Intro

Machine Learning

Microsoft Student Partner
백진헌

Jinheon Baek

Korea University,
Computer Science & Engineering undergraduate student at junior.

Data Mining / Machine Learning / Deep Learning / Cloud Platform

[현재]
Microsoft Student Partners (2016.09 ~)
고려대학교 데이터마이닝 연구실 학부연구생 (2018.03 ~)
Jobshopper Project Manager(Back-end / Machine Learning / Server, …) (2018.07 ~)

[이전]
SW Maestro 8th 연수생 (2017.08 ~ 2017.12)

Overview

1. Concept of Machine Learning
2. Supervised Learning
3. Unsupervised Learning

Concept of Machine Learning

✓Quick Questionnaire

How many people have **heard** about Machine Learning

Concept of Machine Learning

✓Quick Questionnaire

How many people have **heard** about Machine Learning

How many people have **know** about Machine Learning

Concept of Machine Learning

✓Quick Questionnaire

How many people have **heard** about Machine Learning

How many people have **know** about Machine Learning

How many people are **using** Machine Learning

Concept of Machine Learning

✓Machine Learning?

“ Filed of study that gives computers the ability to learn without being explicitly programmed.”, Arthur Samuel (1959)

Concept of Machine Learning

✓Machine Learning?

“ A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”, Tom Mitchell (1997)

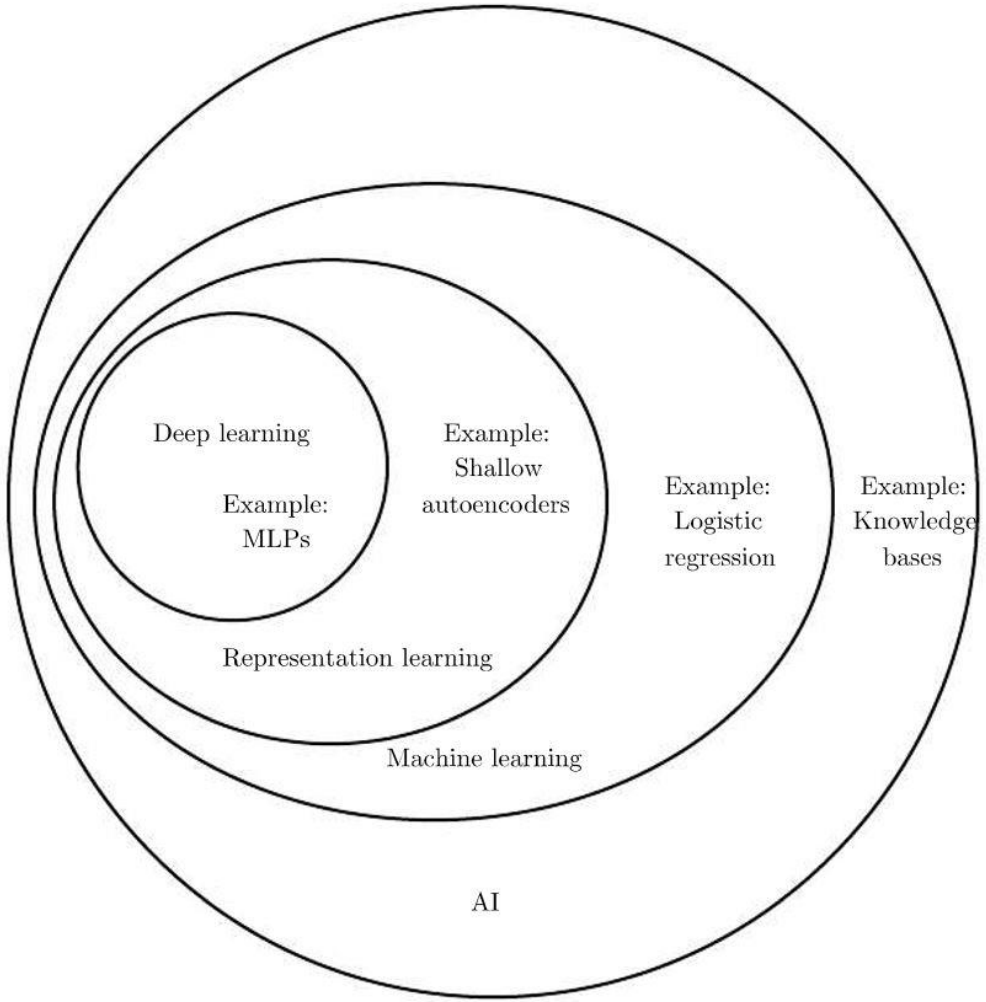
Concept of Machine Learning

✓Machine Learning?

Ex) Spam mail filter,

E(Experience): 스팸인지 아닌지 분류되어 있는 메일을 보는 것
T(Task): 이메일을 스팸인지 아닌지 분류 해보는 것
P(Performance): 스팸인지 아닌지 올바르게 분류되어 있는 메일의 개수

Concept of Machine Learning



Concept of Machine Learning

✓**AI**

“ 사람의 지식을, if - else 문장으로 표현 ”

Knowledge base approach

Concept of Machine Learning

✓AI

None of these projects has led to a major success.

From MIT

Concept of Machine Learning

✓Machine Learning

“ 가공되지 않은 데이터로부터 패턴을 찾는 것 ”

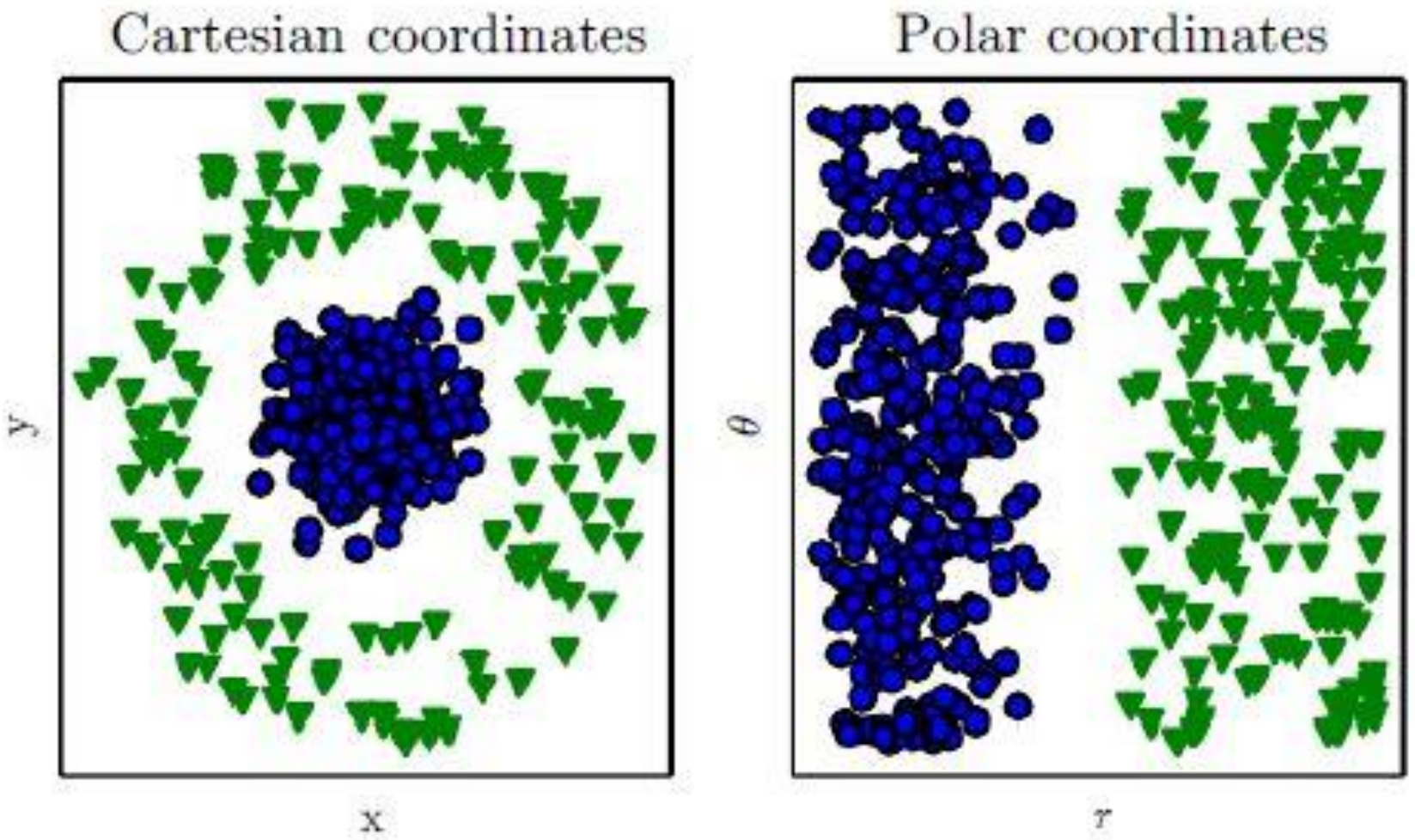
장점: 컴퓨터가 스스로 데이터로부터 패턴을 찾아 나간다.
단점: 컴퓨터에게 주어진 데이터에 과도하게 의존적이다.

Concept of Machine Learning

✓Machine Learning

"Difficult to know what **features** should be extracted."

Concept of Machine Learning



Concept of Machine Learning

✓Representation Learning

Important features 를 추출하는 것이 매우 어려우므로,

컴퓨터만 이해하는 공간을 스스로 만들어,

컴퓨터만 이해하는 공간 안에서 학습을 진행

Concept of Machine Learning

✓Representation Learning – Autoencoder

Autoencoder: encoder + decoder

Encoder: input data into a different representation
Decode: new representation back into the original format

Concept of Machine Learning

✓Representation Learning

데이터로부터,

high-level, abstract features representation

추출이 어려움

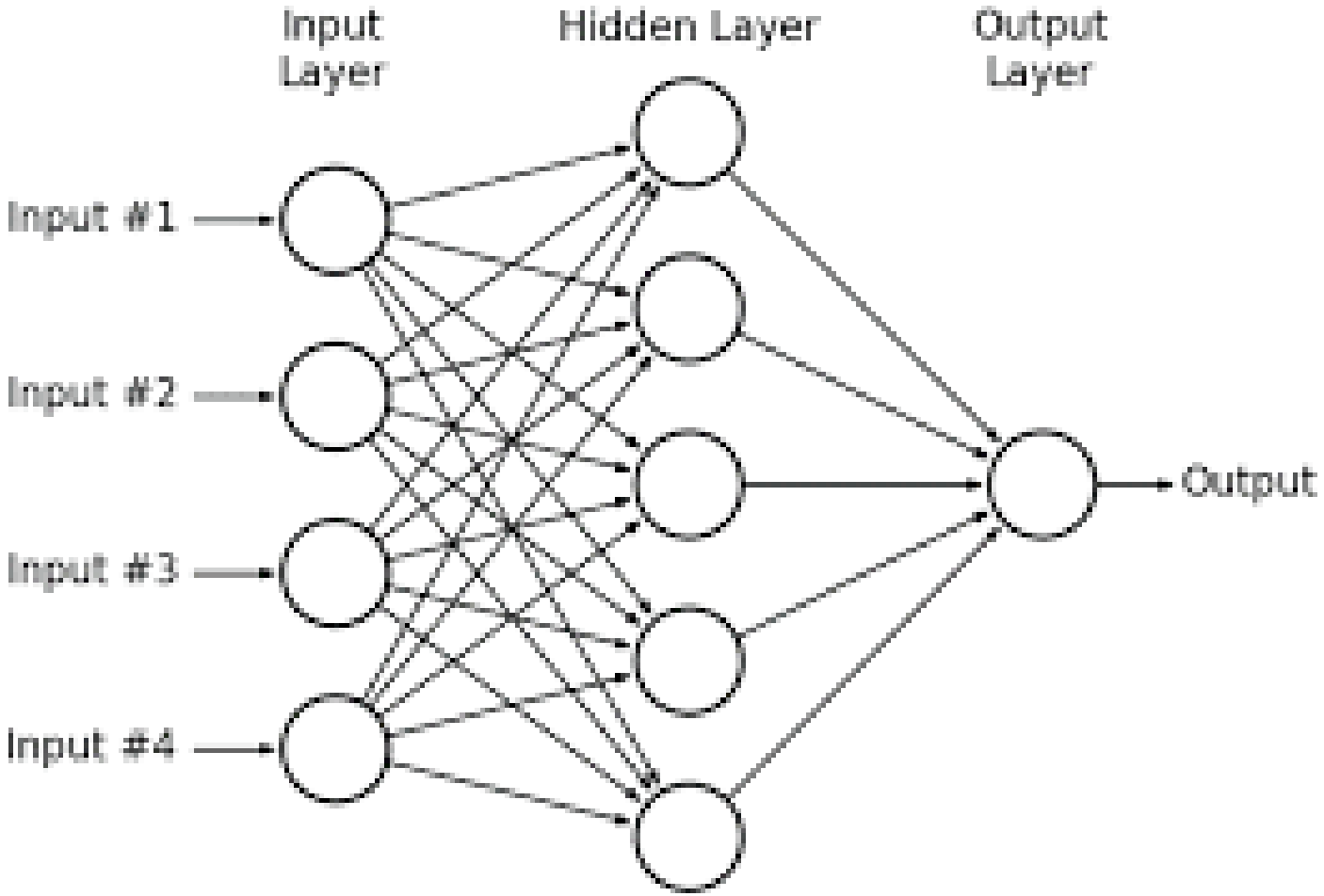
Concept of Machine Learning

✓Deep Learning

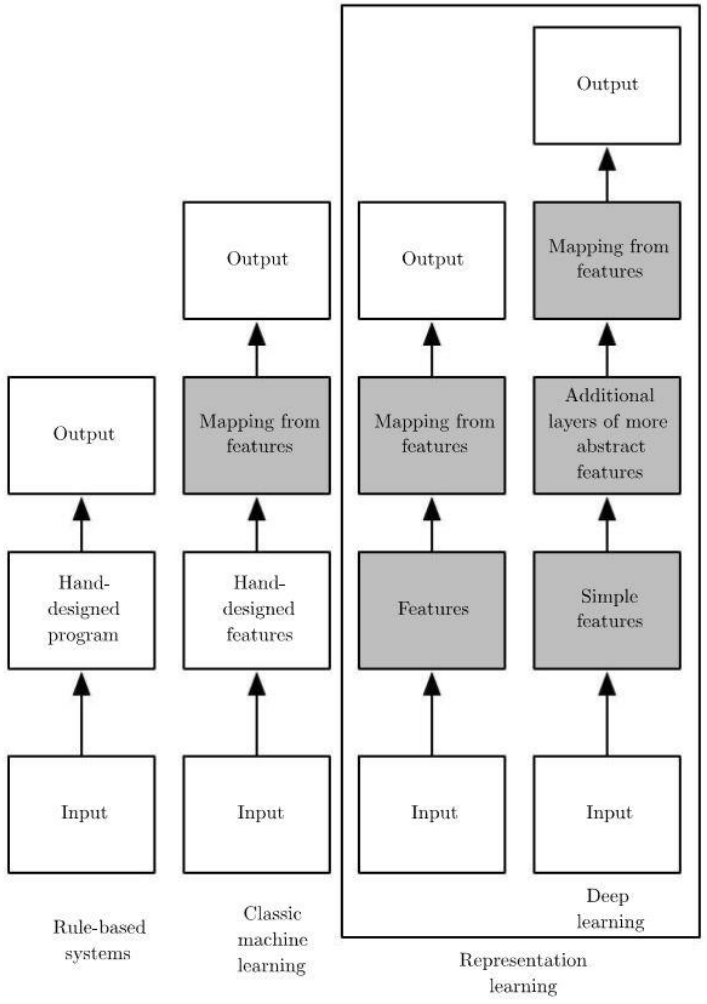
“ 복잡한 모델 표현을 아주 간단한 모델 표현을 쌓아 만든다. ”

Build complex concepts out of simpler concepts (representation)

Concept of Machine Learning



Concept of Machine Learning



Learning

✓Supervised Learning

: 학습에 필요한 데이터와 데이터의 정답이 주어져,
컴퓨터는 데이터의 정답을 기준으로 학습을 수행한다.

