

# Intro

## Machine Learning

Microsoft Student Partner

백진헌

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 8<sup>th</sup> 연수생 (2017.08 ~ 2017.12)

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

### ✓ Machine Learning?

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

### ✓ 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)**



### ✓ Machine Learning?

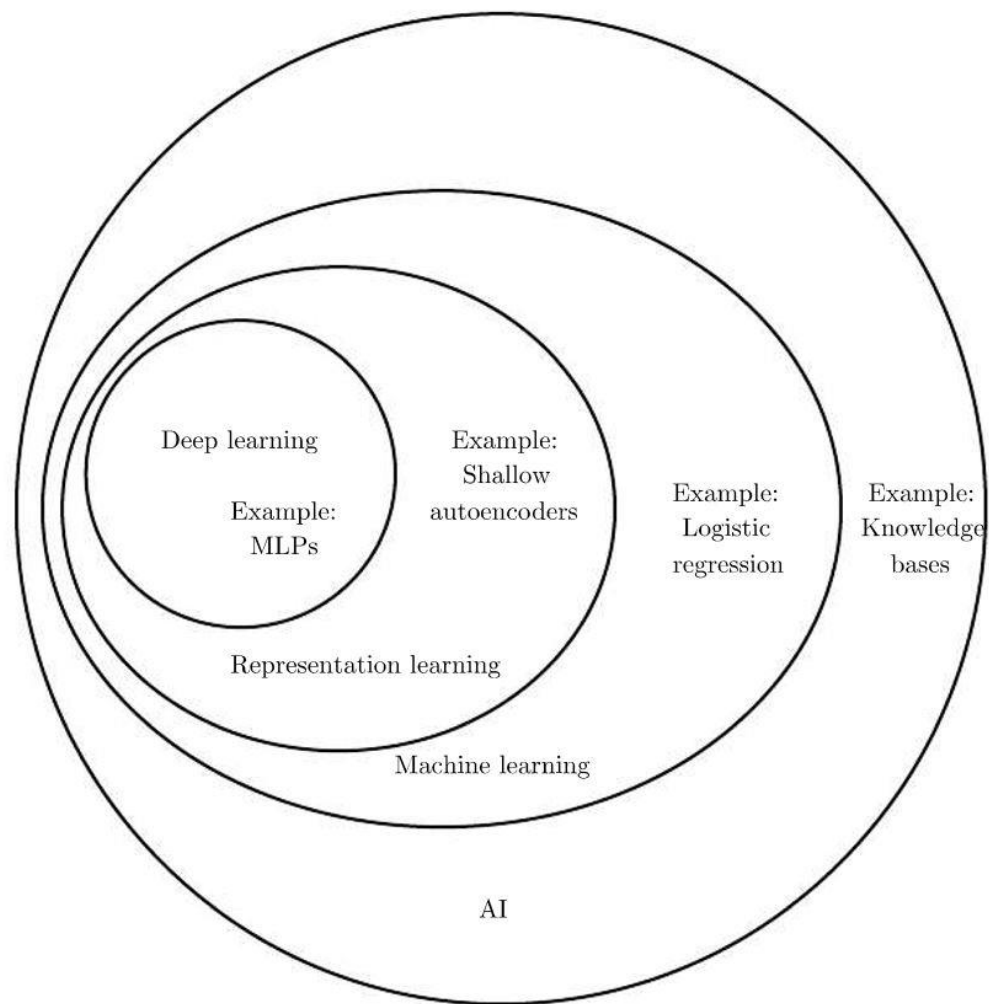
Ex) Spam mail filter,

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



✓ 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

### ✓ Machine Learning

“가공되지 않은 데이터로부터 패턴을 찾는 것”

장점: 컴퓨터가 스스로 데이터로부터 패턴을 찾아 나간다.

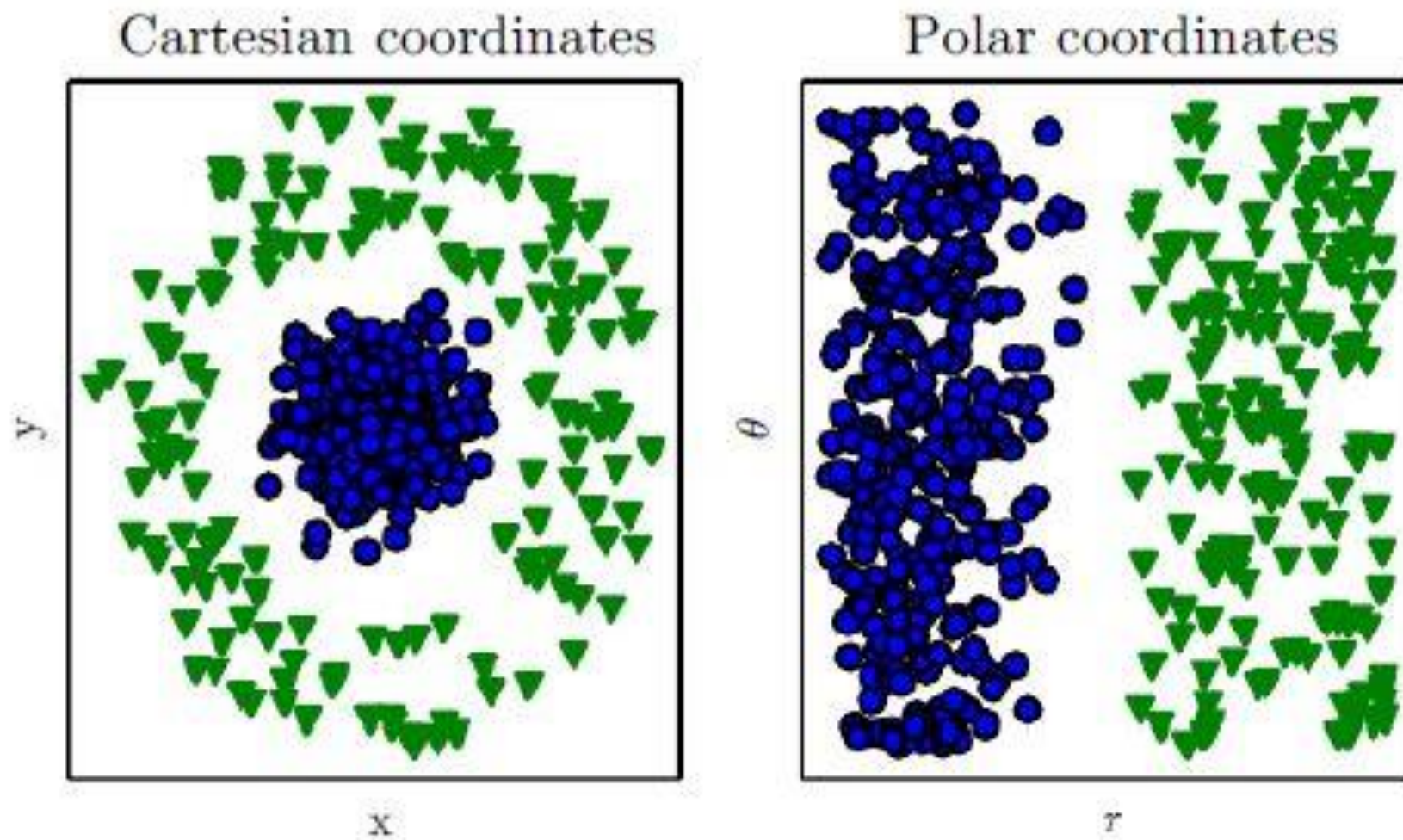
단점: 컴퓨터에게 주어진 데이터에 과도하게 의존적이다.

## Concept of Machine Learning

### ✓ Machine Learning

“Difficult to know what **features** should be extracted.”

# Concept of Machine Learning



### ✓ Representation Learning

Important features 를 추출하는 것이 매우 어려우므로,  
컴퓨터만 이해하는 공간을 스스로 만들어,  
컴퓨터만 이해하는 공간 안에서 학습을 진행



