	수확철강수량	평균온도	와인가격
13	35	35	9.5
22	25	25	3.5
25	21	21	3.2
11	18	18	3.5
47	45	45	4.7
.	.	.	.

겨울강수량	수확철강수량	평균온도	와인가격
35	25	21	?

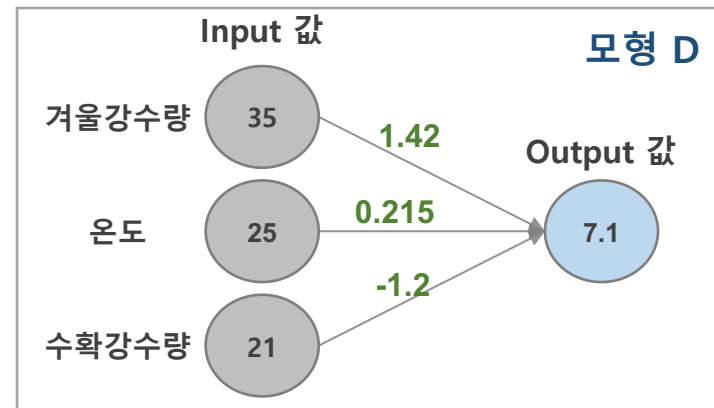
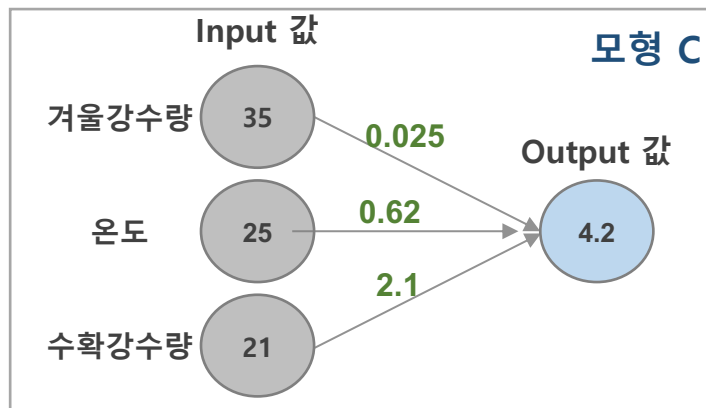
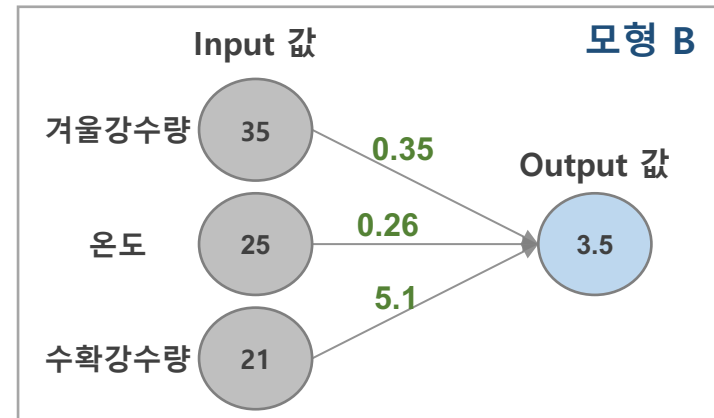
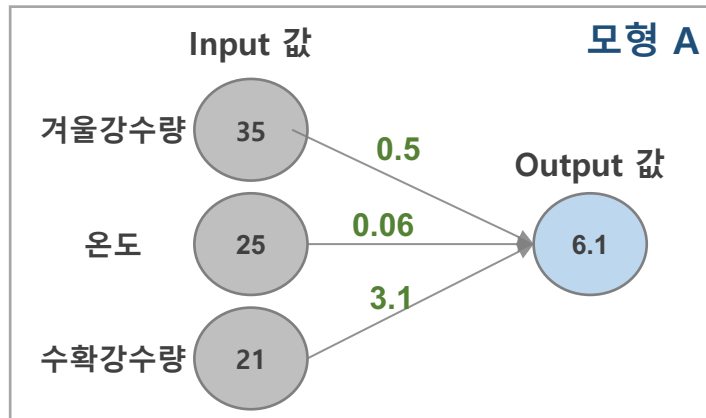


White Box vs. Black Box

Black Box: 더 정확할지언정 인간이 이해하기 힘들다.

$$\text{가격} = 15.4\$ = 0.6 \times \text{A} + 0.1 \times \text{B} + 0.25 \times \text{C} + 0.05 \times \text{D}$$