: Experience a dataset containing features,
but each example is also associated with a label or target.

Learning

✓Unsupervised Learning

: 가지고 있는 데이터를 이용하여,
현재 데이터가 보이는 구조적인 특성을 학습을 통해 알아본다.

: Experience a dataset containing many features,
then learn useful properties of the structure of this dataset.

Supervised Learning

✓Supervised Learning

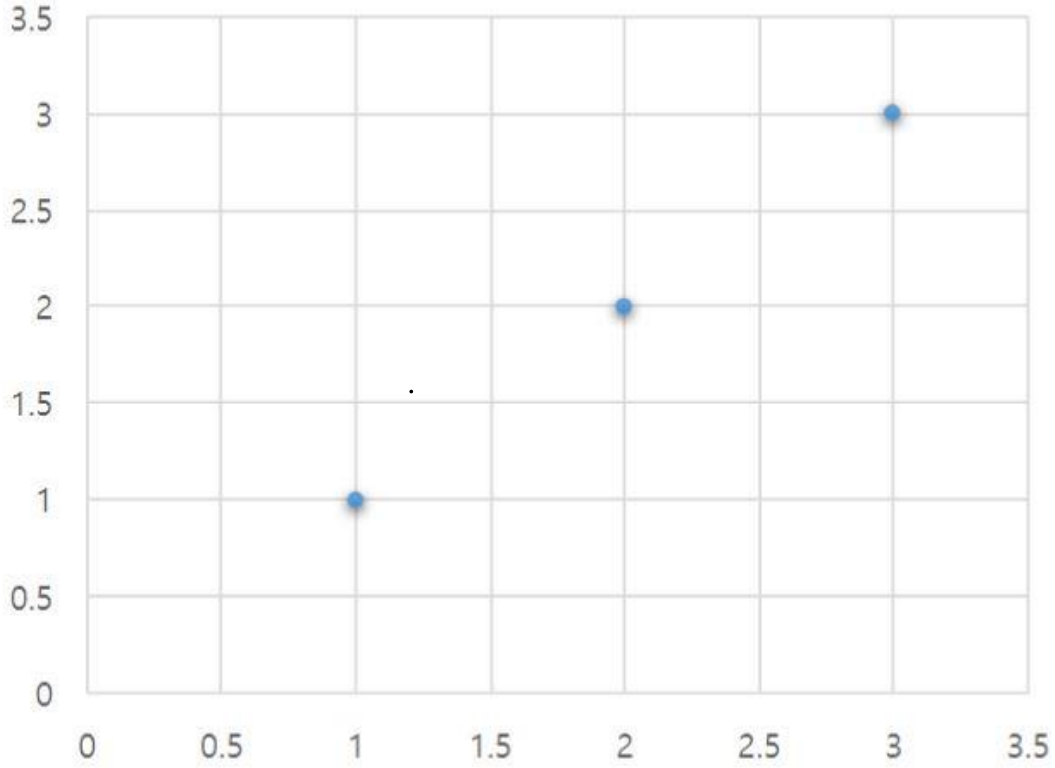
: Regression (Continuous Value – 실수형)

: Classification (Discrete Value – 정수형)