### ✓ Representation Learning – Autoencoder

**Autoencoder: encoder + decoder**

**Encoder:** input data into a different representation

**Decode:** new representation back into the original format

### ✓ Representation Learning

데이터로부터,

**high-level, abstract features representation**

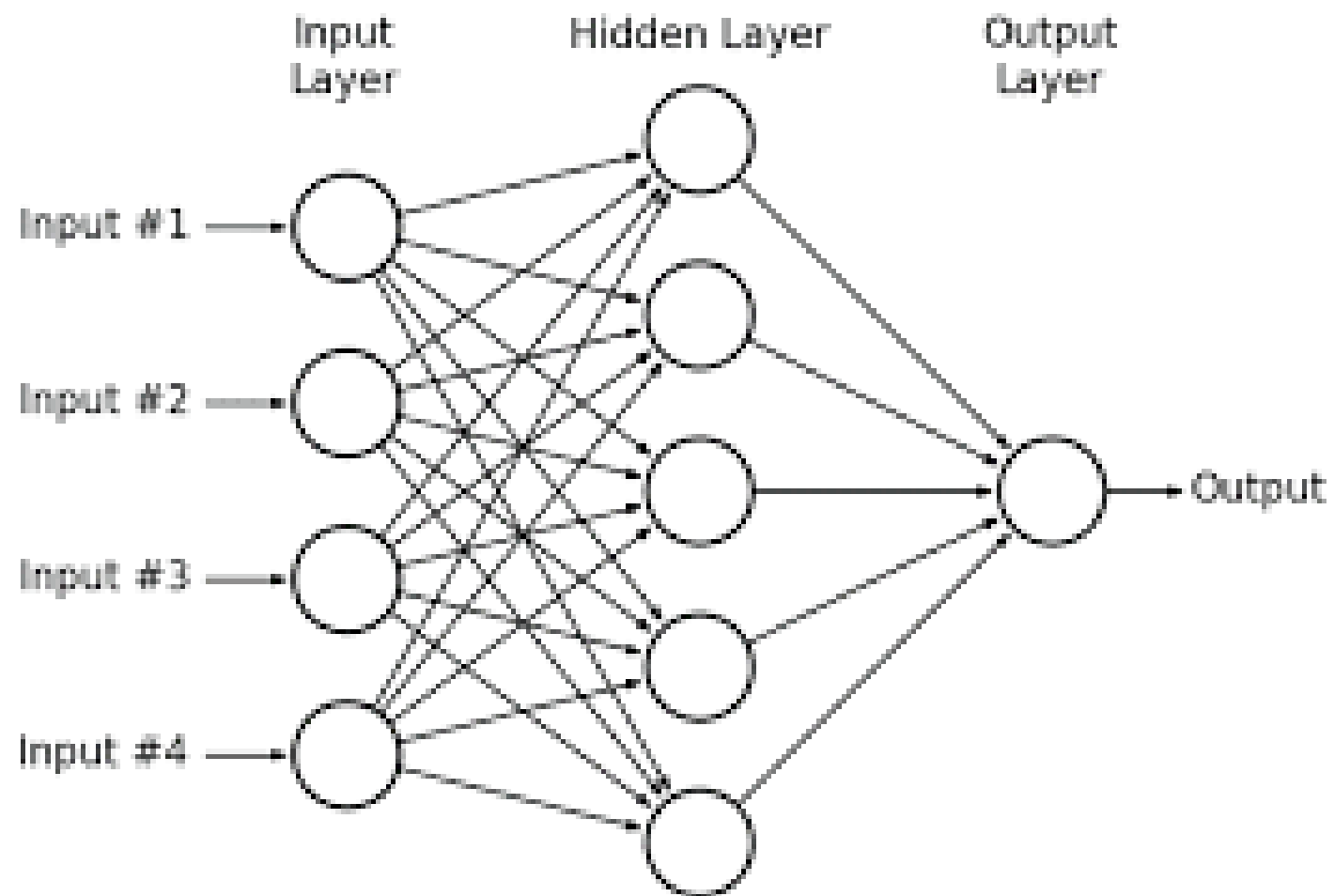
추출이 어려움

### ✓ Deep Learning

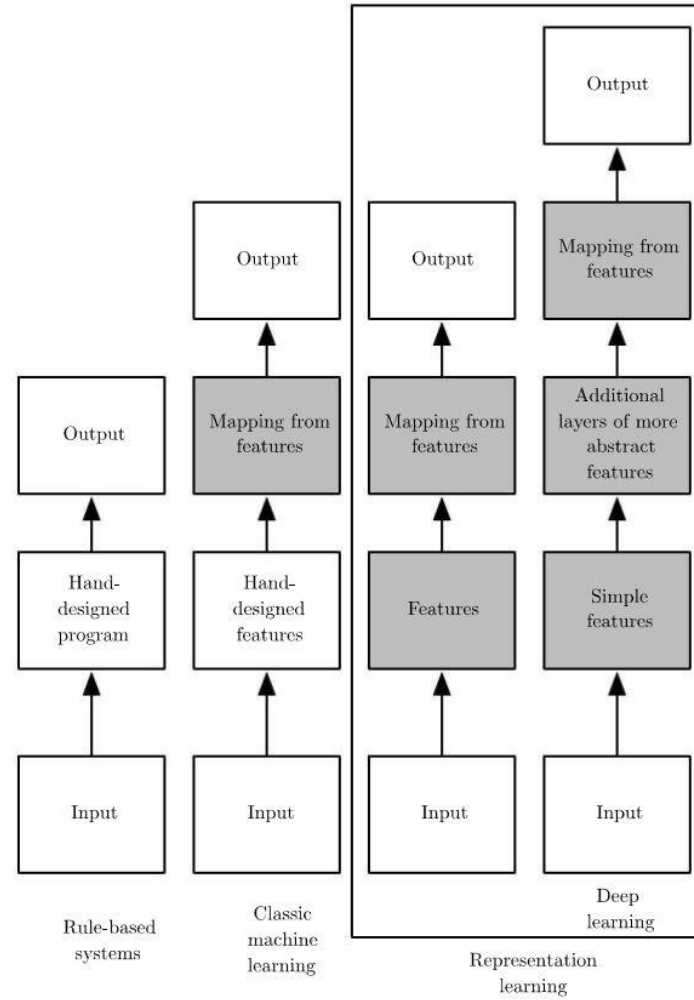
**“복잡한 모델 표현을 아주 간단한 모델 표현을 쌓아 만든다.”**

**Build complex concepts out of simpler concepts (representation)**

# Concept of Machine Learning



# Concept of Machine Learning



### ✓ Supervised Learning

- : 학습에 필요한 데이터와 데이터의 정답이 주어져,  
컴퓨터는 **데이터의 정답**을 기준으로 학습을 수행한다.
- : Experience a dataset containing features,  
but each example is also associated with a label or target.

### ✓ 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)

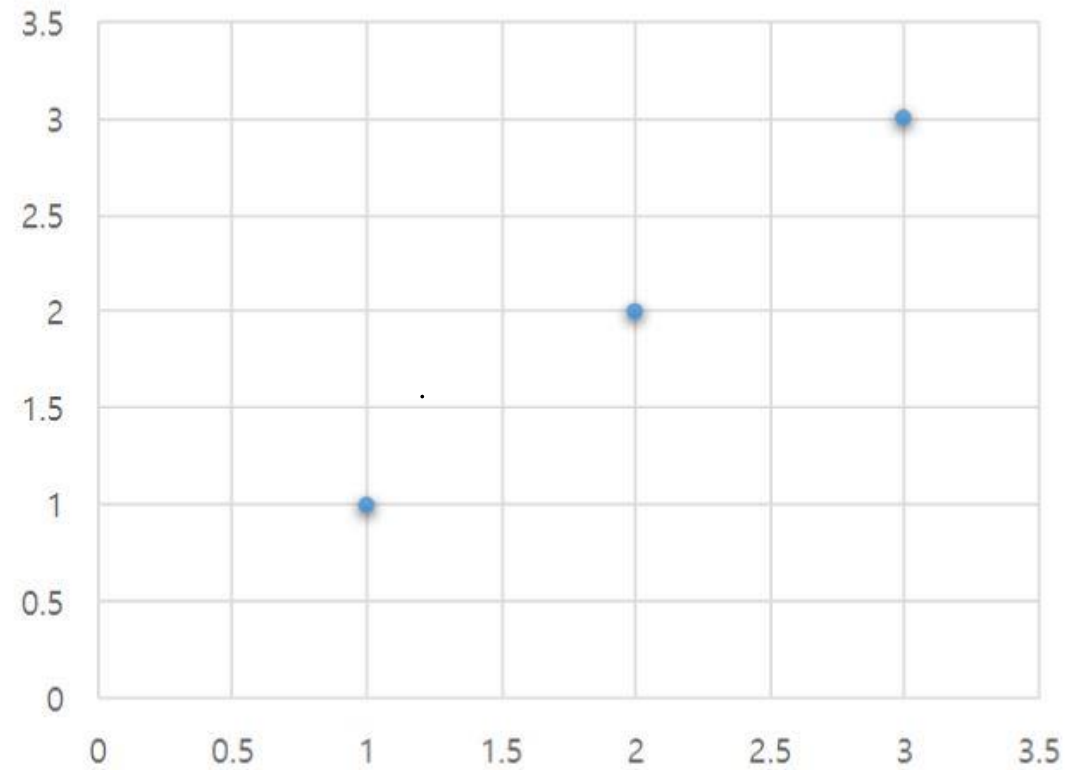
: Classification (Discrete Value - 정수형)