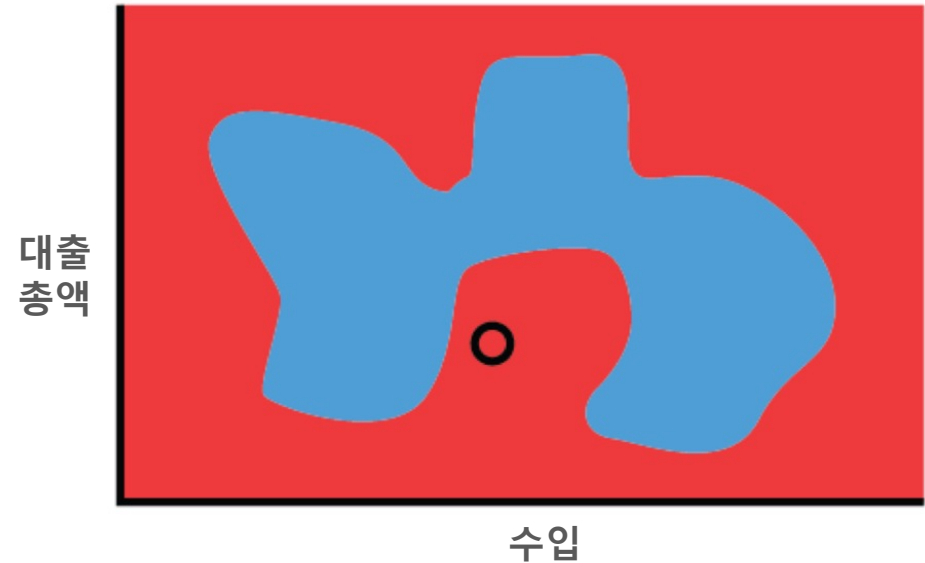
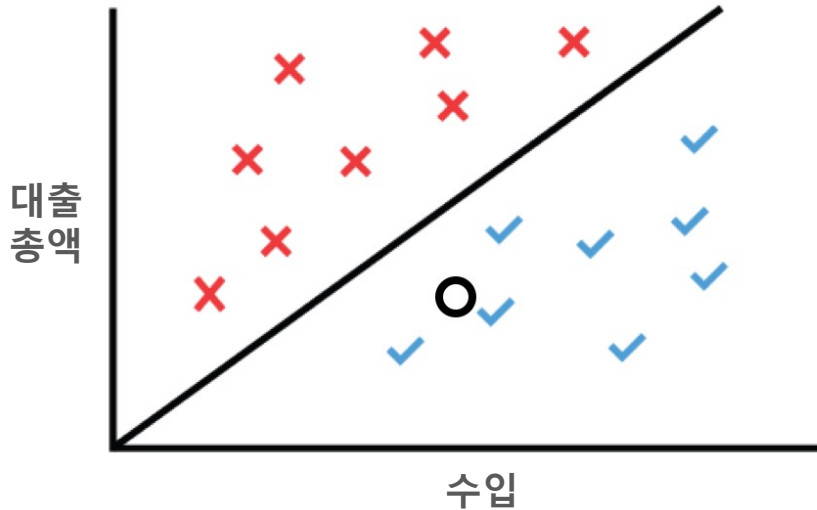
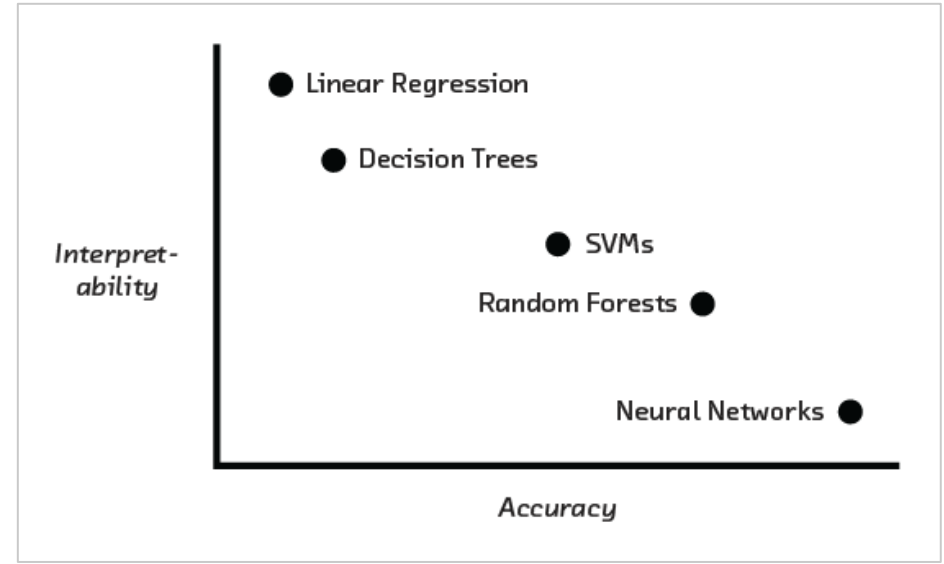
(A: 6.1, B: 3.5, C: 4.2, D: 7.1)



Model Interpretability

정확함 vs. 올바름

Model should be Accurate
for the Right Reason



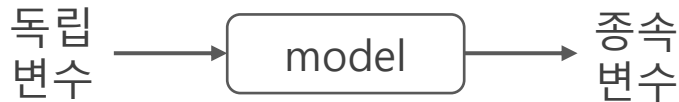
Statistics vs. Machine Learning

“Machine Learning is essentially a form of **applied statistics**”

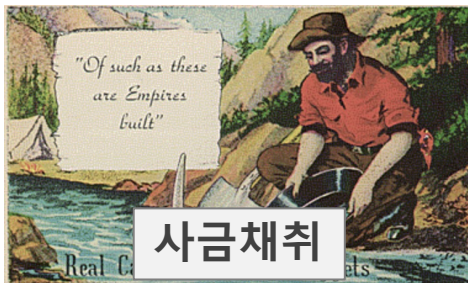
“Machine Learning is statistics scaled up to **big data**”



통계 분석



- **연역적 추론**(Deductive Reasoning): 가설수립 → 가설검증
- **모형의 타당성**: 가설 일반화를 위한 모형의 타당성·재현성이 중요



기계학습



- **귀납적 추론**(Inductive Reasoning): 데이터 → 패턴
- **모형의 유용성**: 의사결정에 활용하기 위한 모형의 유용성이 중요



둘다 데이터를 통해 문제를 해결하는 데 사용됨
기술과 방법의 차이라기보다는 분석 목표의 차이