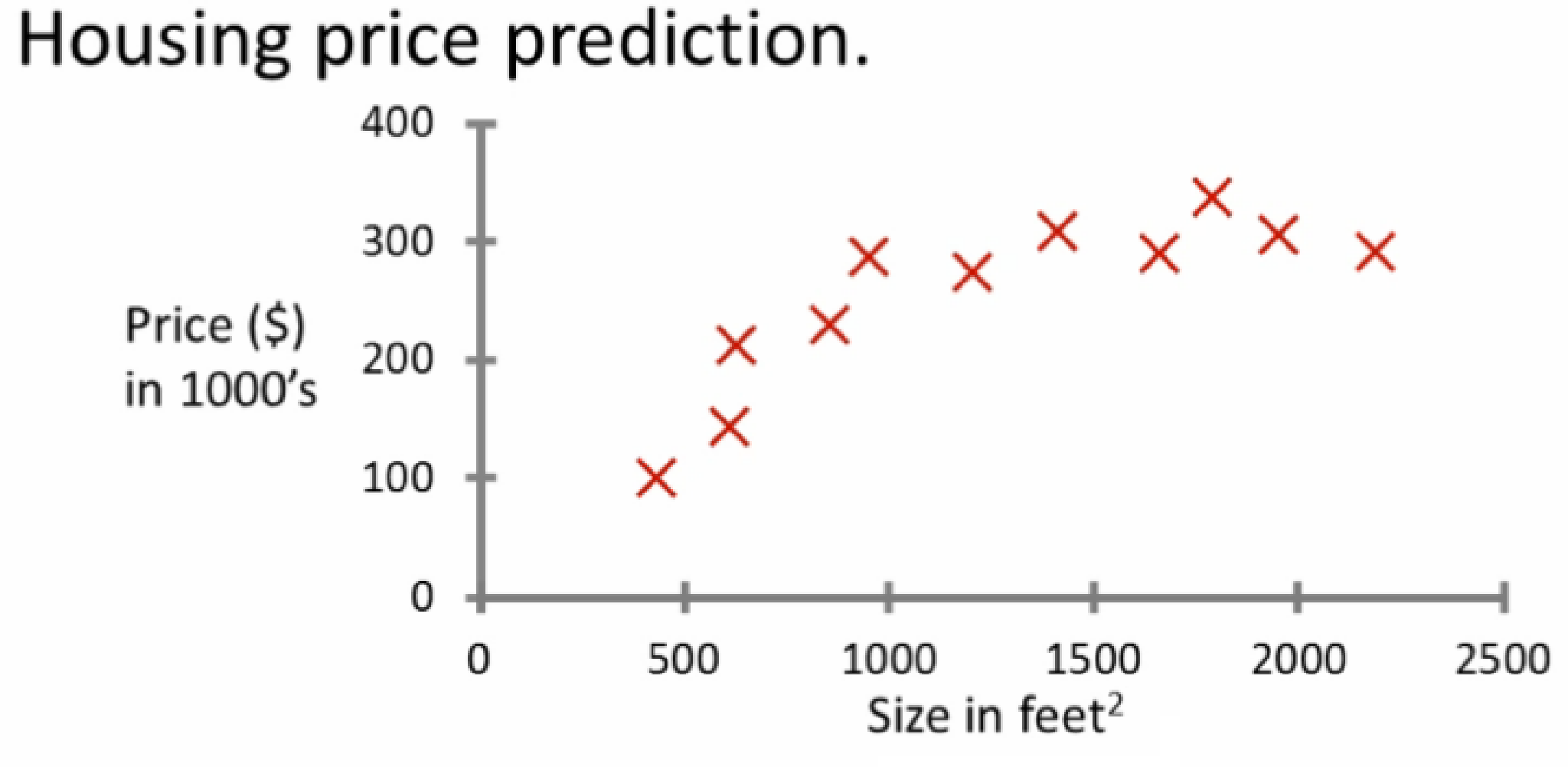
Supervised Learning

✓Supervised Learning – Regression

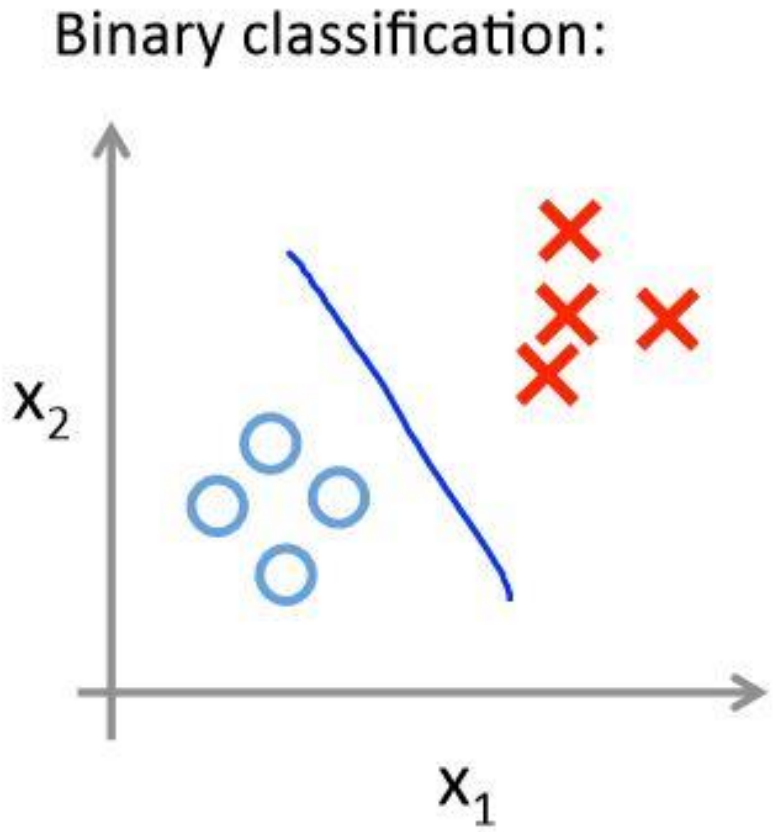
x	y
1	1
2	2
3	3



Supervised Learning



✓Supervised Learning – Classification

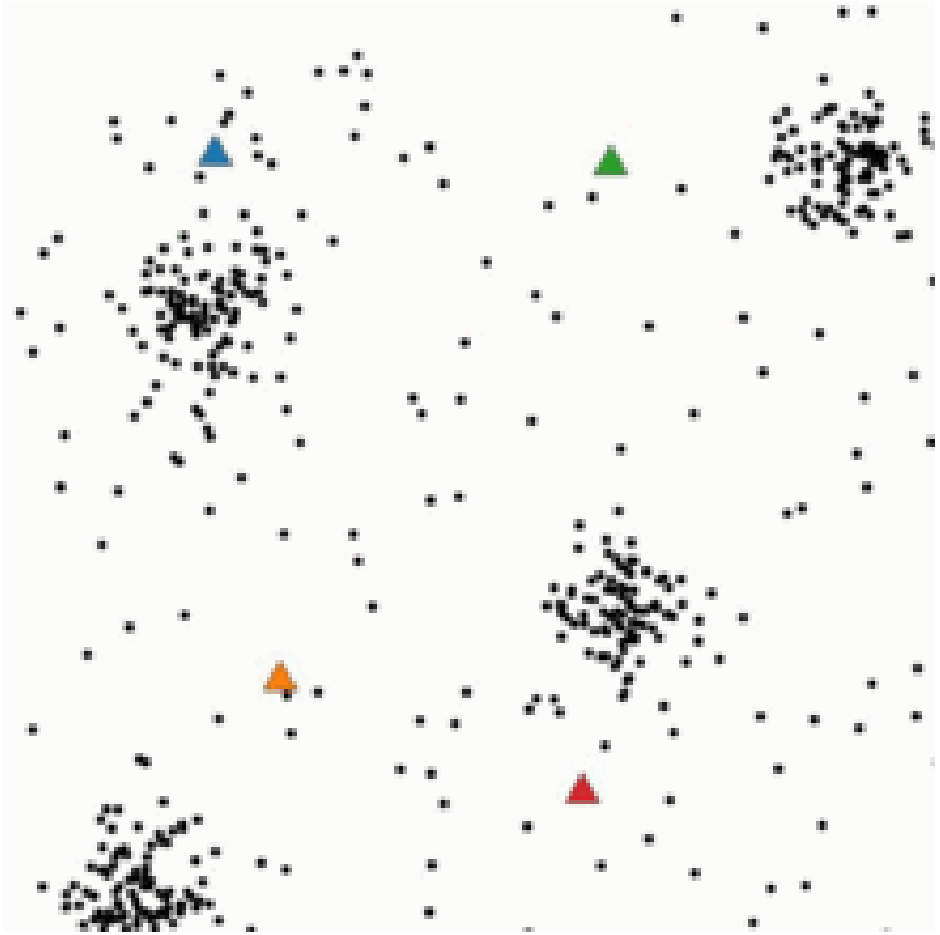


Unsupervised Learning

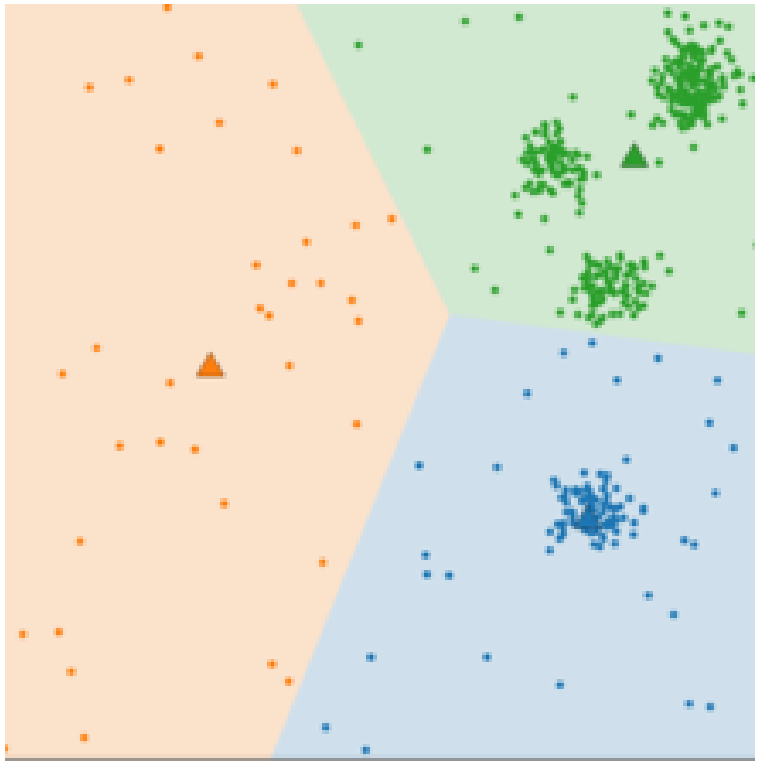
✓Unsupervised Learning

: Clustering

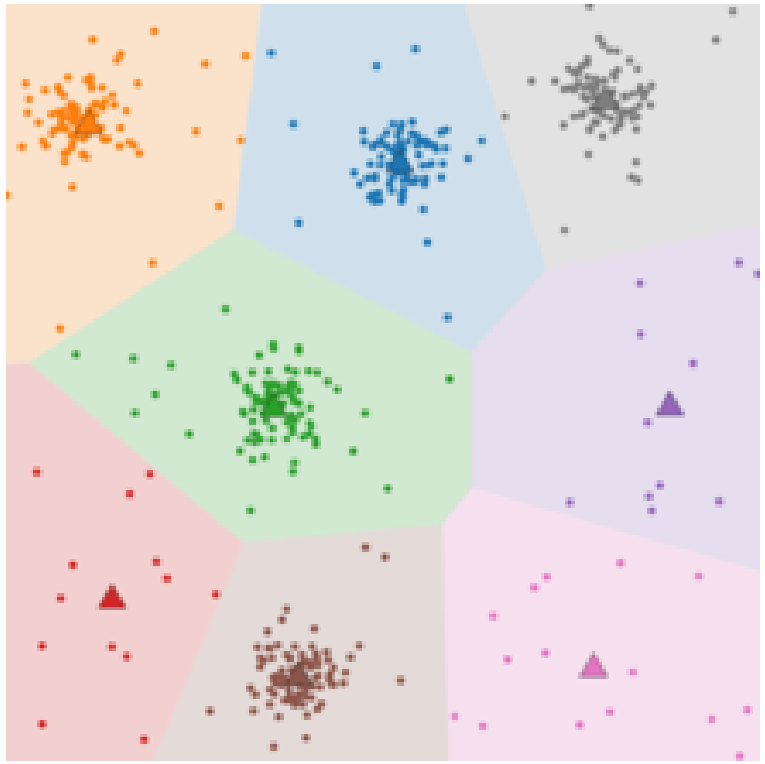
✓Unsupervised Learning – Clustering



Learning

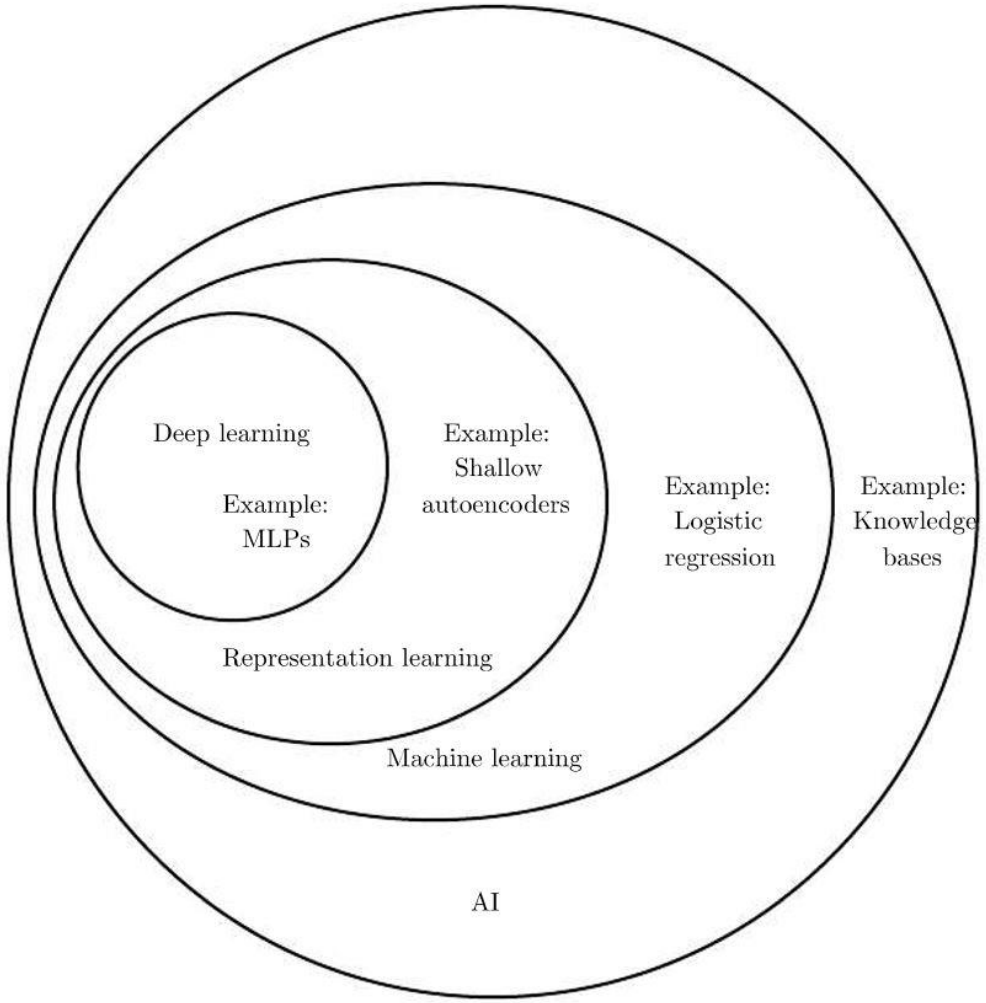


K = 3



K = 8

Warm Up


Deep learning
Example:
MLPs
Example:
Shallow
autoencoders
Example:
Logistic
regression
Example:
Knowledge
bases
Representation learning
Machine learning
AI

Warm Up

- ✓Supervised Learning
- ✓Unsupervised Learning

Q & A

Supervised Learning & HOL

Machine Learning

Microsoft Student Partner
백진헌

백문이 불여일견

Overview

1. **Linear Regression**
   (HOL with Azure Machine Learning)

2. **Neural Network**
   (HOL with Azure Notebook)

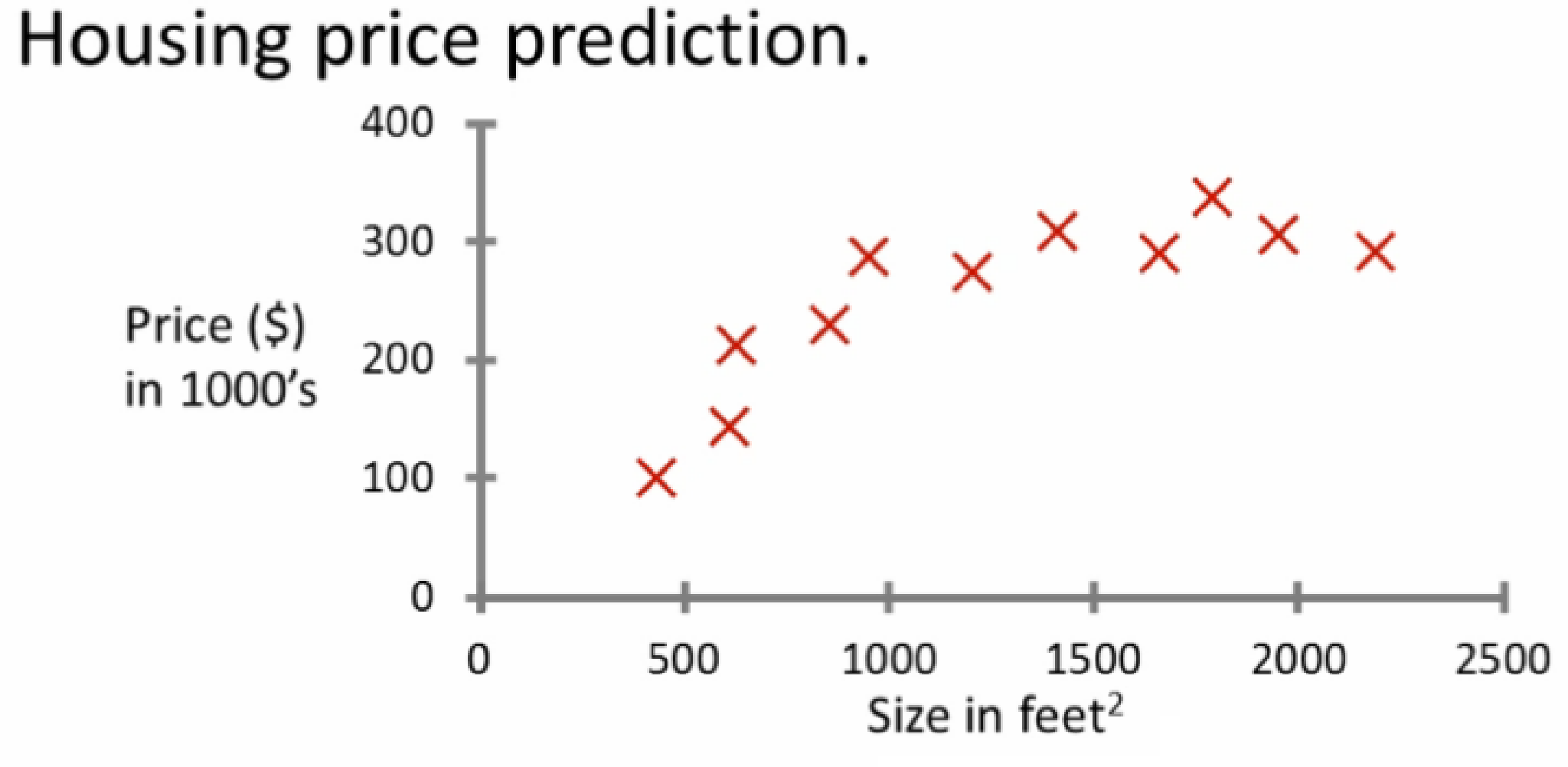
Supervised Learning

[Machine Learning]

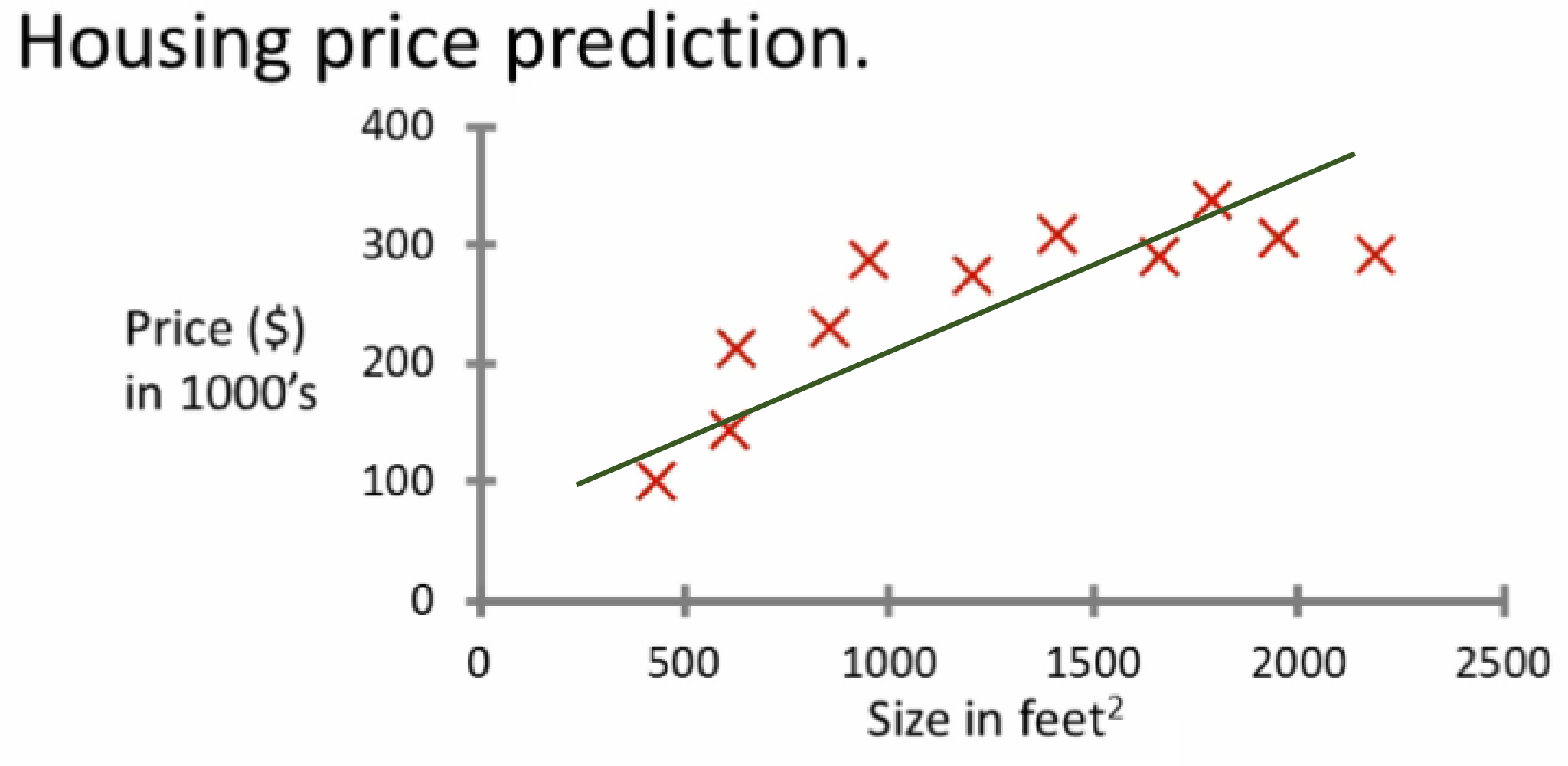
✓Regression

: Linear Regression

Supervised Learning



Supervised Learning



Supervised Learning

✓HOL

Automobile Price Data

Supervised Learning

✓HOL

“Azure Machine Learning Studio”

(https://studio.azureml.net/)