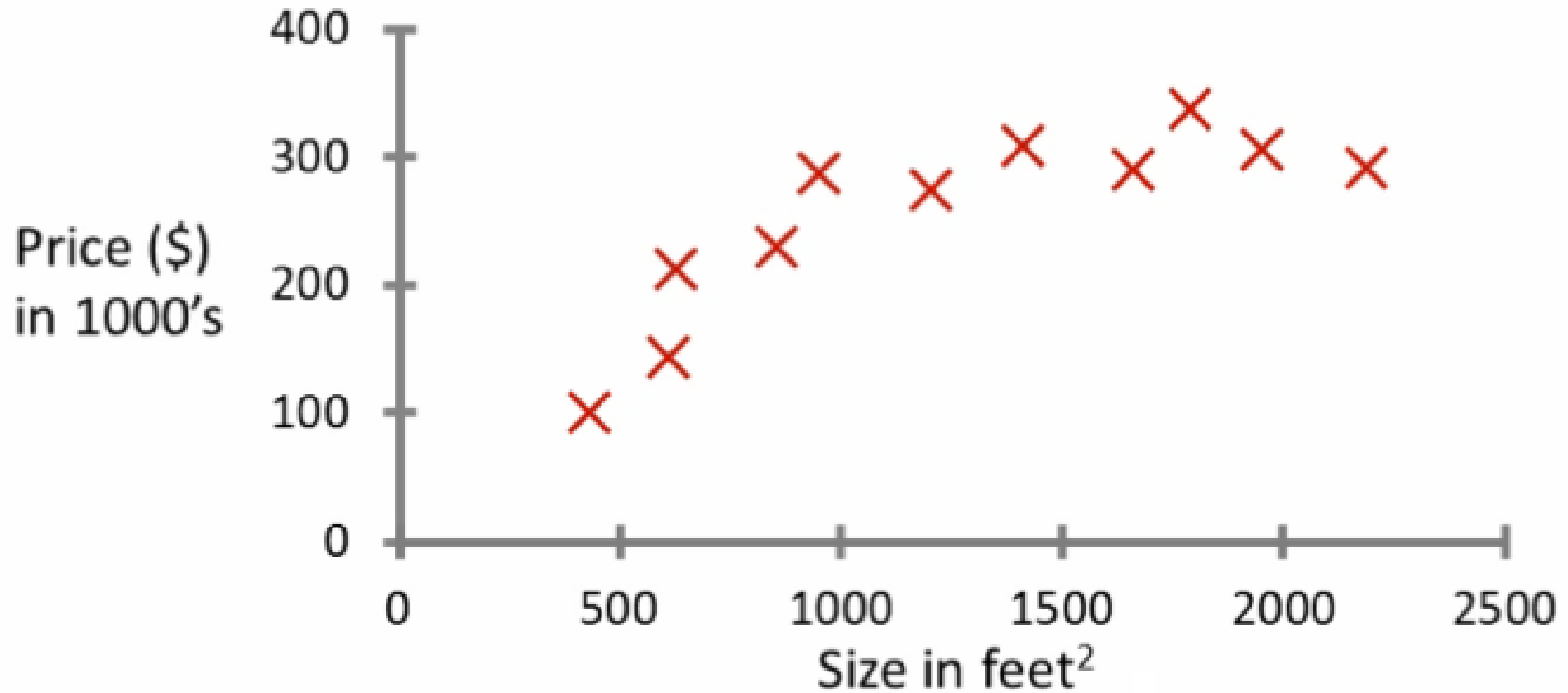
# Supervised Learning

## ✓ Supervised Learning – Regression

x	y
1	1
2	2
3	3



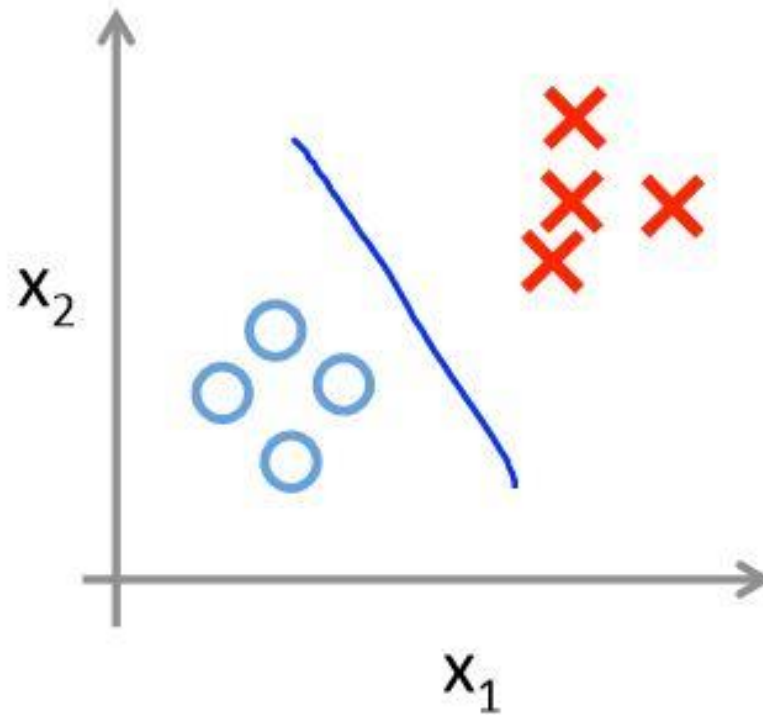
## Housing price prediction.



# Supervised Learning

## ✓ Supervised Learning – Classification

Binary classification:

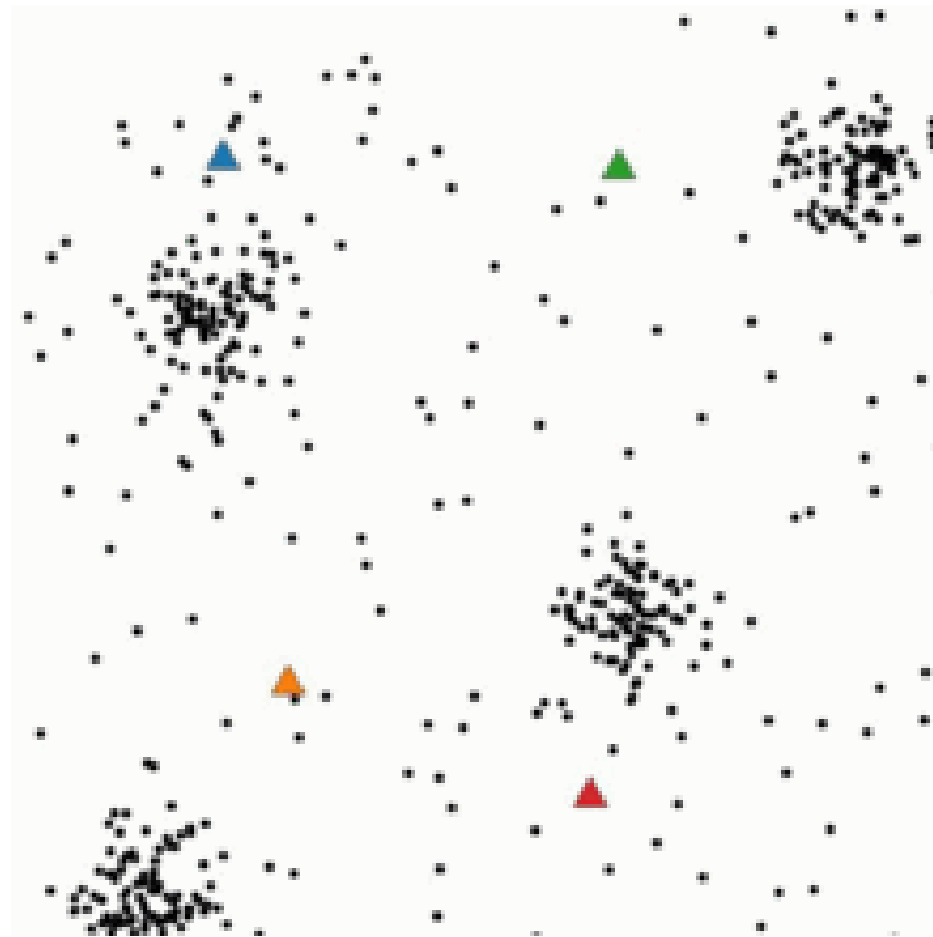


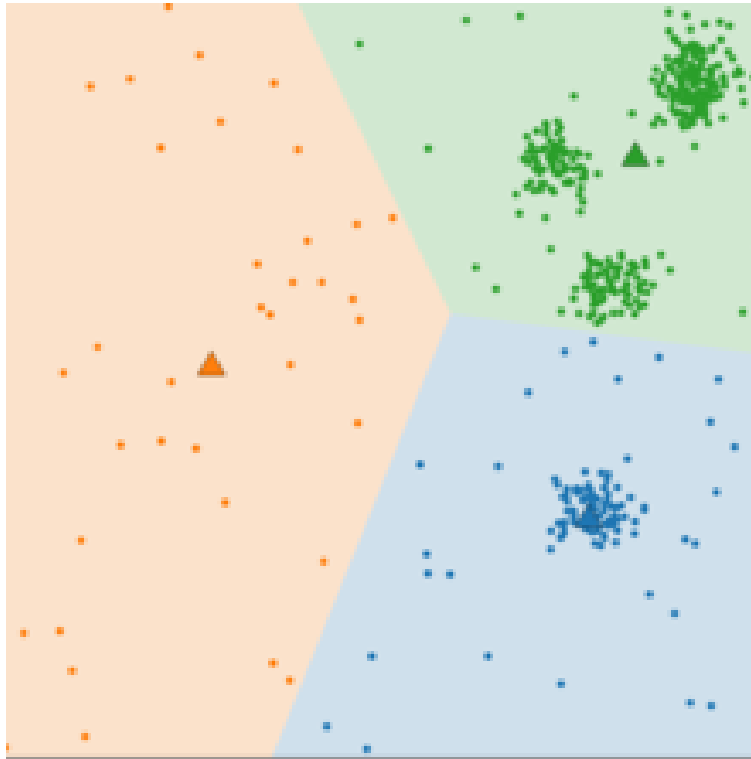
# Unsupervised Learning

✓ Unsupervised Learning

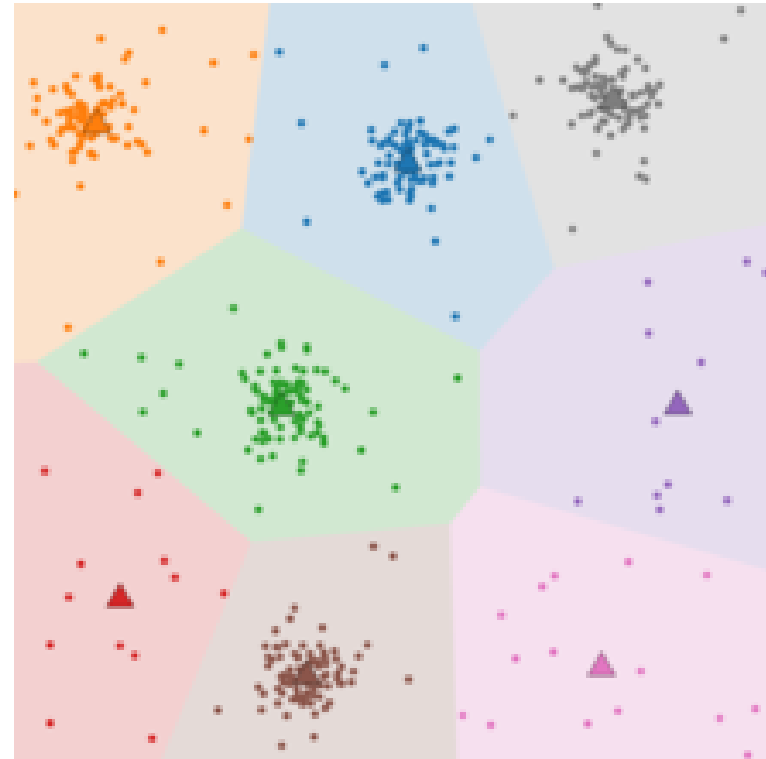
: Clustering

## ✓ Unsupervised Learning – Clustering



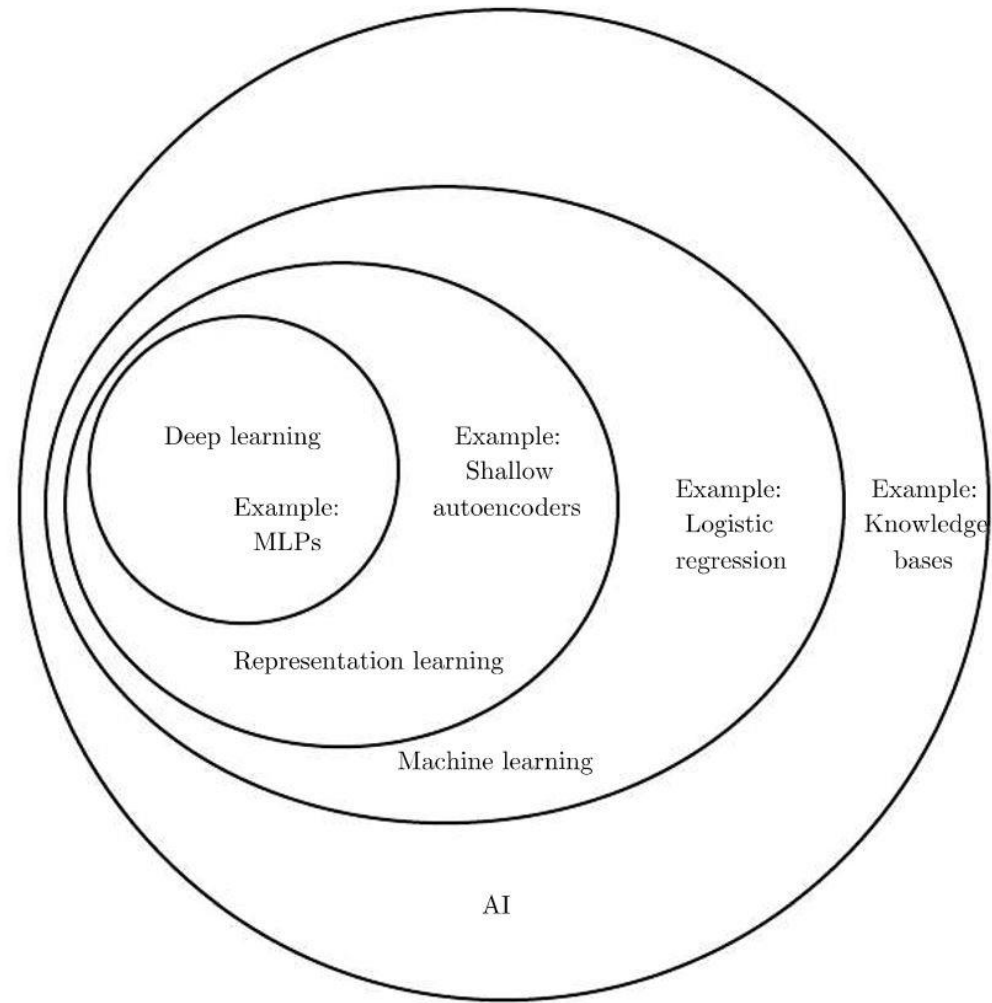


$K = 3$



$K = 8$

# Warm Up



## Warm Up

✓ Supervised Learning

✓ Unsupervised Learning



# | Q & A

# Supervised Learning & HOL

Machine Learning

Microsoft Student Partner

백진헌

# 백문이 불여일견

## 1. Linear Regression

(HOL with Azure Machine Learning)

## 2. Neural Network

(HOL with Azure Notebook)