Supervised Learning

✓HOL

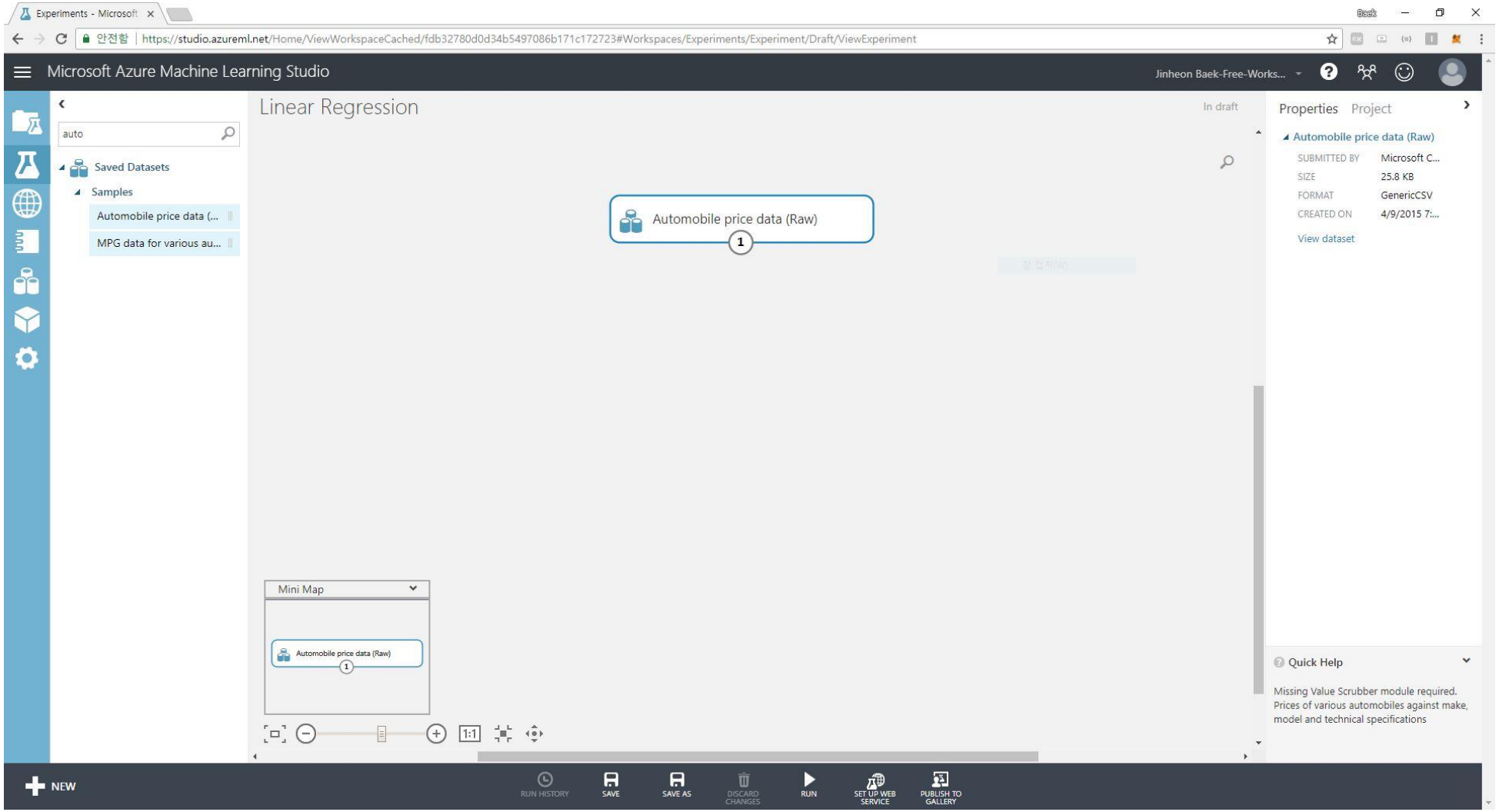
1) New

2) EXPERIMENT

3) Blank Experiment

Supervised Learning

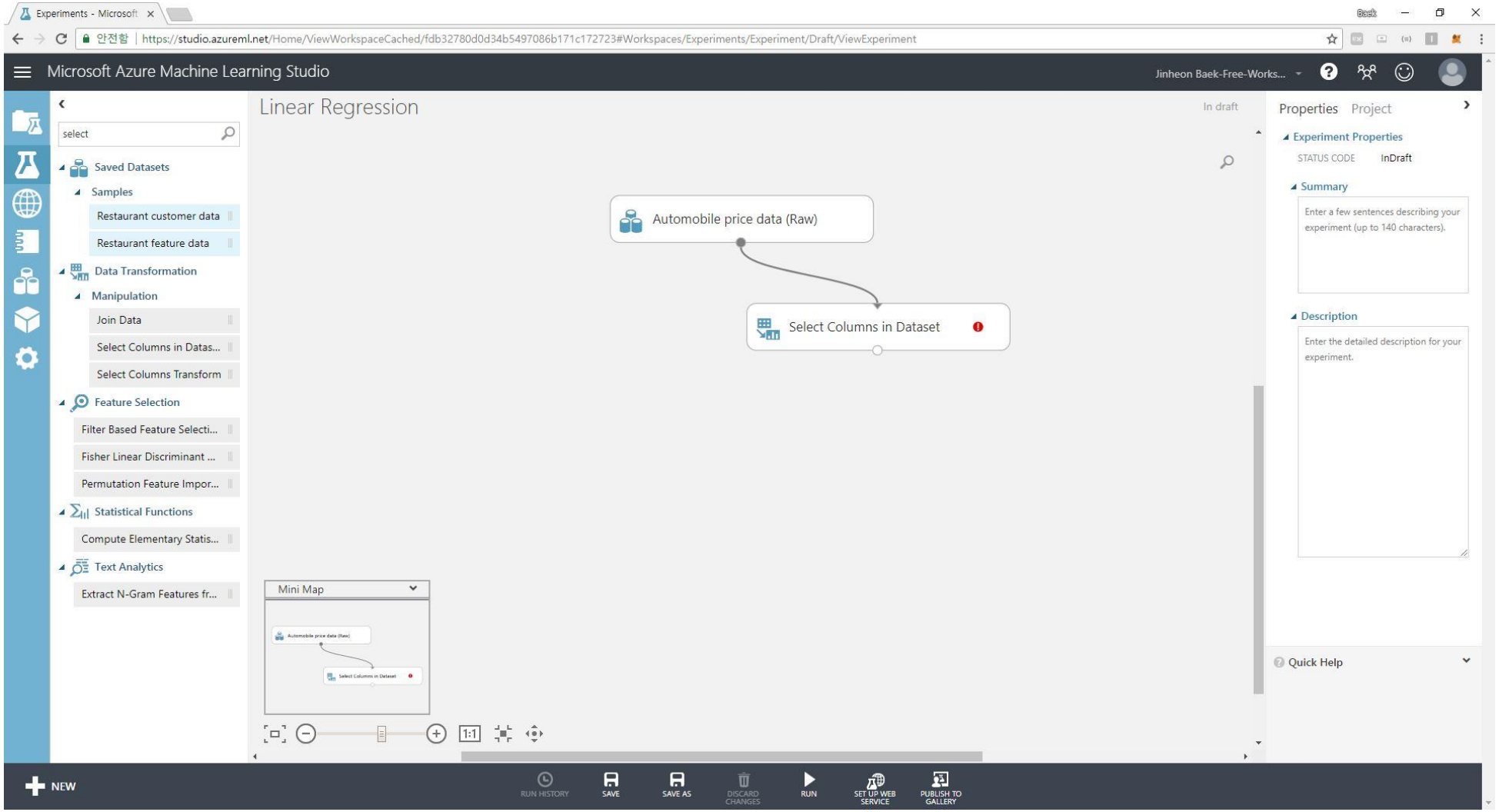
검색창: Automobile price data 선택



Supervised Learning

검색창: Select Columns in Dataset 선택

Automobile price data -> Select Columns ... 이어주기

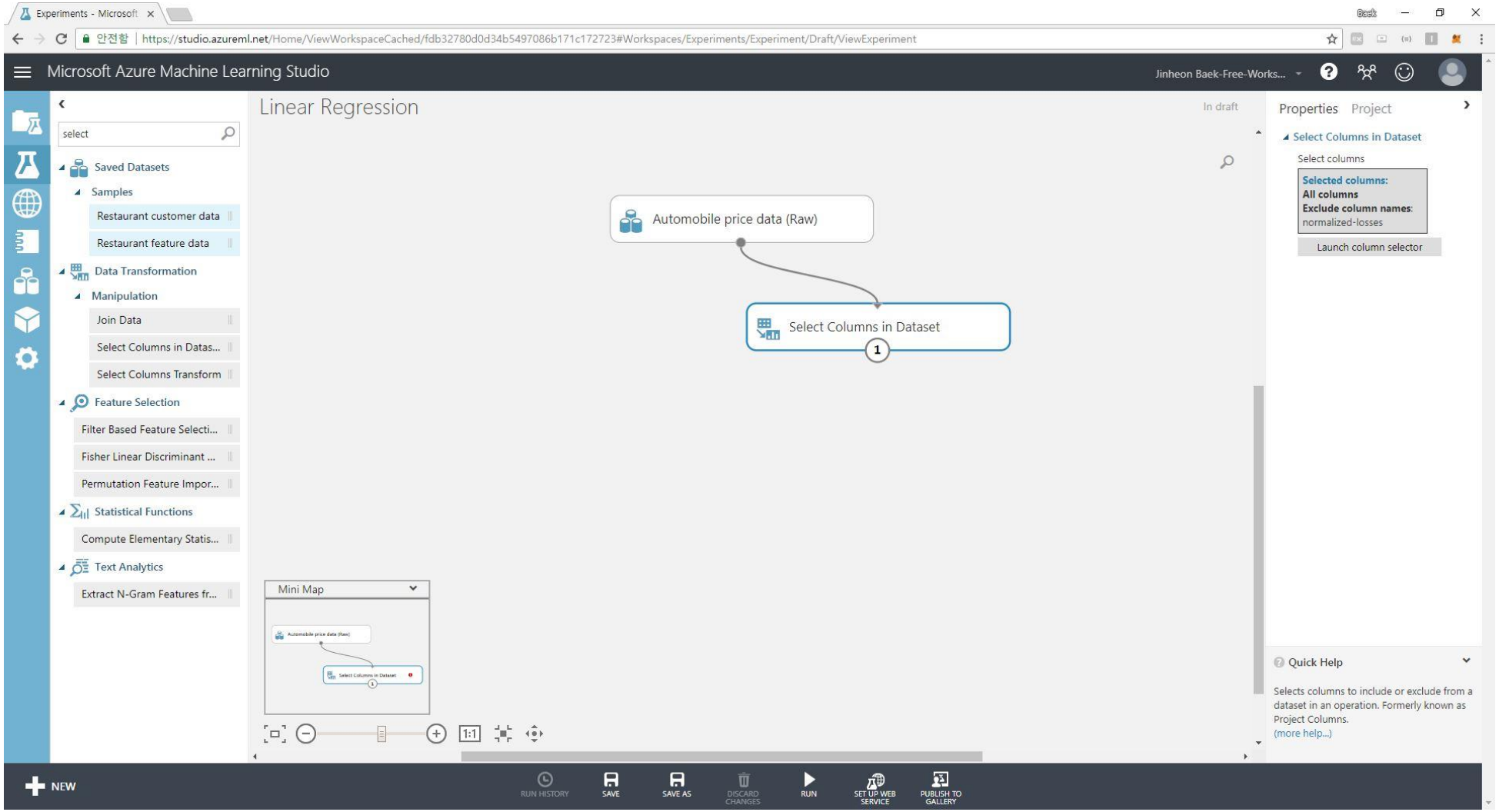


Supervised Learning

Select Columns in Dataset 클릭 (오른쪽 메뉴 확인)

Launch column selector 클릭

With rules -> All columns -> Exclude -> normalized-losses (column names)

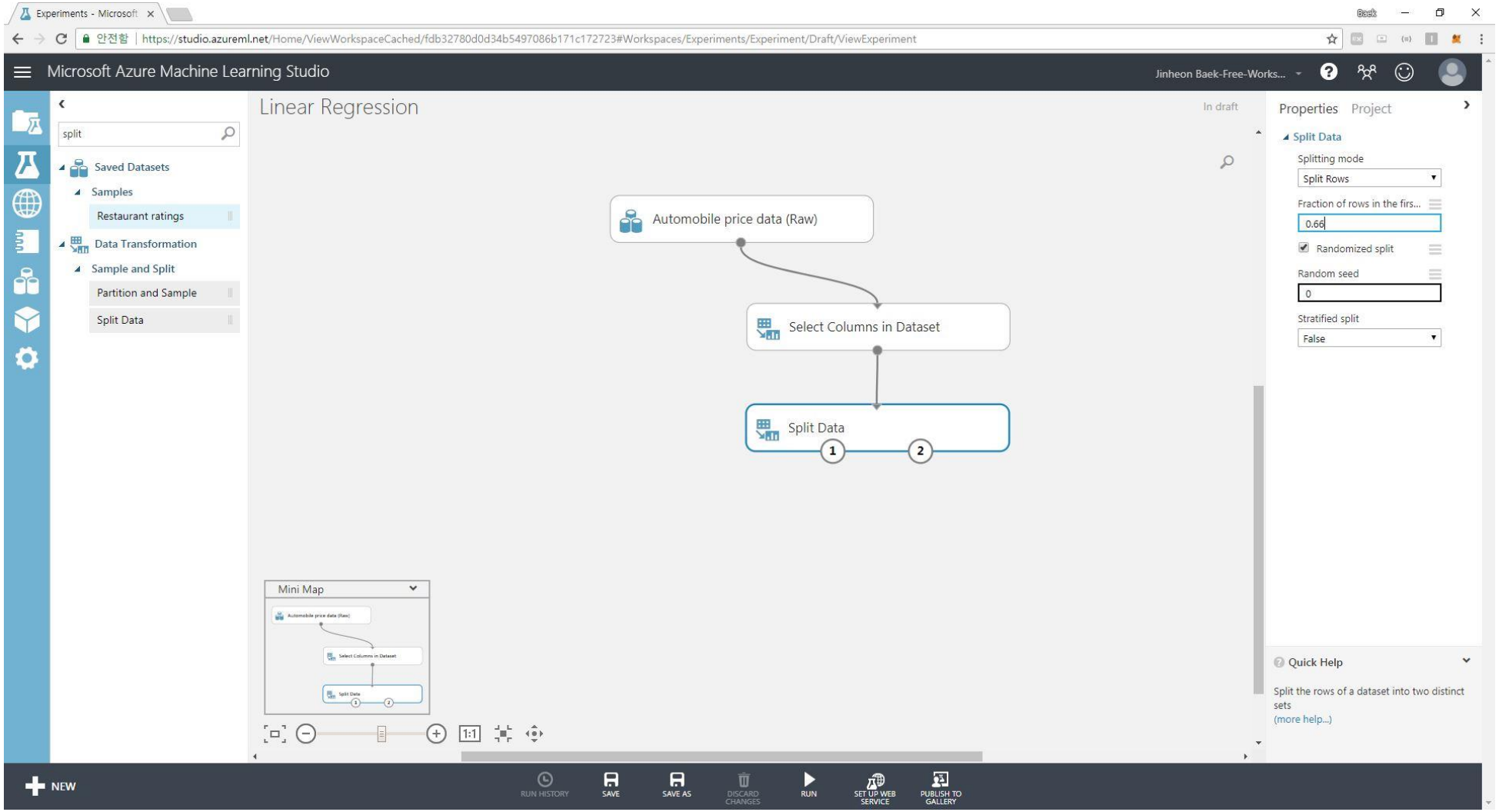


Supervised Learning

검색창: Split Data

Select Columns in Dataset -> Split Data (연결)

Split Data 오른쪽 메뉴 -> Fraction of rows ... (0.66 설정)

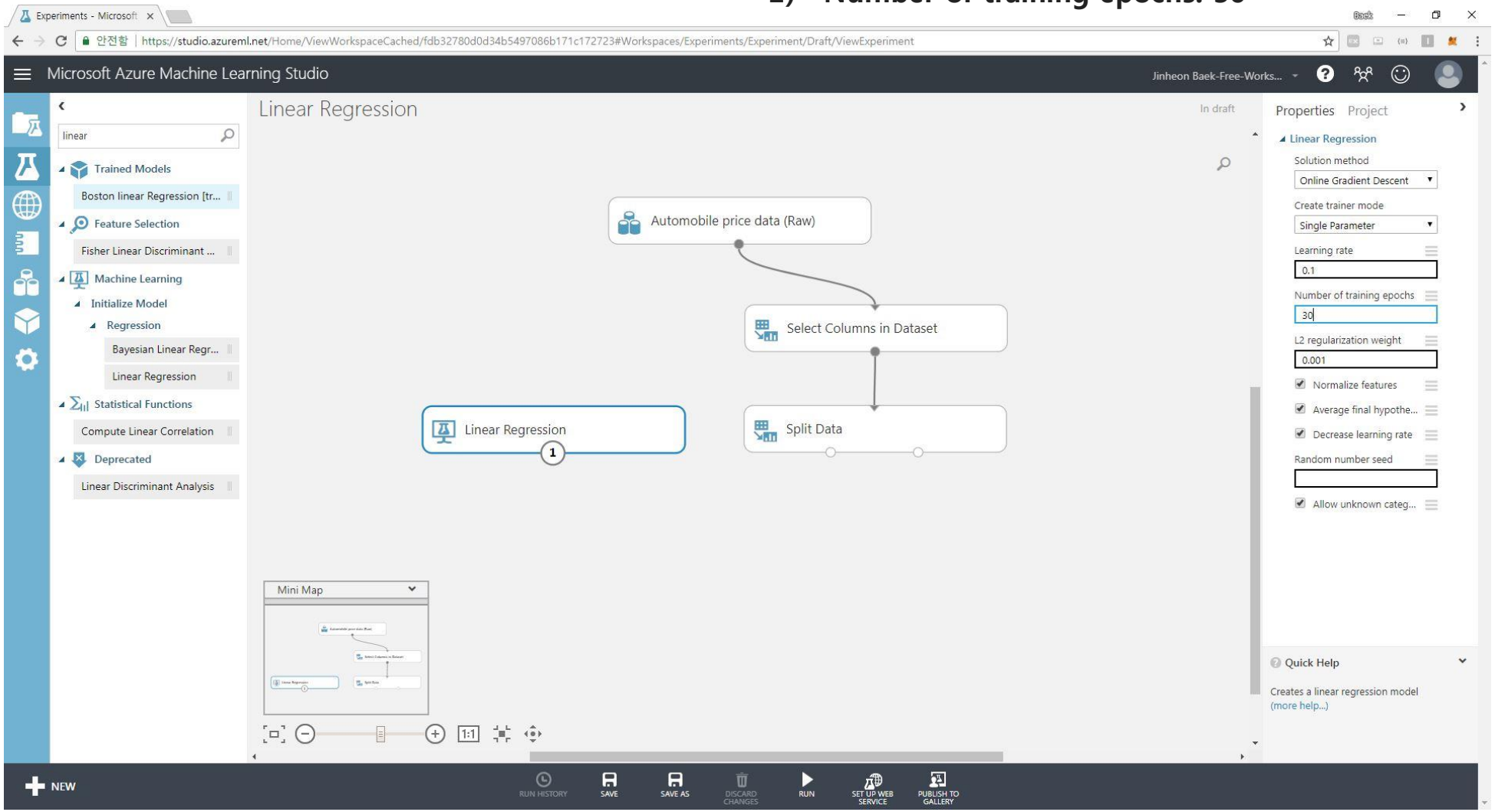


Supervised Learning

검색창: Linear Regression

Linear Regression 오른쪽 메뉴 ->

1) **Solution Method: Online Gradient Descent**
2) **Number of training epochs: 30**