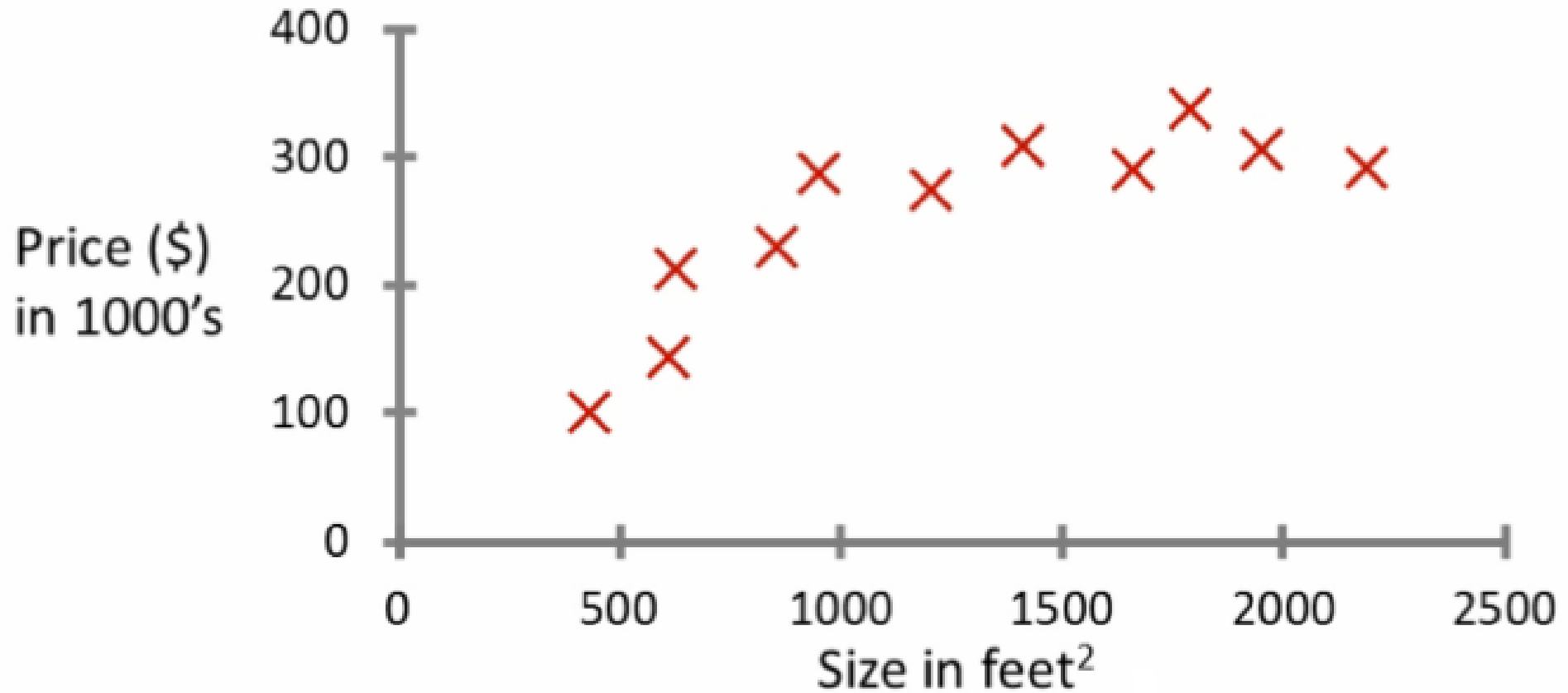
# Supervised Learning

## [Machine Learning]

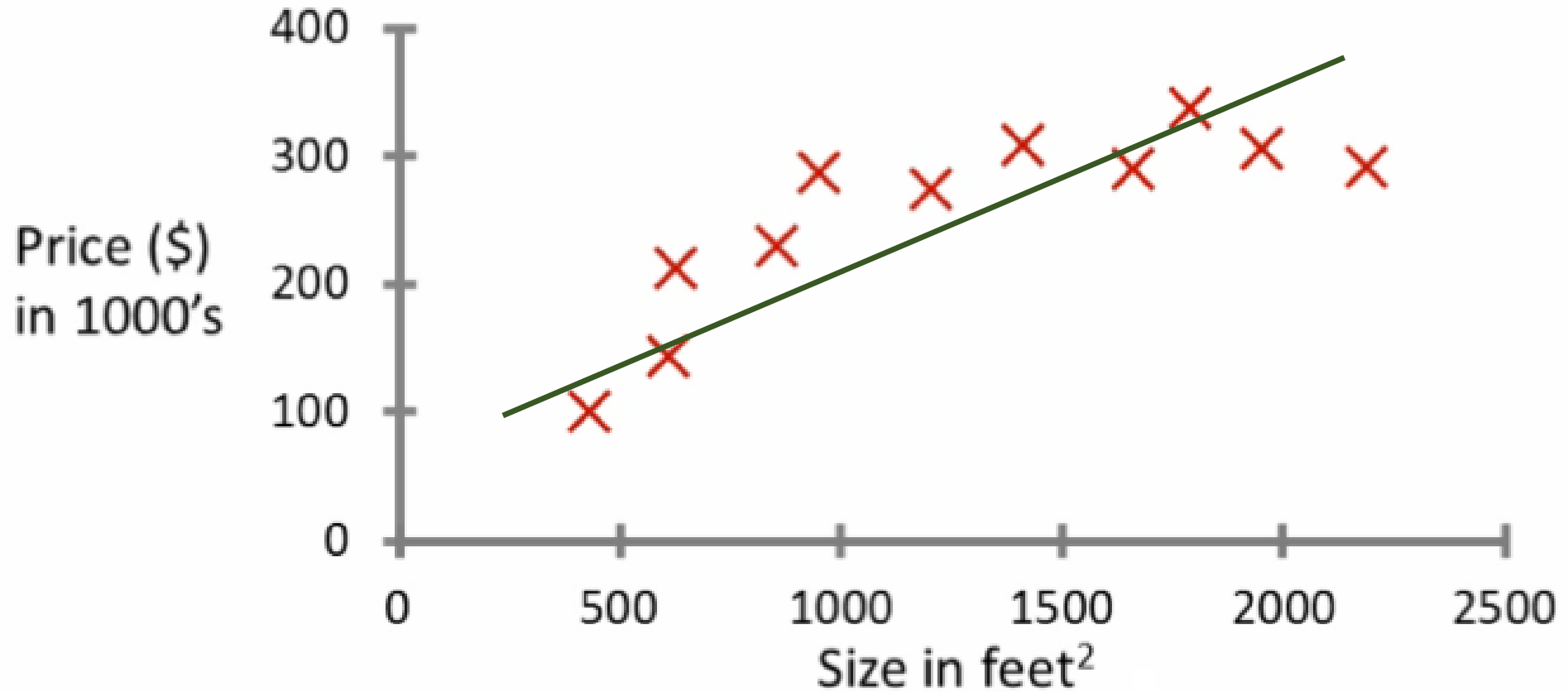
✓ Regression

: Linear Regression

## Housing price prediction.



## Housing price prediction.



# 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

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar shows the title 'Microsoft Azure Machine Learning Studio' and the user 'Jinheon Baik-Free-Works...'. The main workspace is titled 'Linear Regression' and contains a single module labeled 'Automobile price data (Raw)' with a circled '1' next to it. A search bar at the top left of the workspace contains the text 'auto'. On the left sidebar, under 'Saved Datasets', the 'Automobile price data (...)' dataset is selected. The right sidebar shows the 'Properties' tab for the selected dataset, listing details: SUBMITTED BY (Microsoft C...), SIZE (25.8 KB), FORMAT (GenericCSV), and CREATED ON (4/9/2015 7...). A 'Quick Help' section at the bottom right provides a warning: 'Missing Value Scrubber module required. Prices of various automobiles against make, model and technical specifications'. The bottom toolbar includes icons for '+ NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

# Supervised Learning

검색창: Select Columns in Dataset 선택

Automobile price data -> Select Columns ... 이어주기

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace is titled "Linear Regression" and shows a workflow with two steps: "Automobile price data (Raw)" and "Select Columns in Dataset". The "Select Columns in Dataset" step has a red error icon. The left sidebar contains a search bar and a list of modules under various categories like "Data Transformation", "Feature Selection", and "Statistical Functions". The right sidebar shows the "Properties" tab with "Experiment Properties" and "Summary" sections. The bottom toolbar includes buttons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

# Supervised Learning

Select Columns in Dataset 클릭 (오른쪽 메뉴 확인)

Launch column selector 클릭

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

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow for a Linear Regression model. The first step is 'Automobile price data (Raw)', which feeds into the 'Select Columns in Dataset' step, marked with a circled '1'. The left sidebar contains a search bar and a list of operations under categories like 'Data Transformation', 'Feature Selection', and 'Text Analytics'. The right sidebar shows the 'Properties' panel for the 'Select Columns in Dataset' step, with the 'Selected columns' set to 'All columns' and 'Exclude column names' set to 'normalized-losses'. A 'Launch column selector' button is visible. The bottom status bar includes icons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

# Supervised Learning

검색창: Split Data

Select Columns in Dataset -> Split Data (연결)

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

Microsoft Azure Machine Learning Studio interface showing a workflow for Linear Regression. The workflow consists of three steps: 'Automobile price data (Raw)', 'Select Columns in Dataset', and 'Split Data'. The 'Split Data' step is highlighted with a blue border and numbered 1 and 2. The 'Properties' panel on the right shows the 'Split Data' configuration: Splitting mode is 'Split Rows', Fraction of rows in the first set is '0.66', Randomized split is checked, Random seed is '0', and Stratified split is 'False'. The 'Mini Map' at the bottom left shows the overall workflow structure. The bottom toolbar includes buttons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

# Supervised Learning

검색창: Linear Regression

Linear Regression 오른쪽 메뉴 ->

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

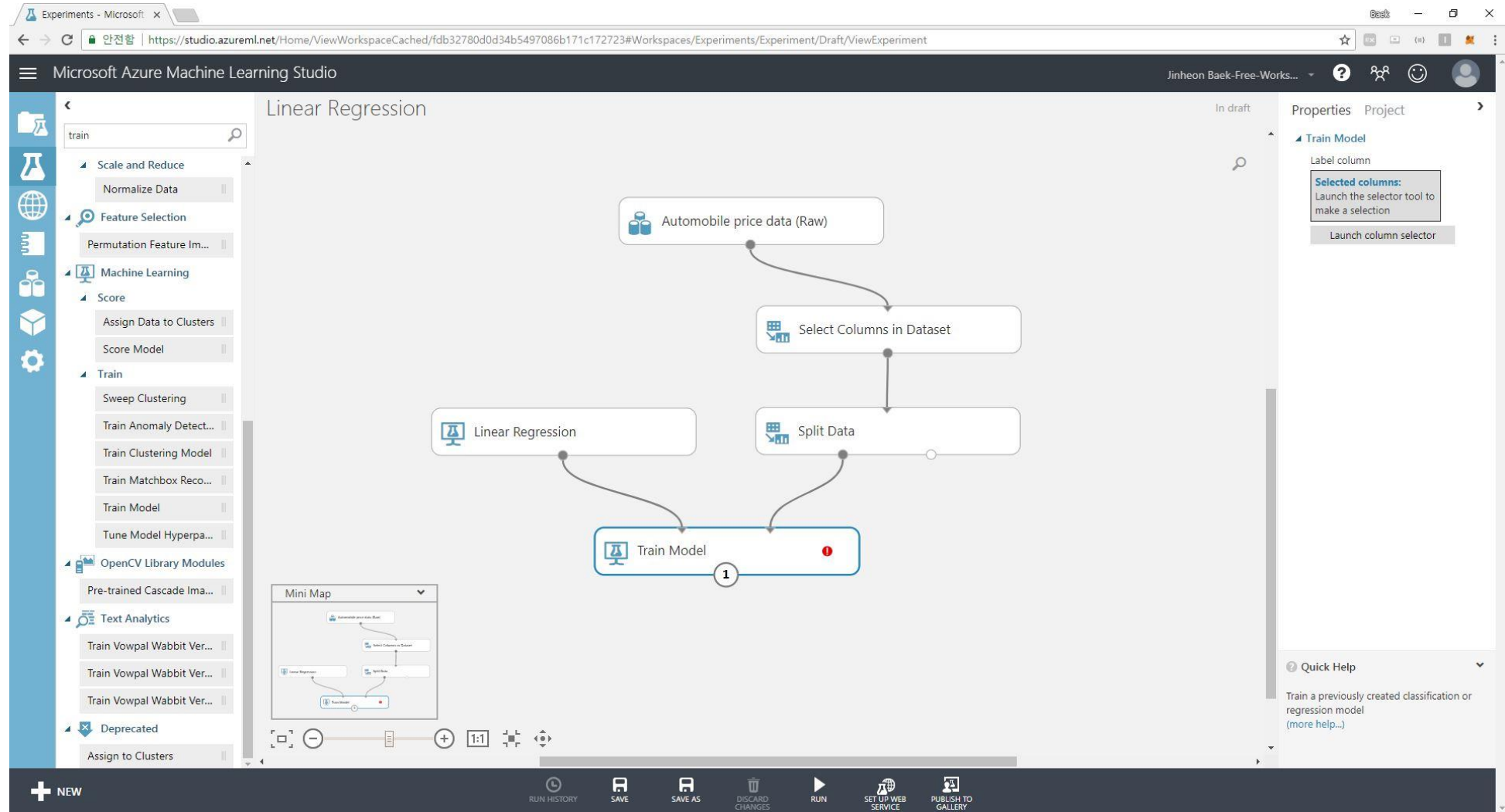
The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow for Linear Regression. The workflow starts with a dataset named 'Automobile price data (Raw)', which is processed by a 'Select Columns in Dataset' node, followed by a 'Split Data' node. The 'Linear Regression' node is highlighted with a blue border and a circled '1'. The left sidebar contains a search bar and a list of modules under the 'Machine Learning' category, including 'Initialize Model', 'Regression', and 'Statistical Functions'. The right sidebar shows the 'Properties' panel for the 'Linear Regression' module, with the 'Solution method' set to 'Online Gradient Descent' and the 'Number of training epochs' set to '30'. The bottom of the interface features a toolbar with icons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