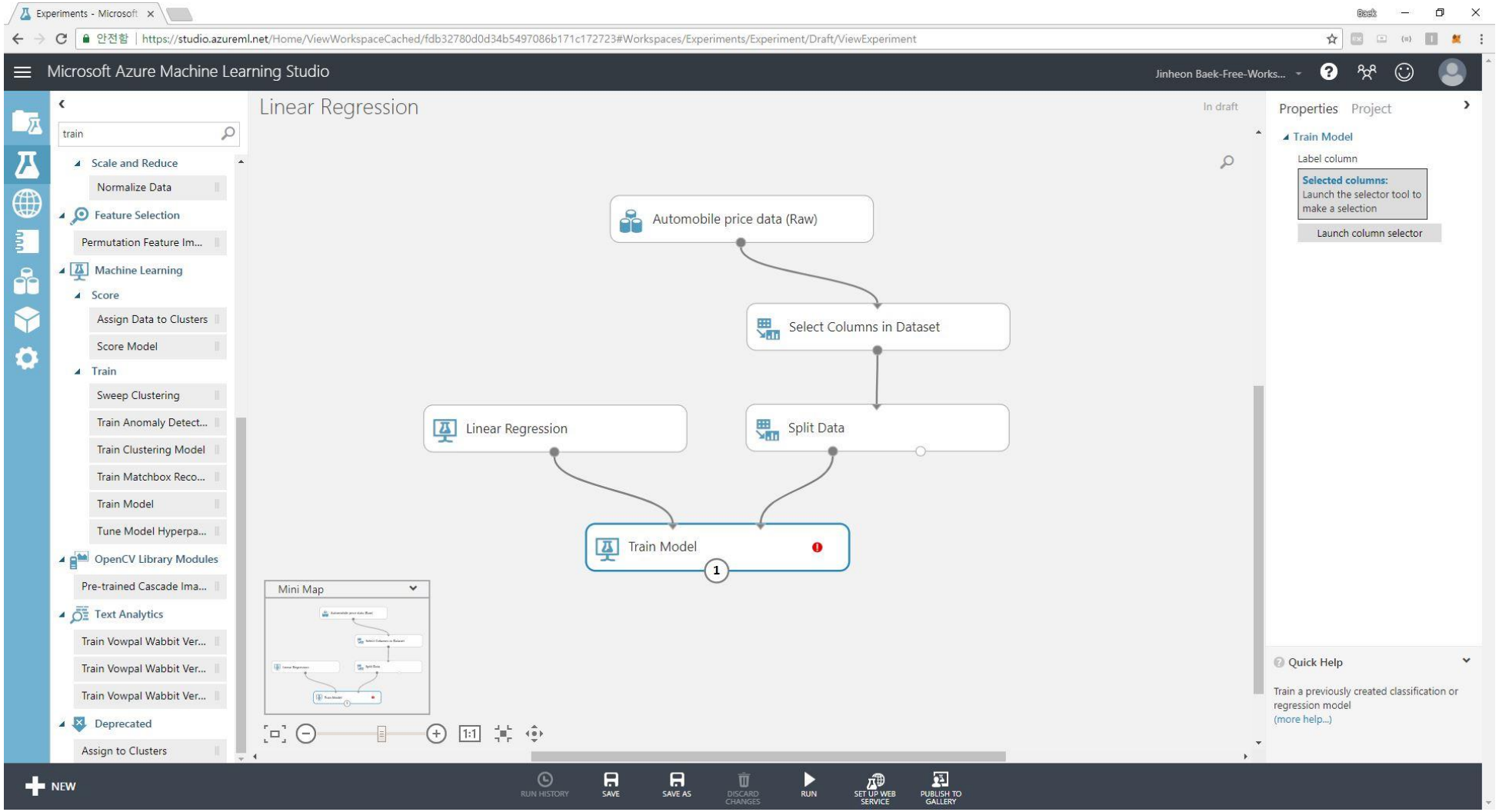


Supervised Learning

검색창: Train Model

Linear Regression -> Train Model (왼쪽 연결)

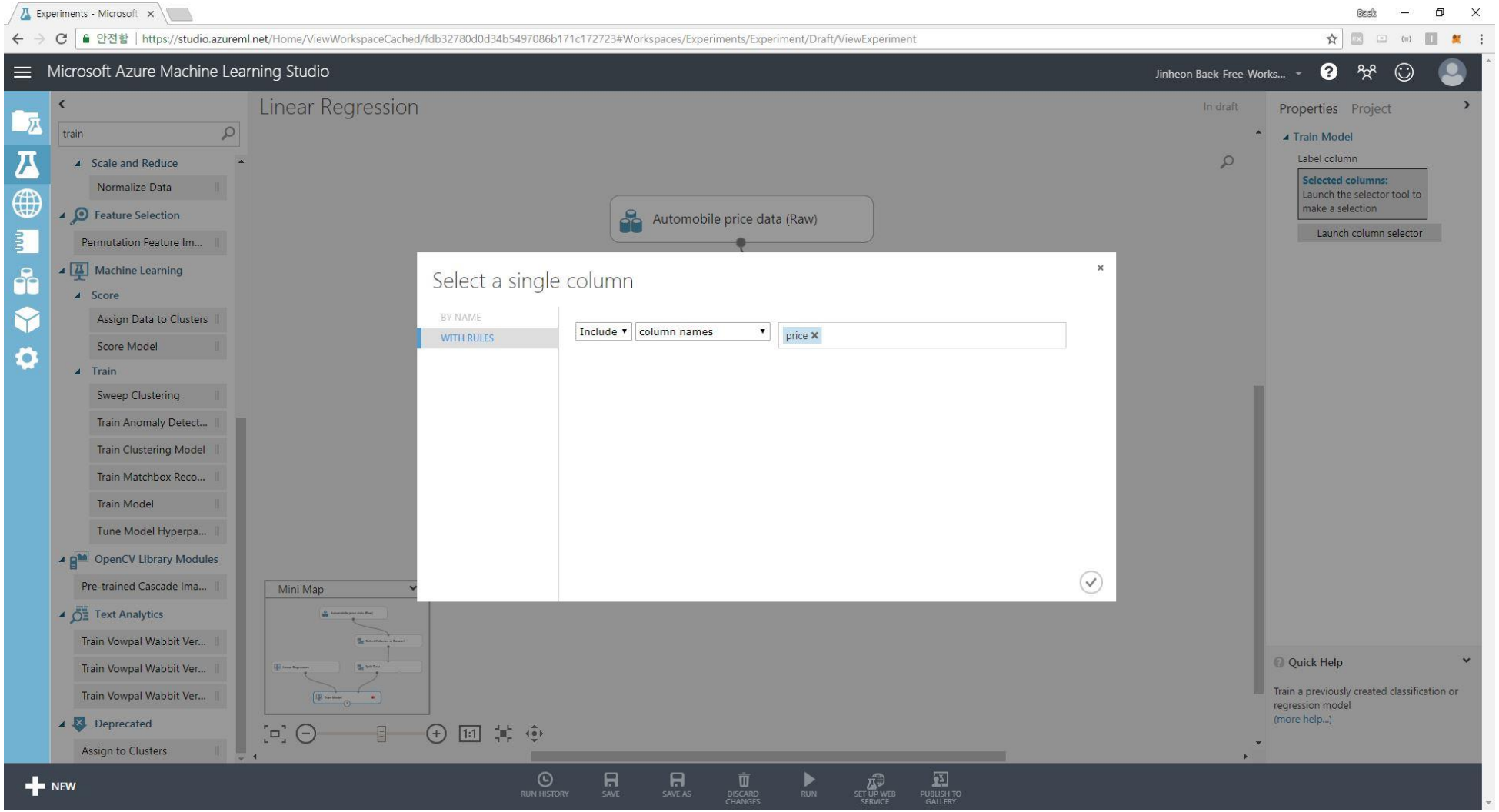
Split Data -> Train Model (오른쪽 연결)



Supervised Learning

Train Model 오른쪽 메뉴 -> Launch column selector

With Rules -> Include -> column names -> price

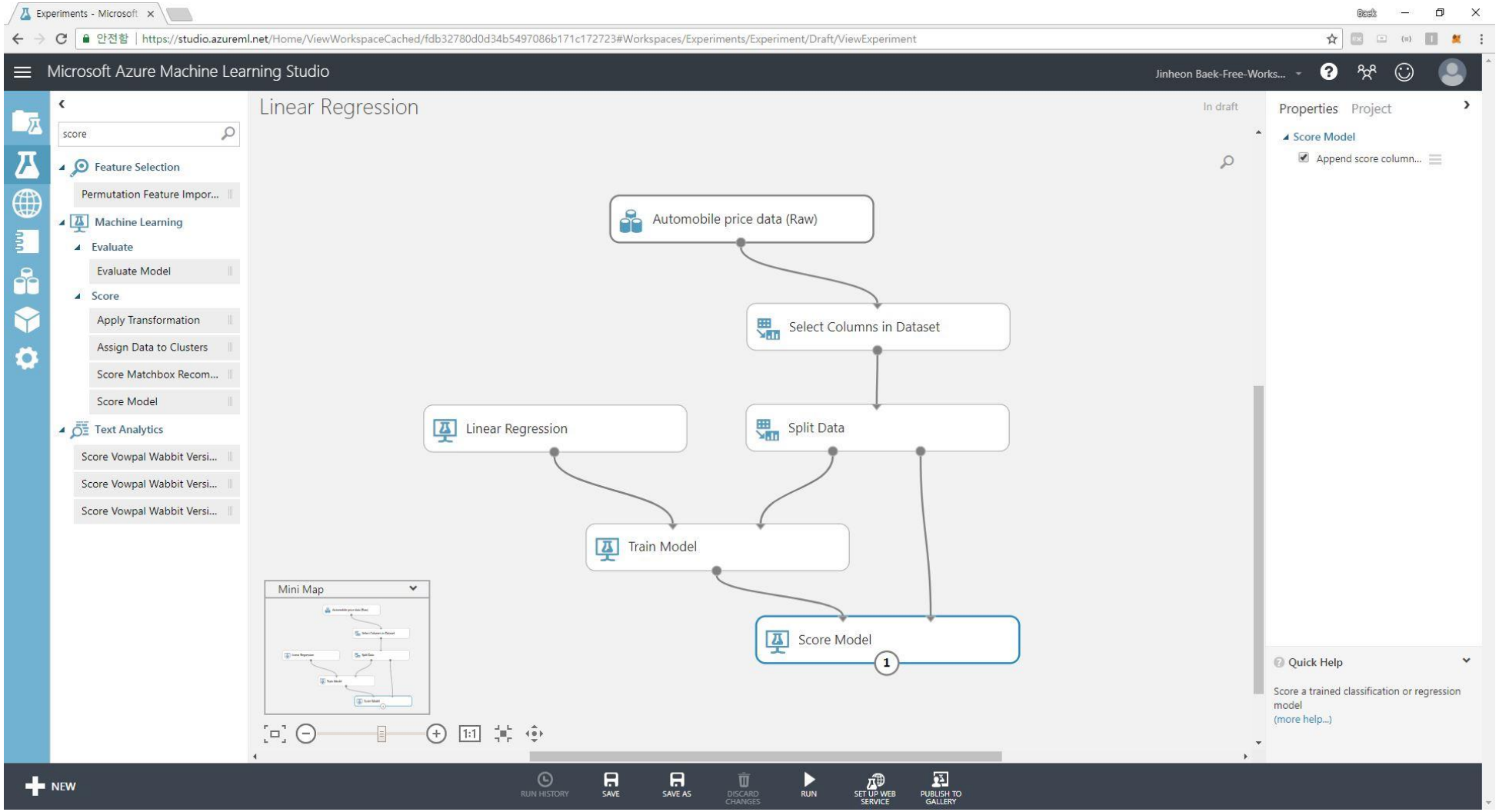


Supervised Learning

검색창: Score Model

Train Model -> Score Model (왼쪽 연결)

Split Data -> Score Model (오른쪽 연결)

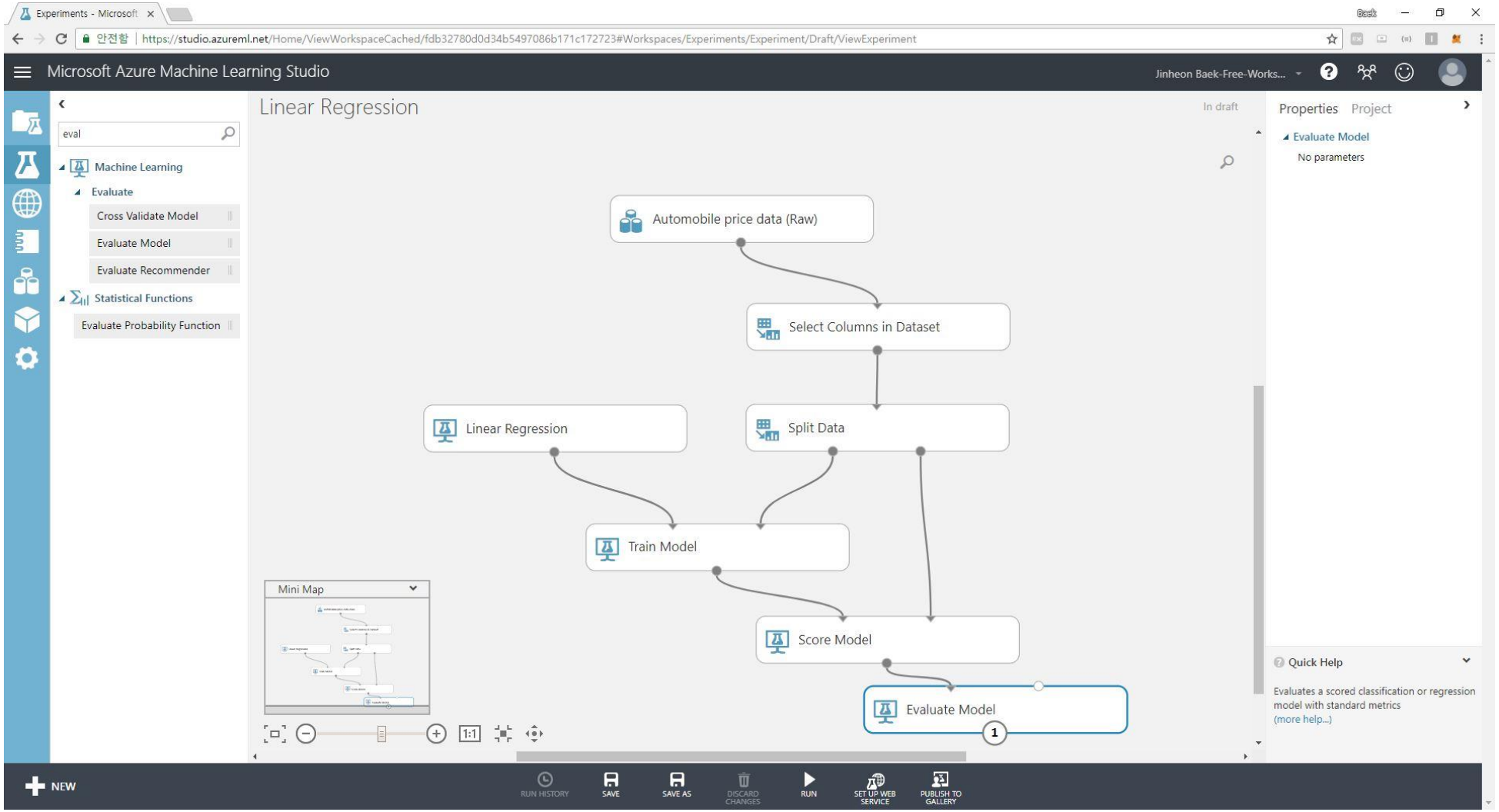


Supervised Learning

검색창: Evaluate Model

Score Model -> Evaluate Model (왼쪽 연결)

하단 Run 실행 후 결과 보기 !