# 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

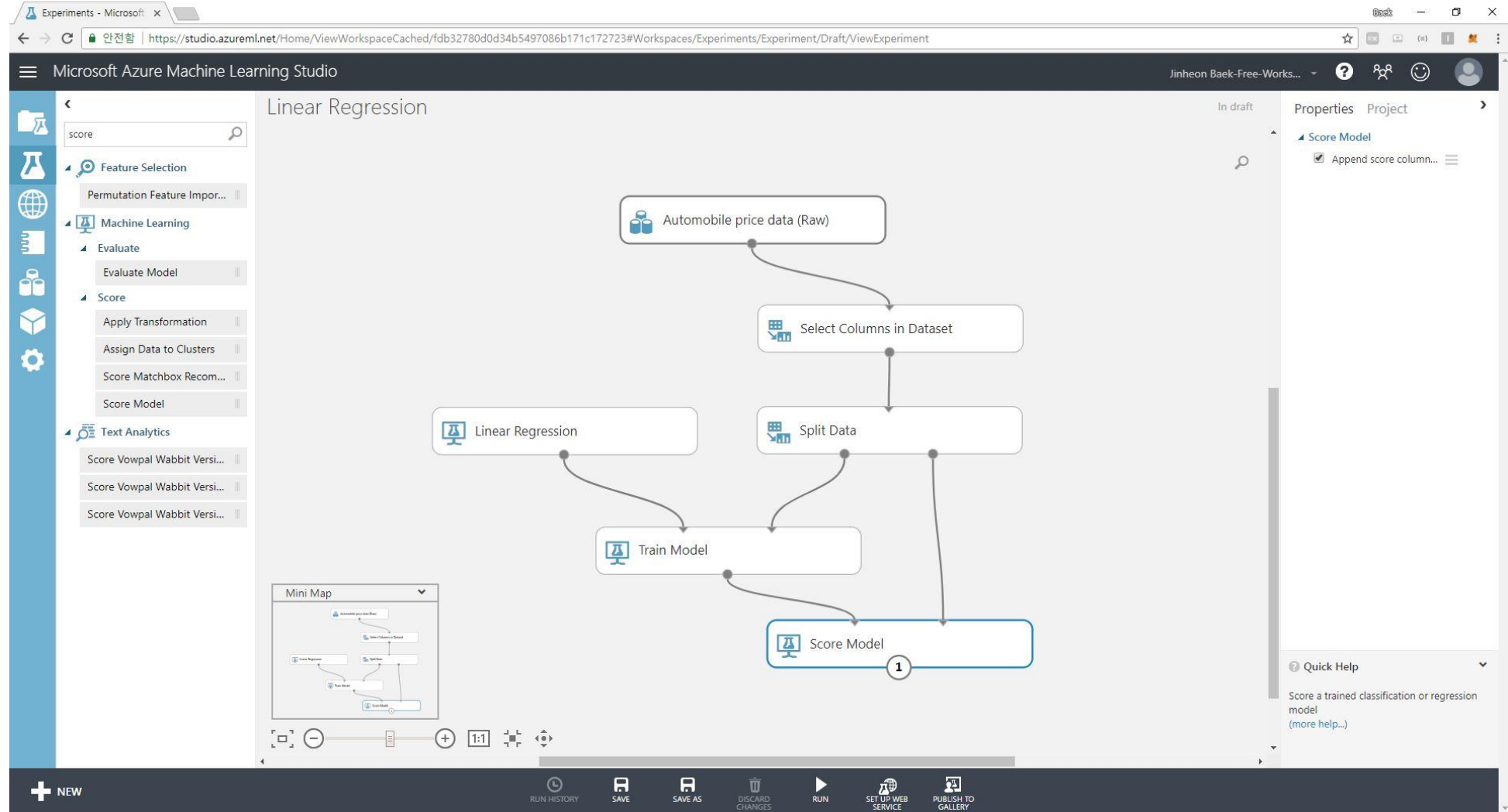
The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a 'Linear Regression' workflow with a data source 'Automobile price data (Raw)'. A 'Train Model' module is present, and a 'Select a single column' dialog is open over it. The dialog has two tabs: 'BY NAME' and 'WITH RULES'. The 'WITH RULES' tab is active, showing a rule configuration: 'Include' (selected from a dropdown) followed by 'column names' (selected from another dropdown) and 'price' (entered in a text field). The right sidebar contains the 'Properties' pane for the 'Train Model' module, which includes a 'Label column' section with a 'Launch column selector' button. The bottom of the interface features a toolbar with icons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

# Supervised Learning

검색창: Score Model

Train Model -> Score Model (왼쪽 연결)

Split Data -> Score Model (오른쪽 연결)

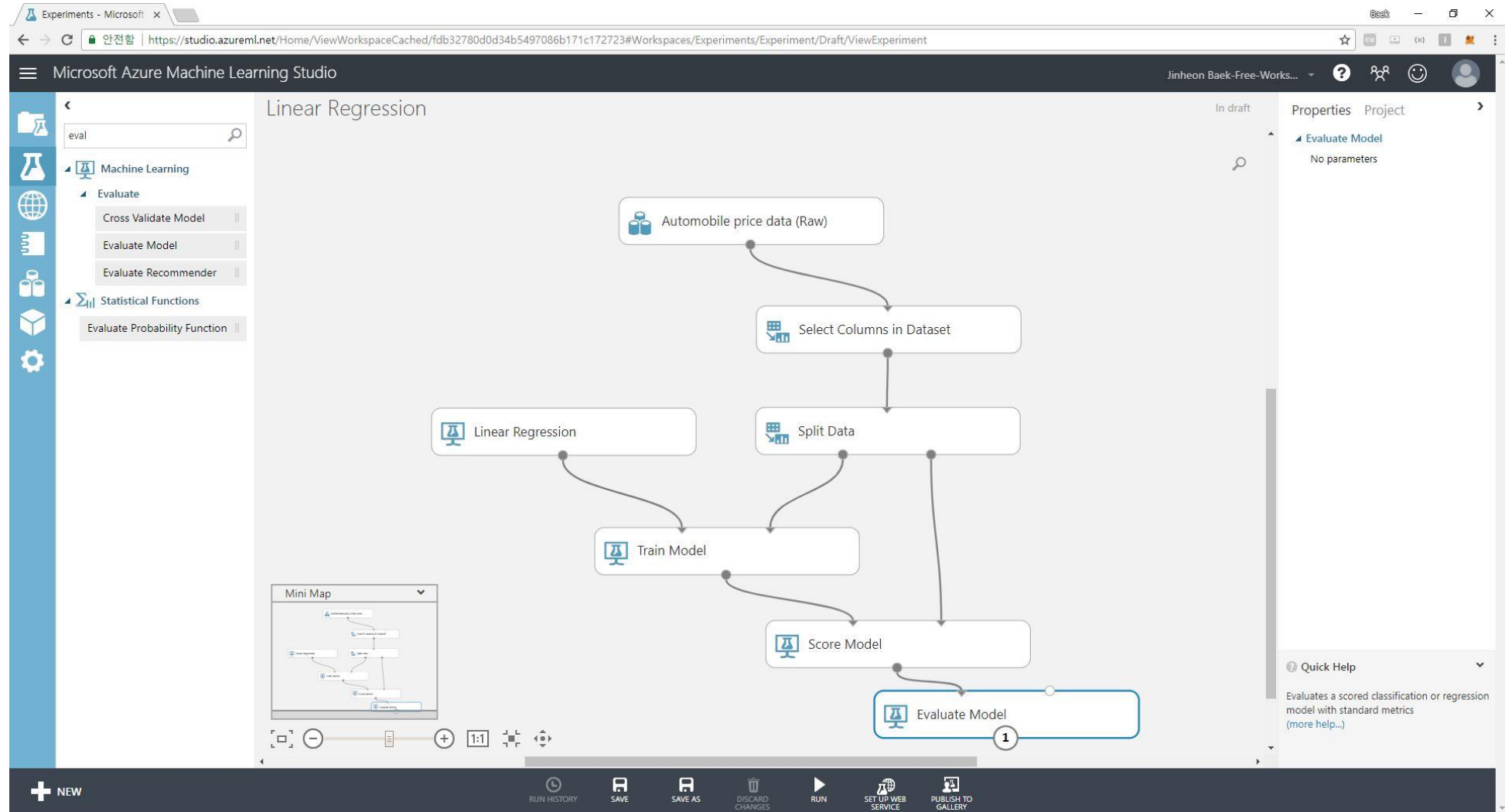


# Supervised Learning

검색창: Evaluate Model

Score Model -> Evaluate Model (왼쪽 연결)

하단 Run 실행 후 결과 보기 !

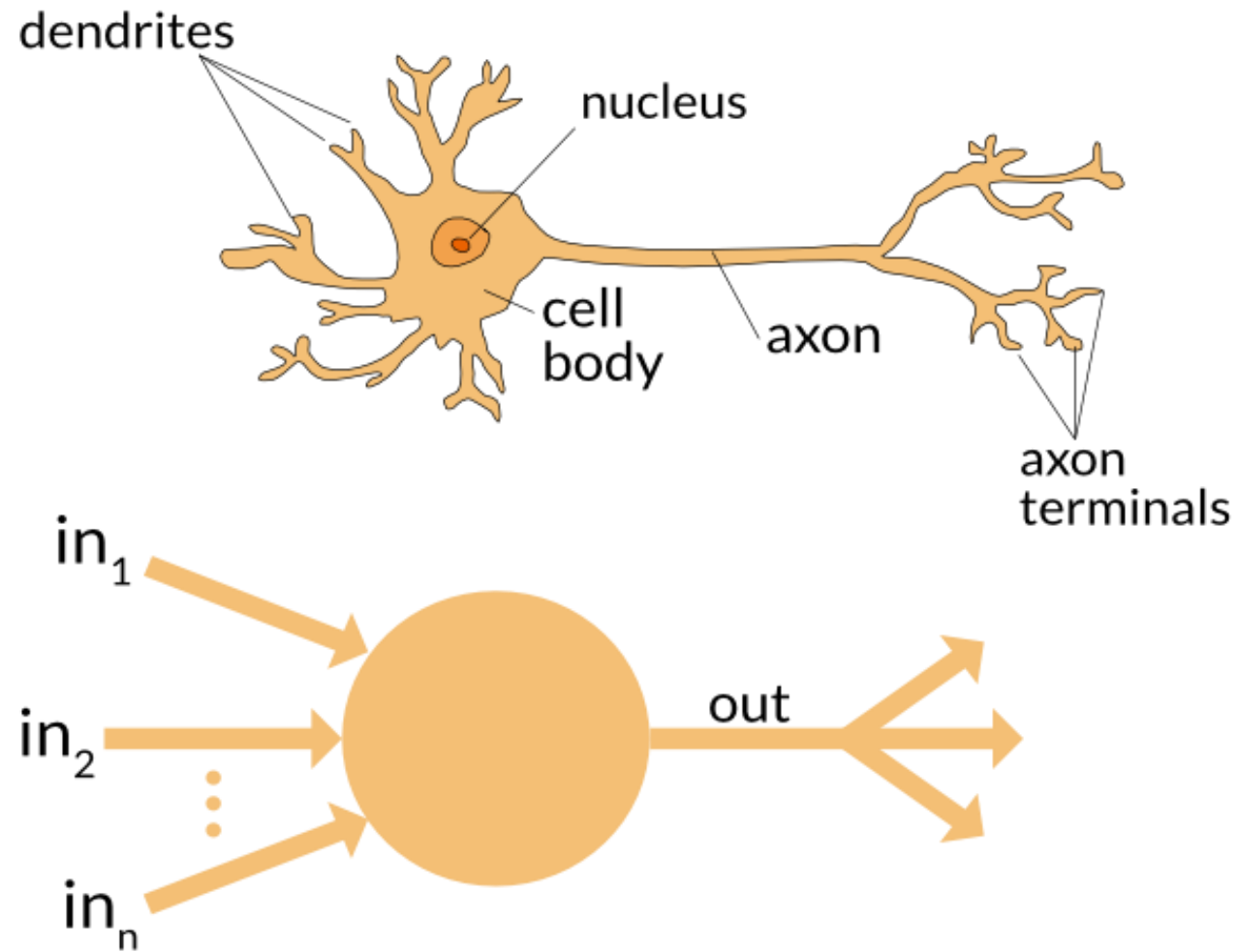


# Supervised Learning

## [Deep Learning]

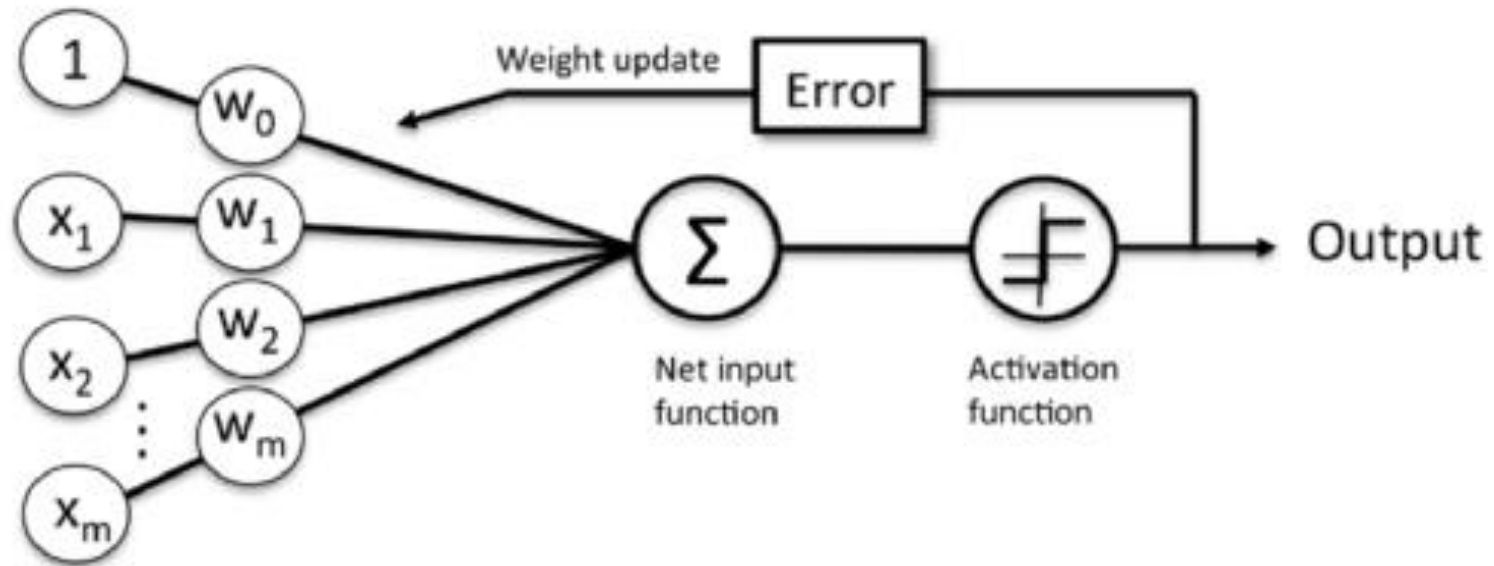
- ✓ Neural Network
  - : Simple MLP (Multi-Layer Perceptron)