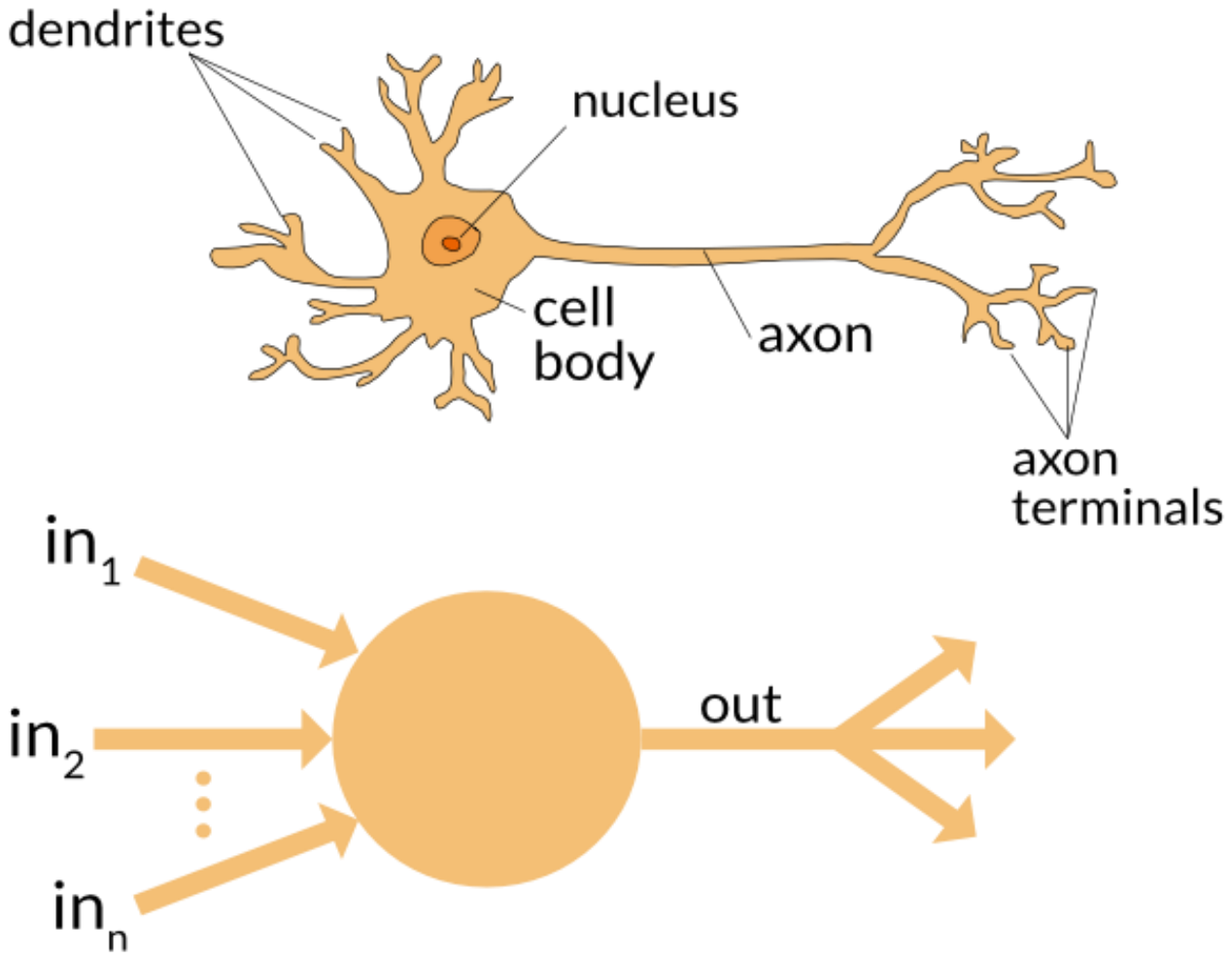


Supervised Learning

[Deep Learning]

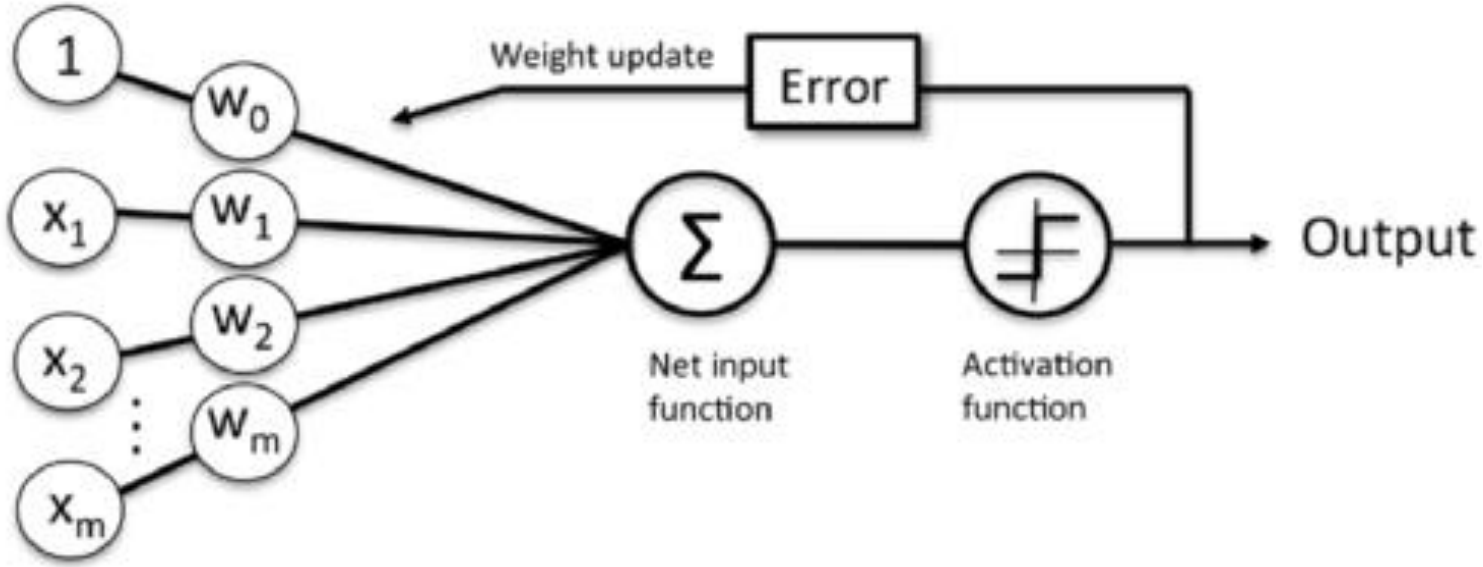
✓Neural Network
: Simple MLP (Multi-Layer Perceptron)

Supervised Learning



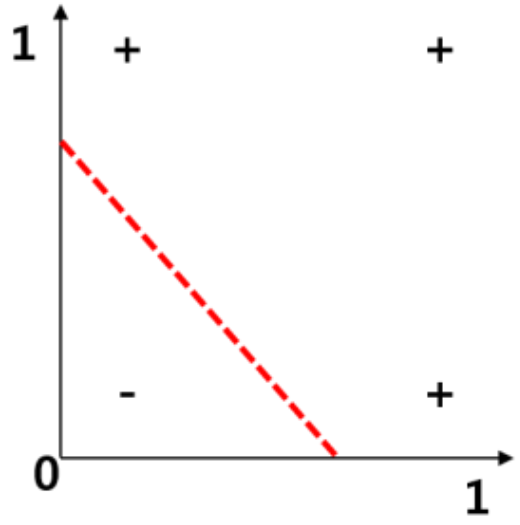
Supervised Learning

Single Layer Neural Network



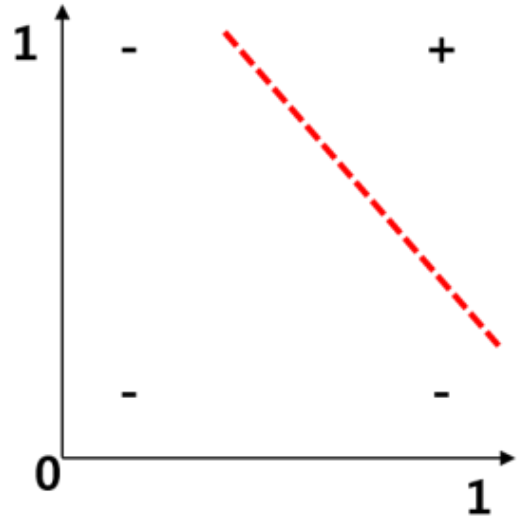
Supervised Learning

OR



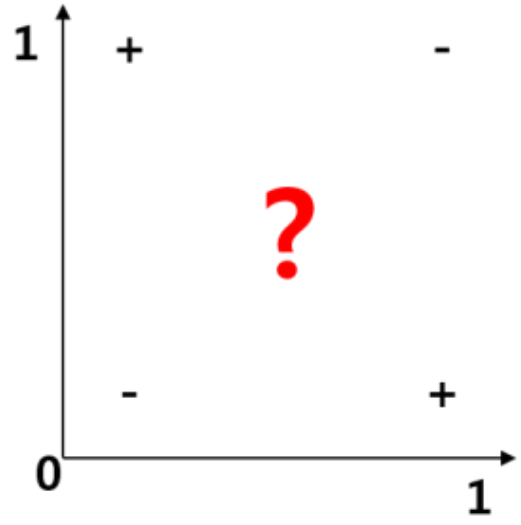
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

AND



x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

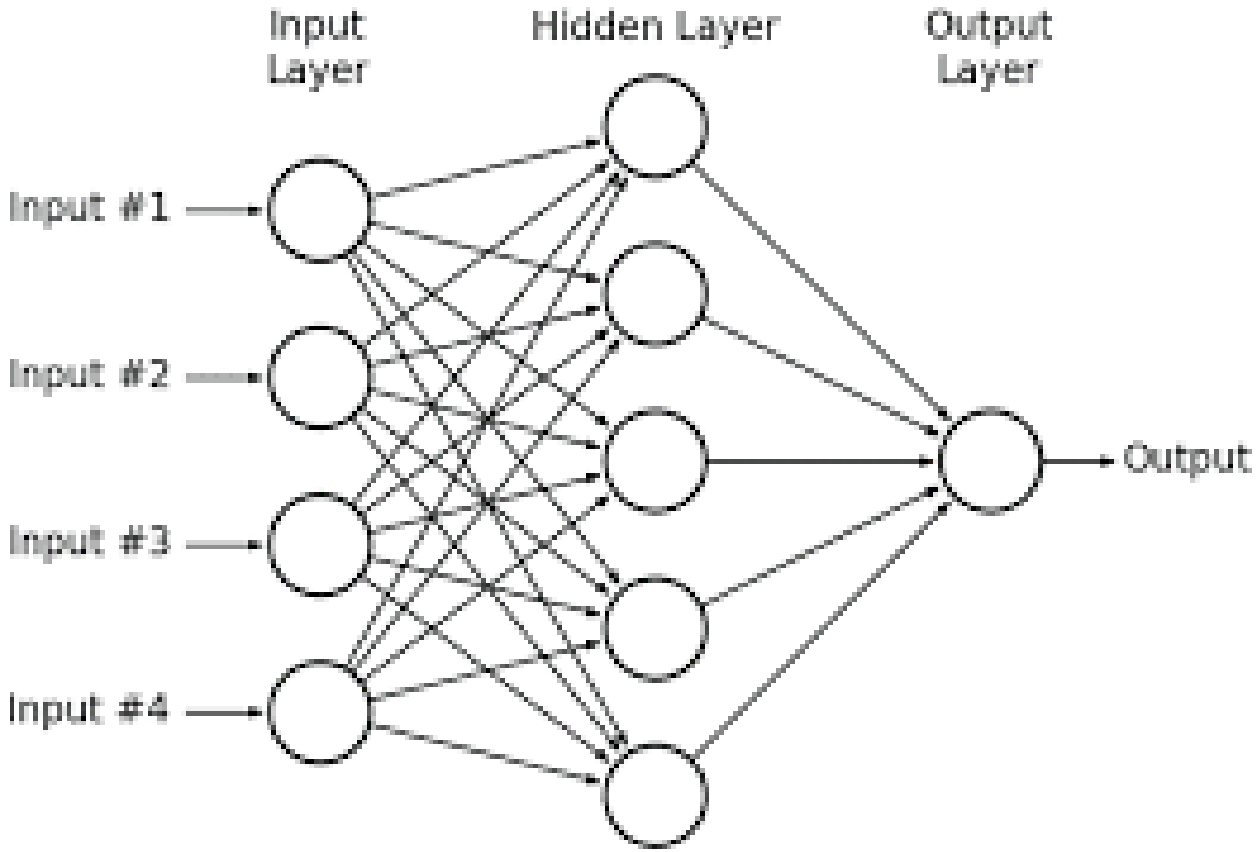
XOR



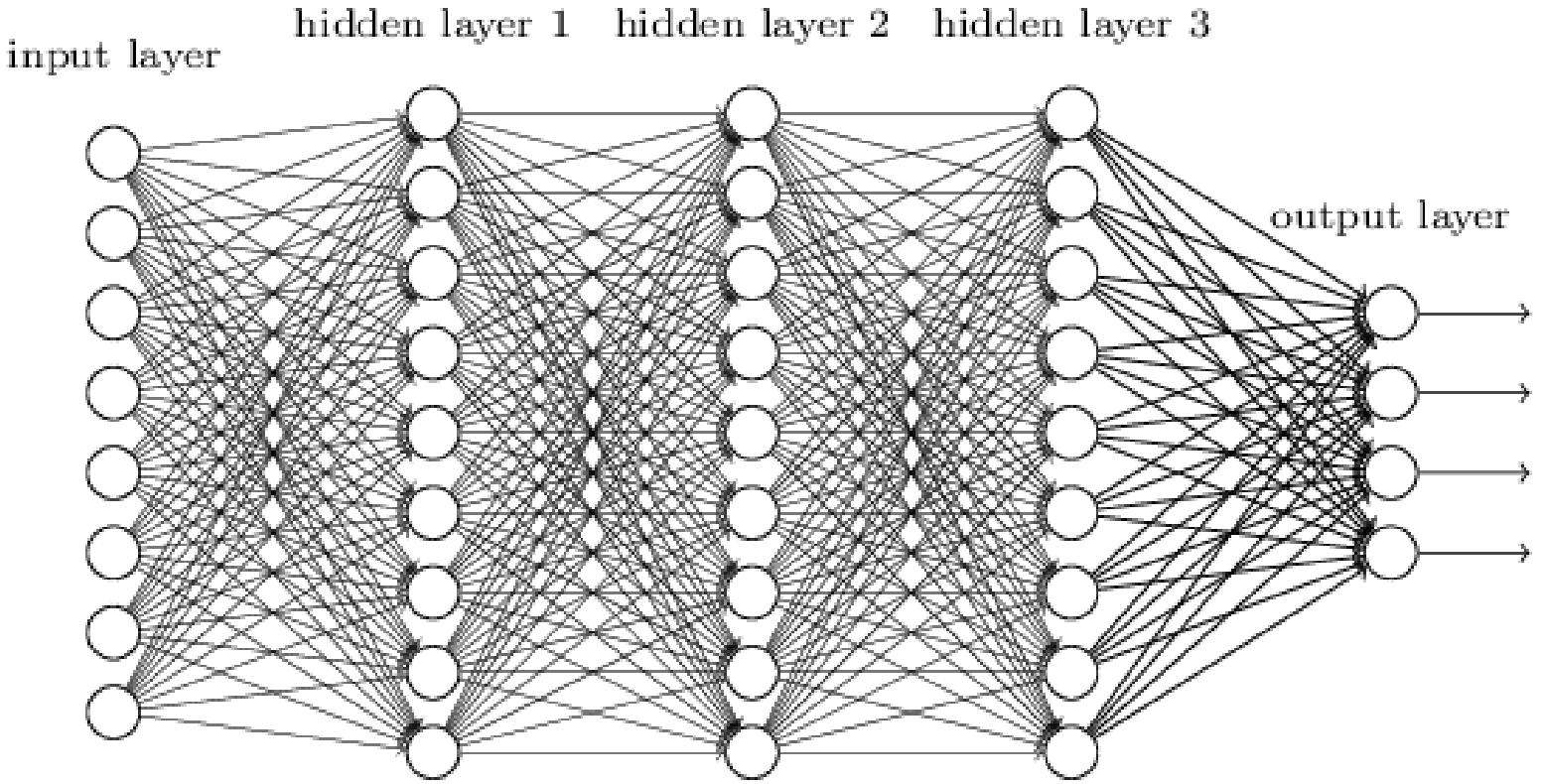
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Supervised Learning

Multi-Layer Neural Network

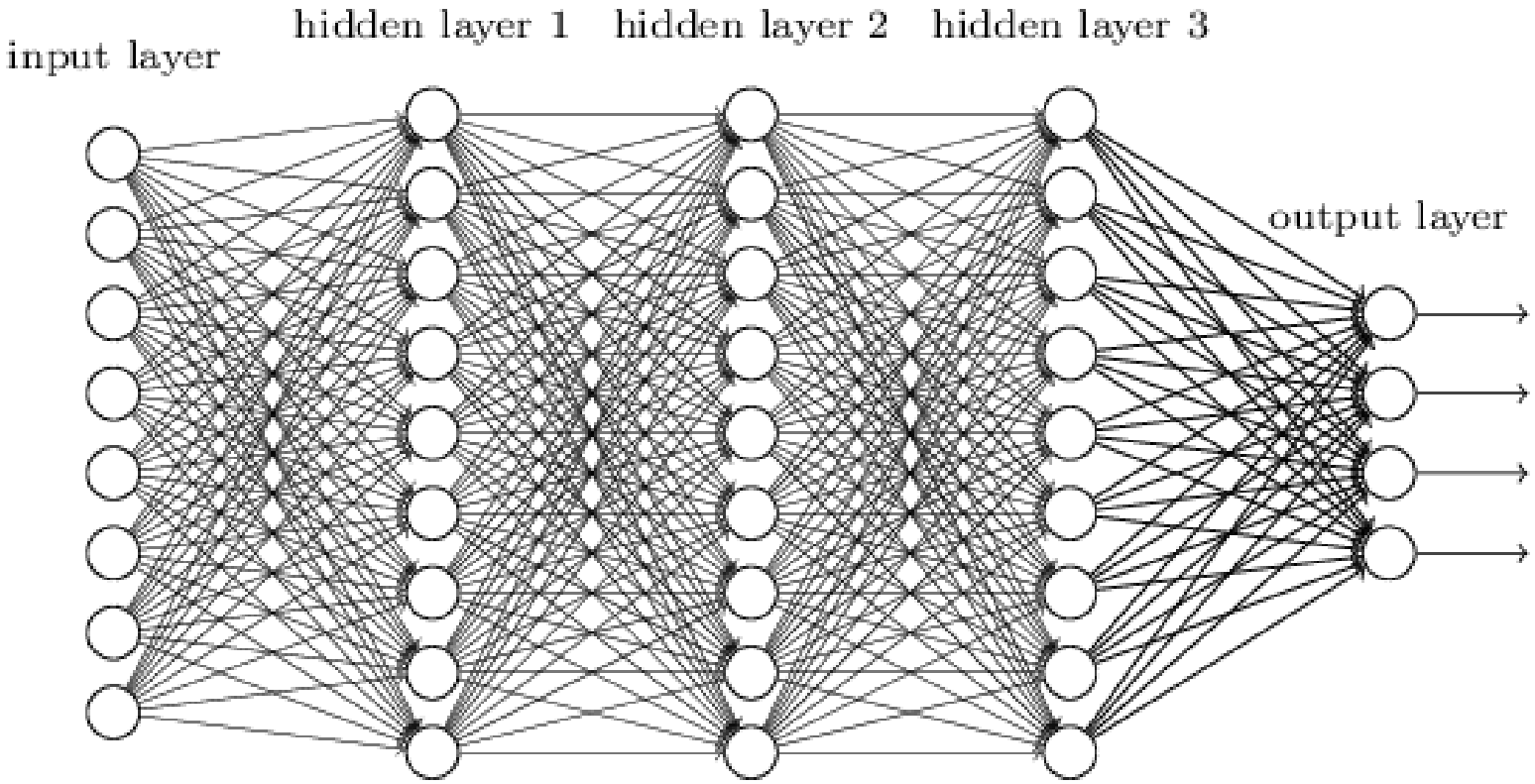


Neural-network



결과 예측

Neural-network



Layer 값 조정

Neural Network

✓HOL

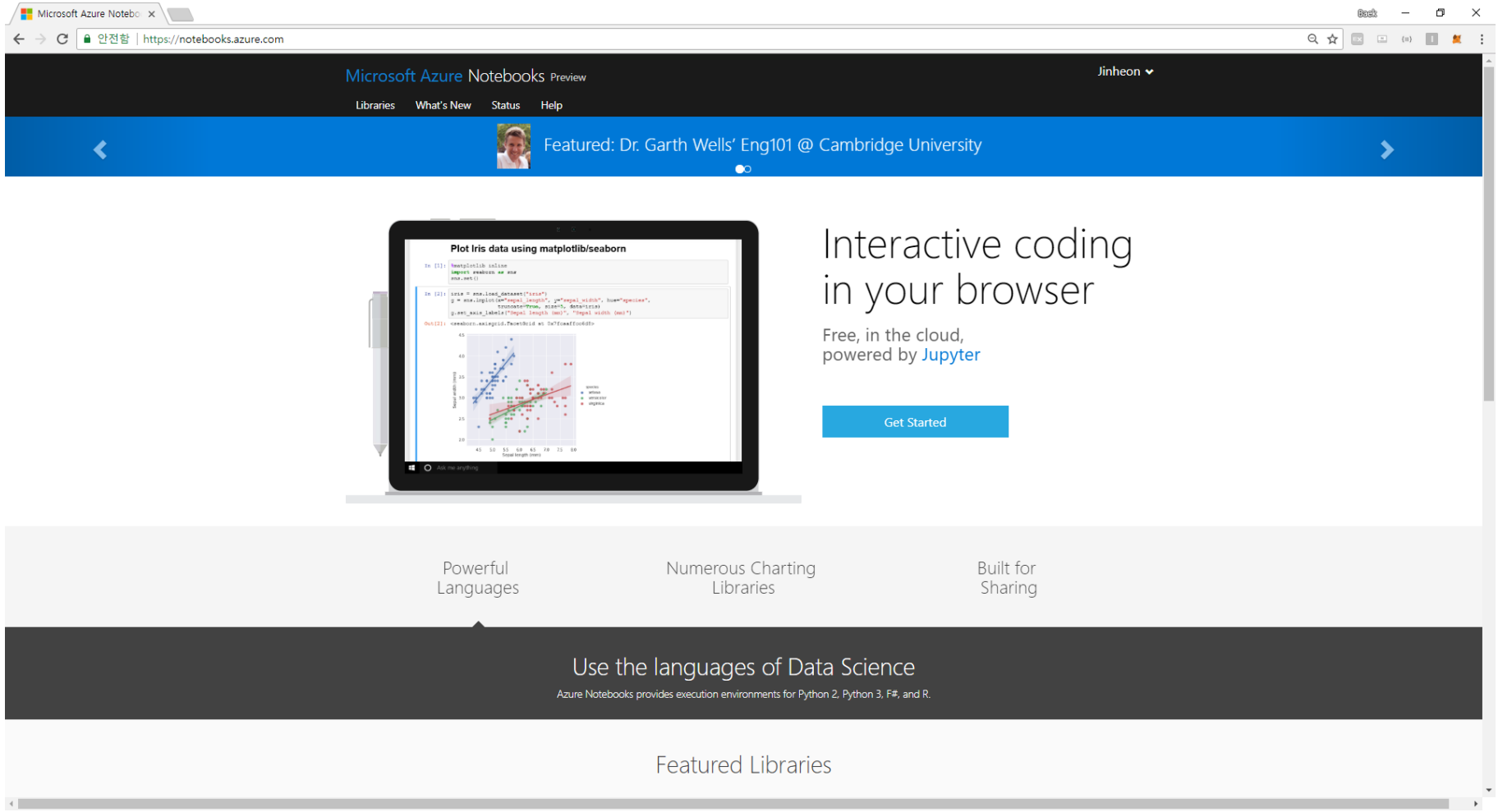
“MNIST Data Set ?”

Neural Network

28 x 28, 필기체 숫자 (0 ~ 9)

Neural Network

✓Azure Notebook (https://notebooks.azure.com)



Neural Network

✓Azure Notebook (https://notebooks.azure.com)

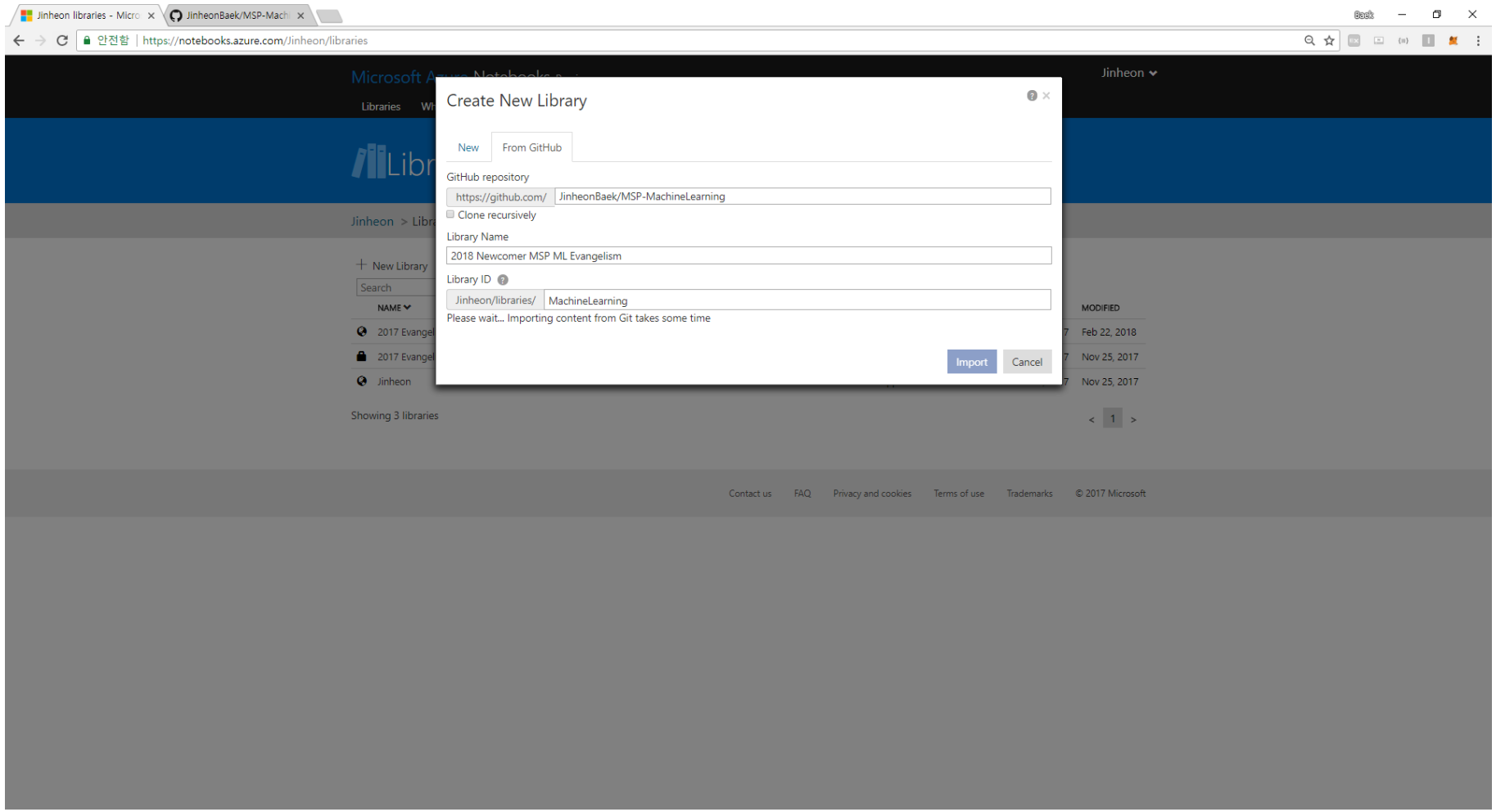
1. Logic
2. Libraries 메뉴
3. New Library

Neural Network

GitHub repository: JinheonBaek/Coding-edu-volunteer-ML-seminarLibrary Name (예시): 20180712 ML Evangelism

Library ID (예시): MachineLearning

✓Azure Notebook (https://notebooks.azure.com)



Neural Network

✓HOL – Challenge

“Testing score is above 0.98”

Neural Network

✓HOL – Challenge

“Testing score is above 0.98”

Fitting

Machine Learning

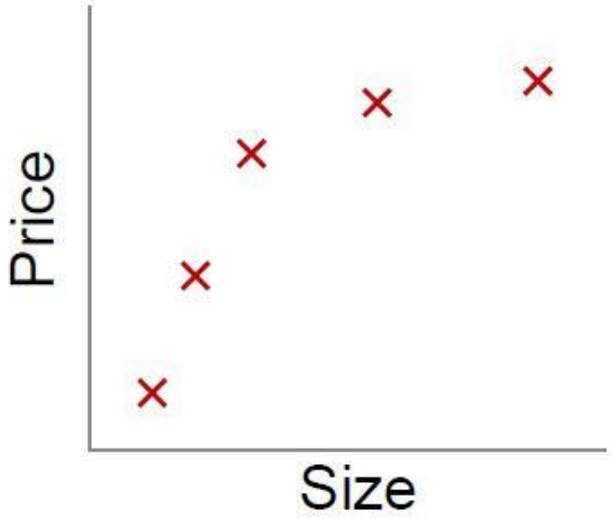
Microsoft Student Partner
백진헌

Fitting

✓**Fitting**

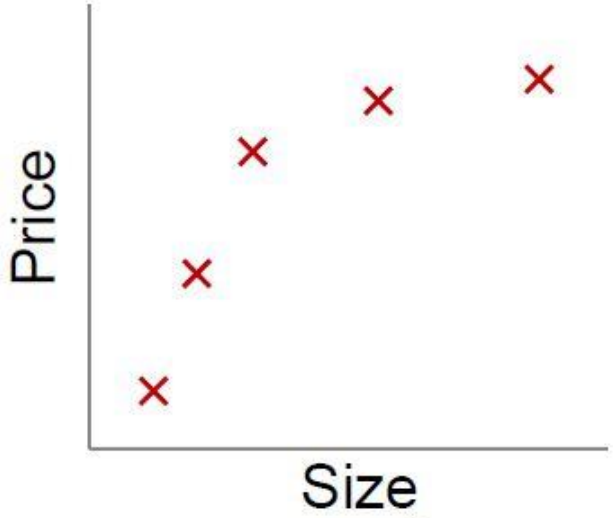
$\theta_0 + \theta_1 x$

High bias
(underfit)



$\theta_0 + \theta_1 x + \theta_2 x^2$

"Just right"

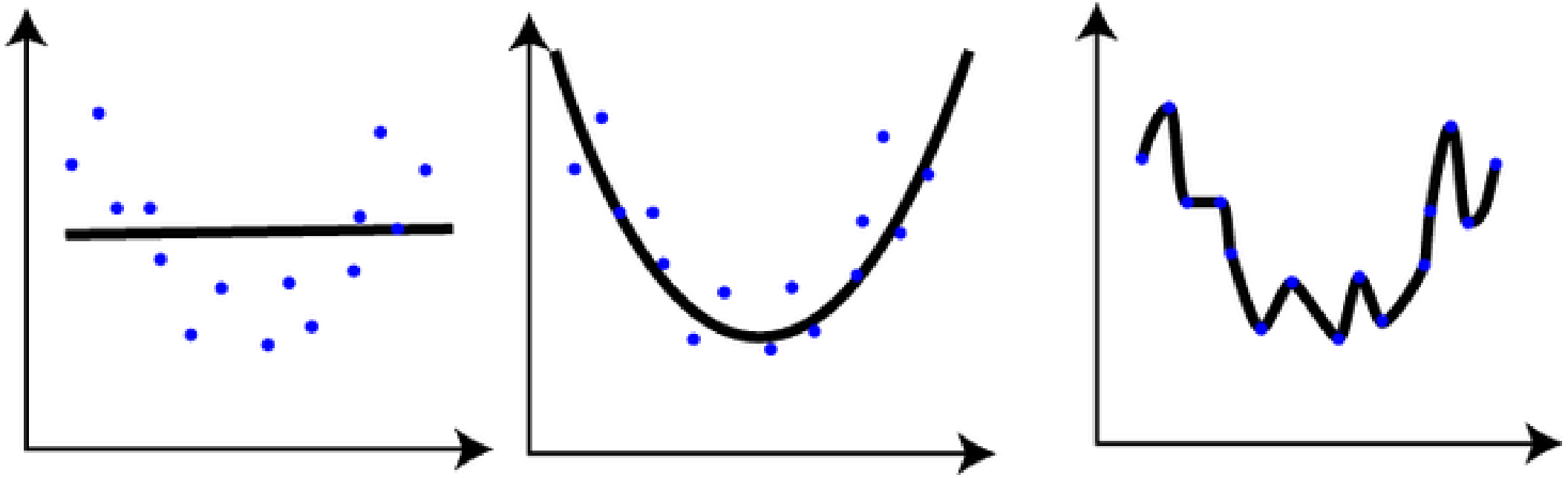


$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High variance
(overfit)

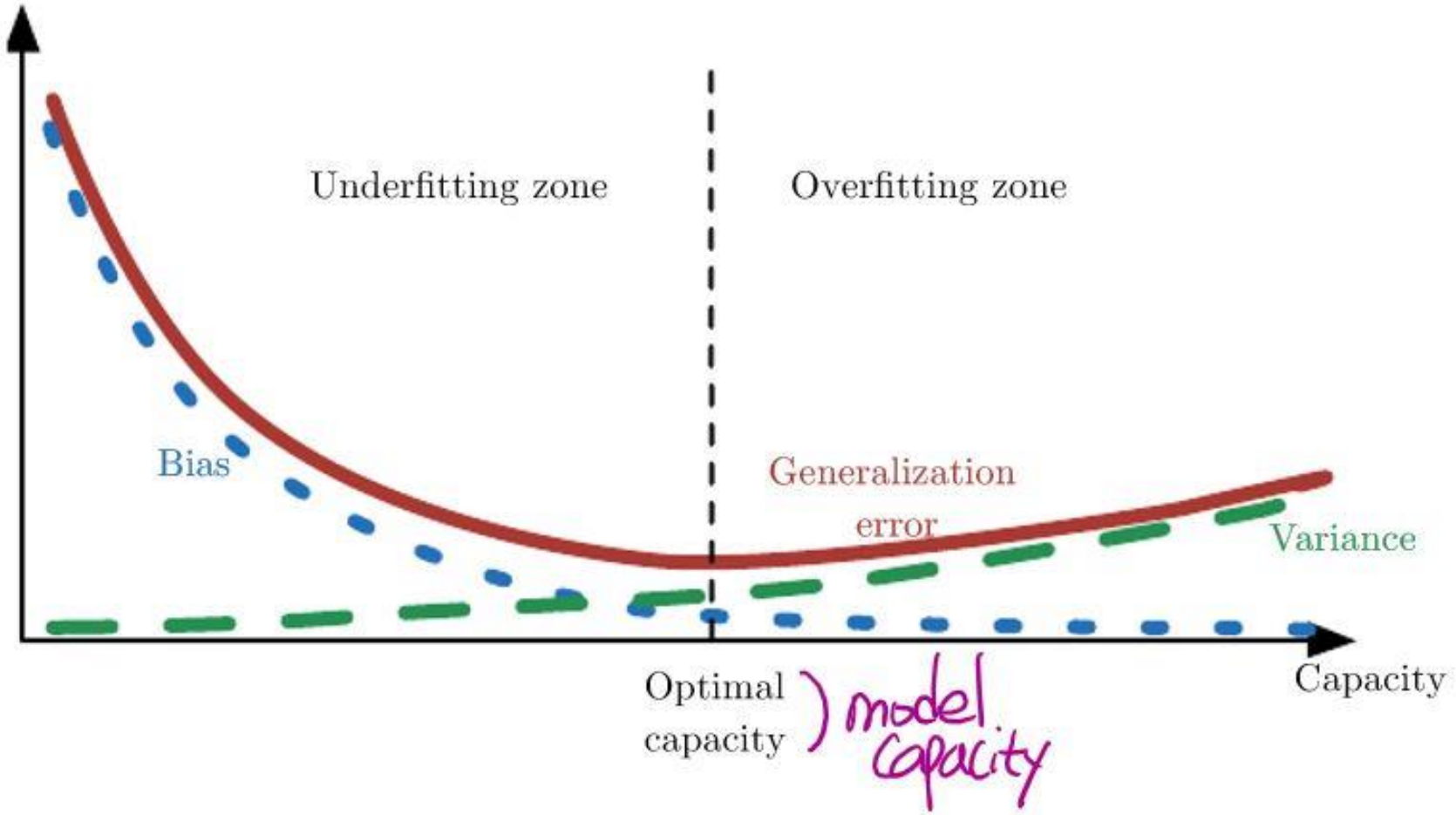
Fitting

✓Fitting



Fitting

✓**Fitting**



Model

Machine Learning

Microsoft Student Partner
백진헌

Models

✓Blackbox vs Descriptive Model

Linear Regression

: Models are descriptive, because one can see which variables are weighed heaviest.

Neural Network

: Models are generally opaque.

Models

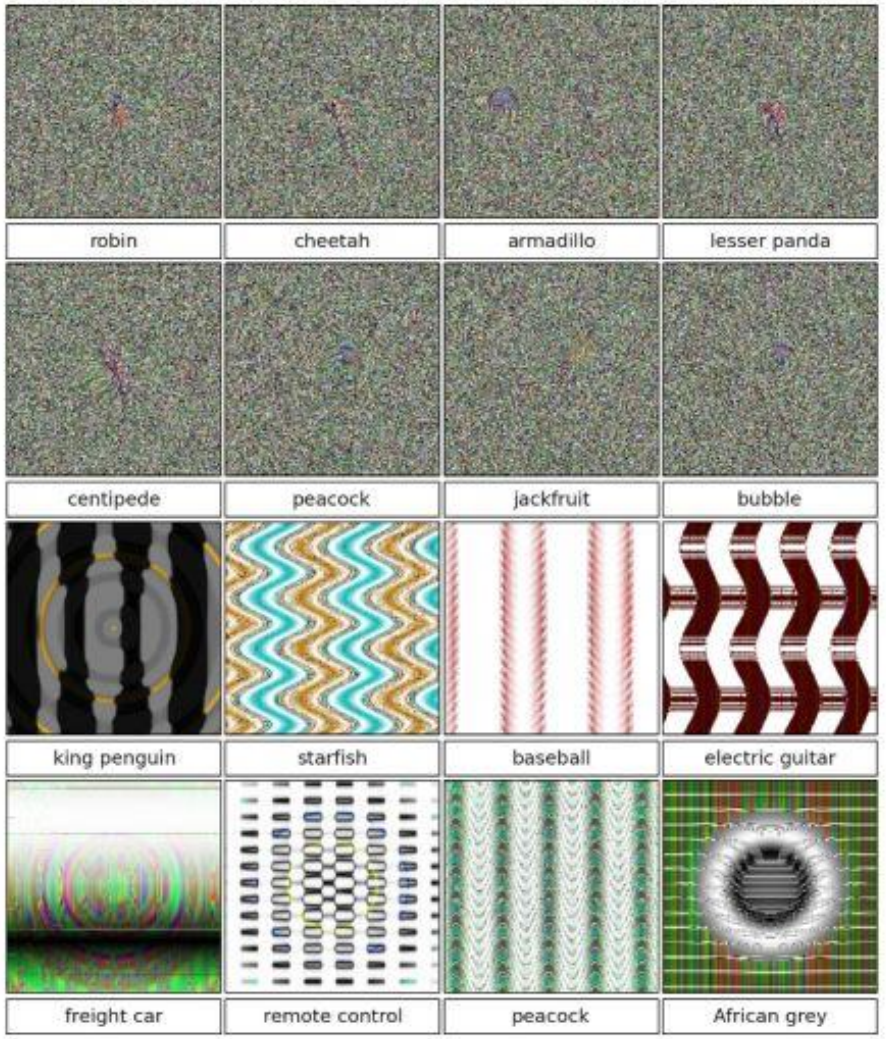
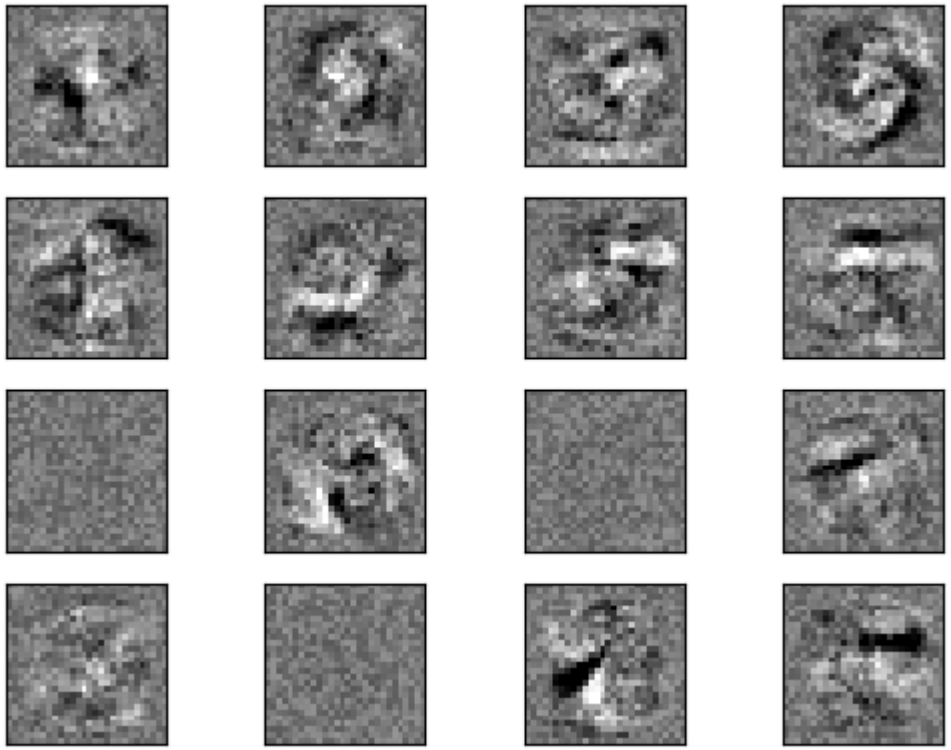


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

Models



Models

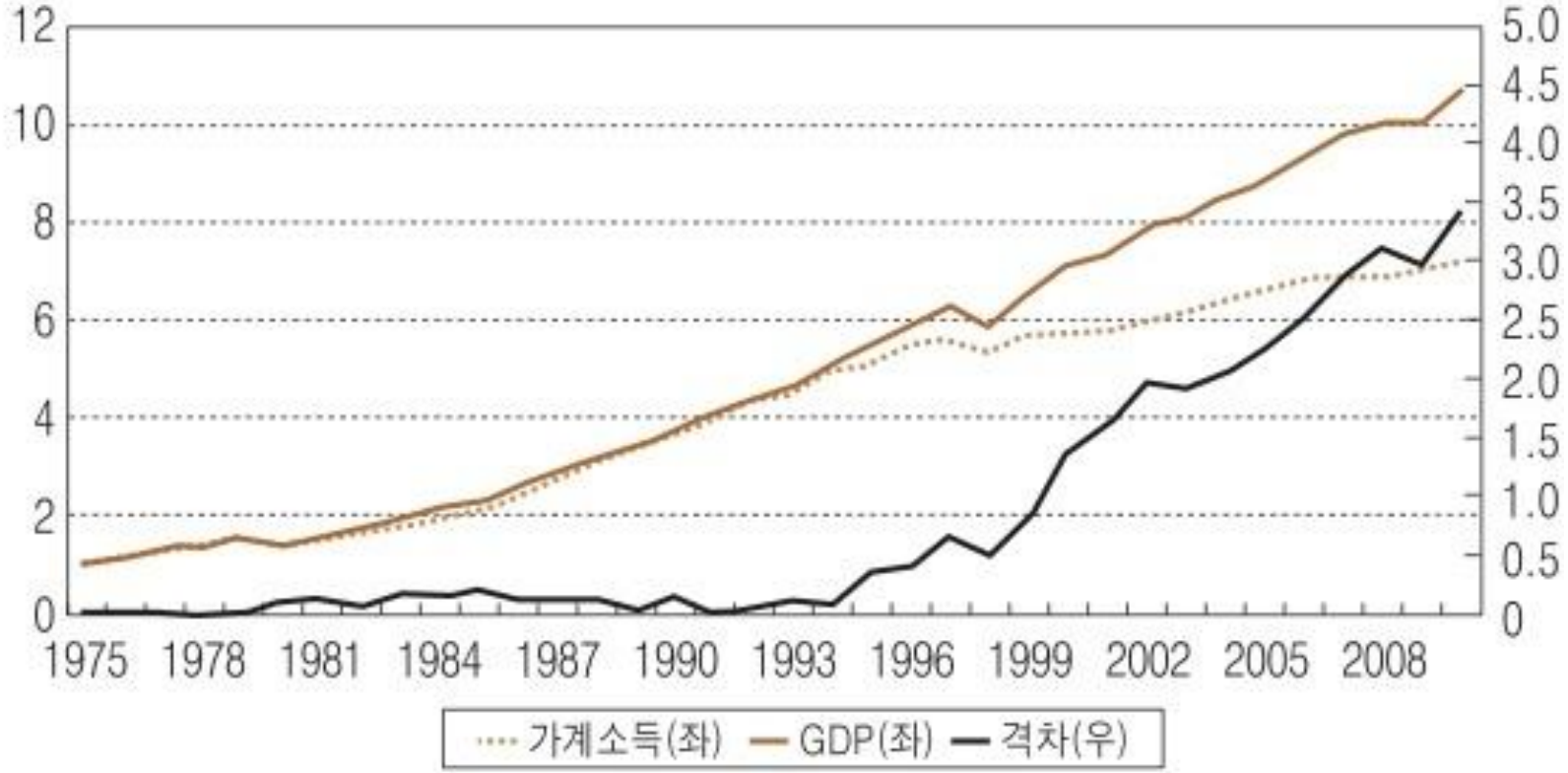
✓First-principle vs Data-driven

First-principle
: Models based on a theoretical explanation.

Data-driven
: Models based on observed data correlations.

Warm Up

〈그림 4-2〉 GDP와 가계소득의 증가 추이 비교



주 : (실질)GDP와 (실질)가계소득은 모두 1975=1로 지수화.

References

- ✓Deep Learning Book – MIT
- ✓Machine Learning – Coursera
- ✓Korea Univ. Cose432 (Data Science)

Q & A

Thanks