# Supervised Learning



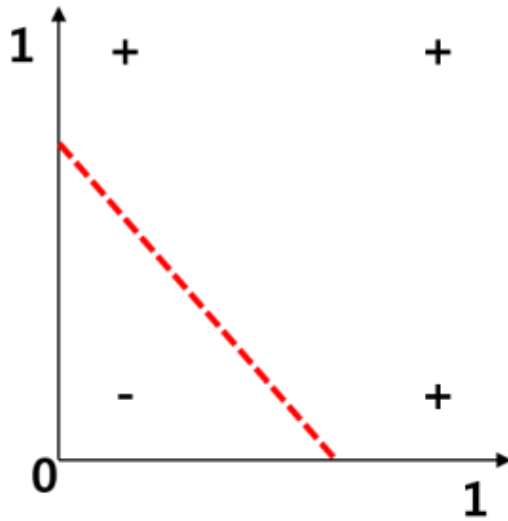
# Supervised Learning

## Single Layer Neural Network



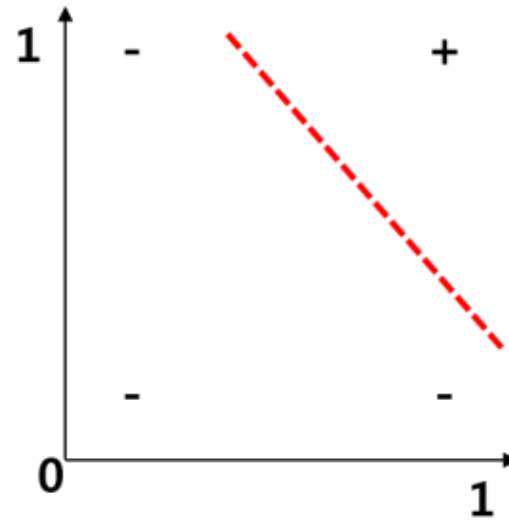
# Supervised Learning

OR



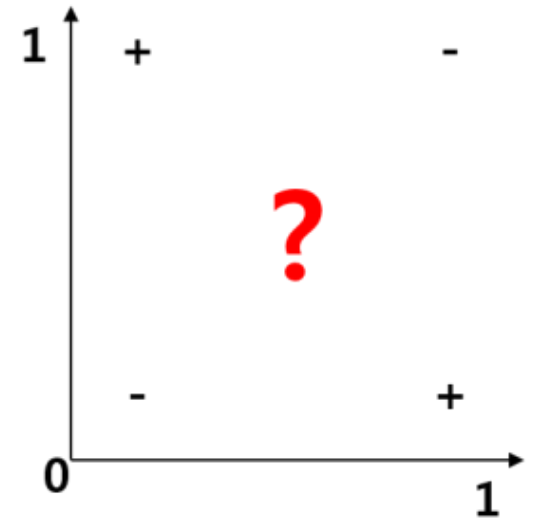
$x_1$	$x_2$	$y$
0	0	0
0	1	1
1	0	1
1	1	1

AND



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

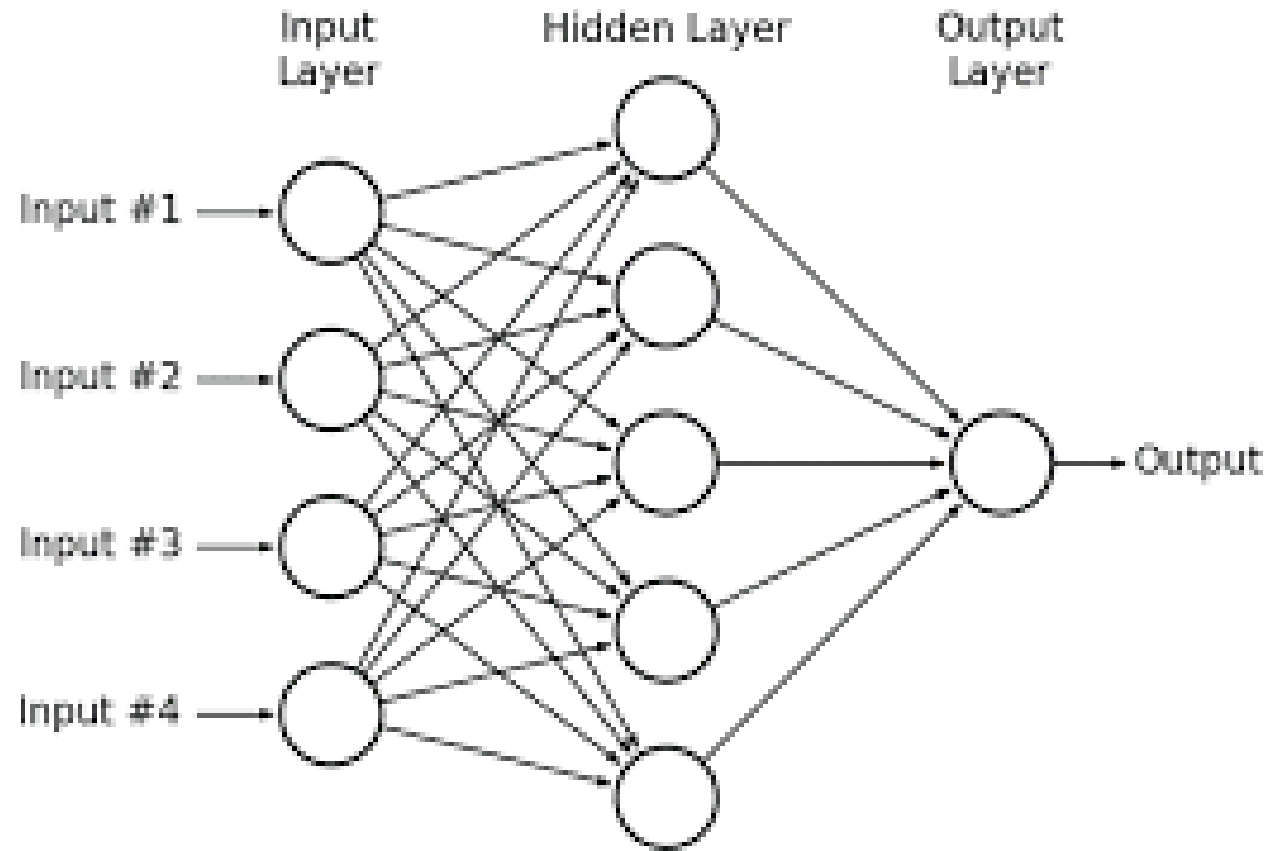
XOR



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

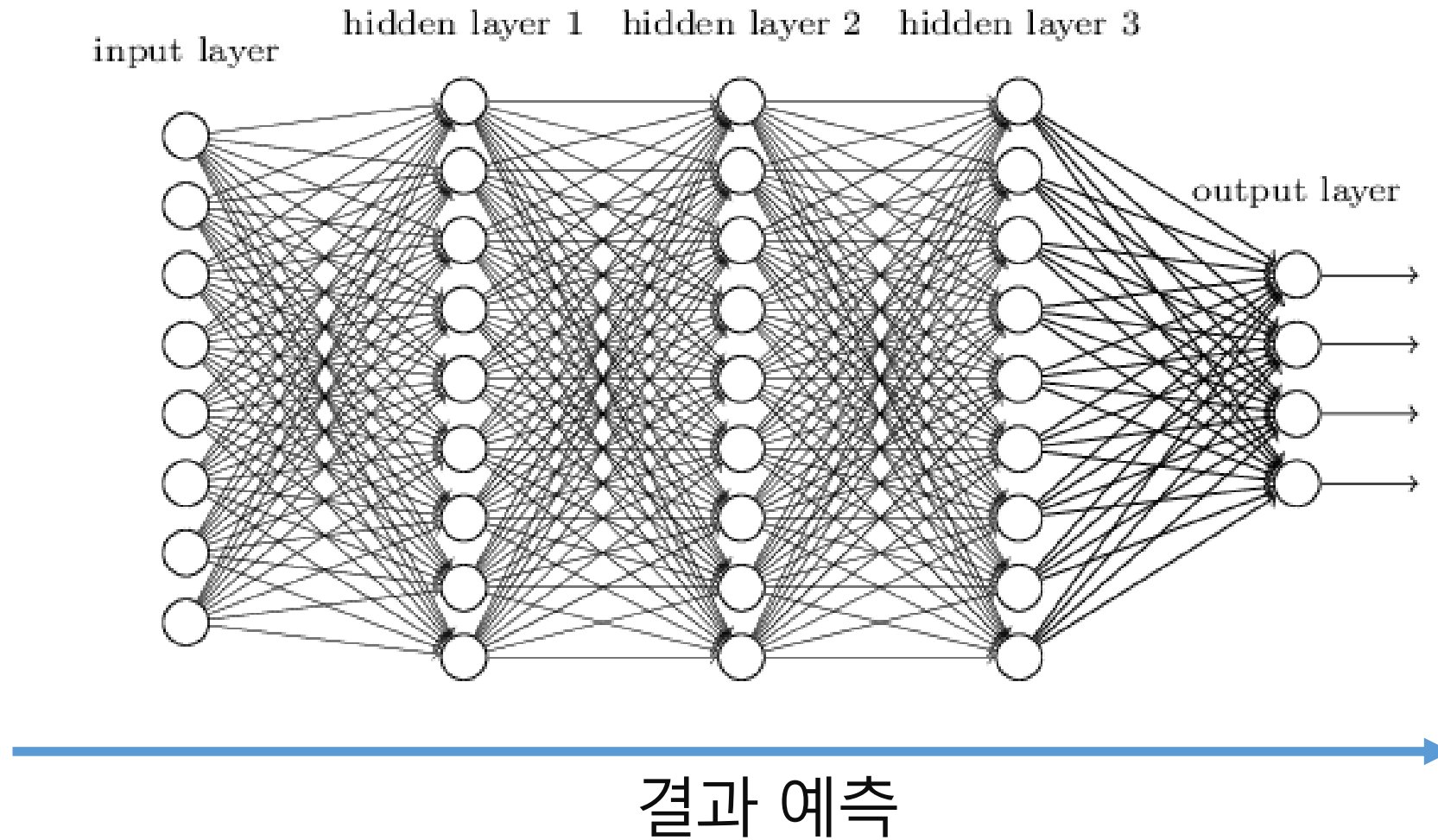
# Supervised Learning

## Multi-Layer Neural Network

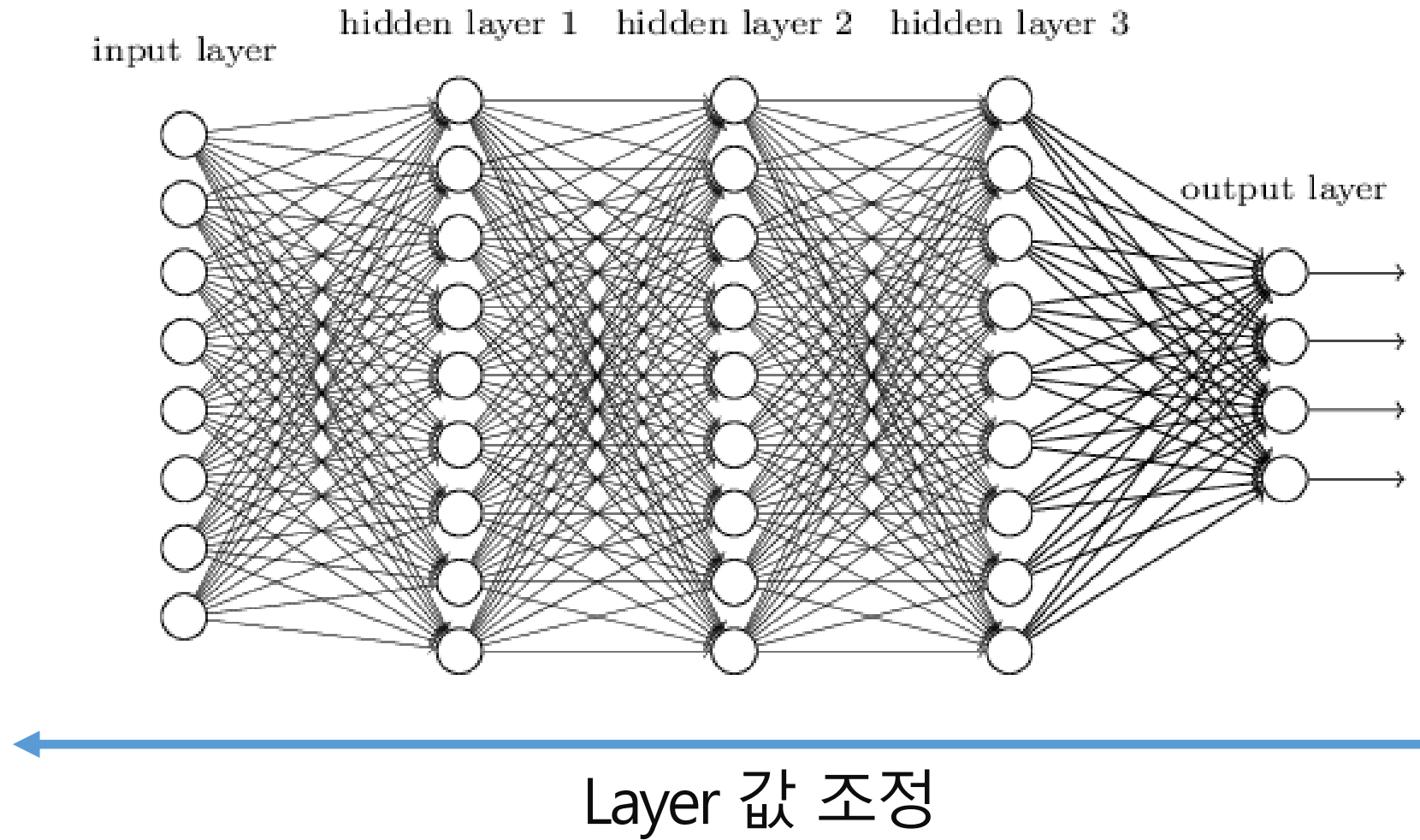




# Neural-network



# Neural-network



✓HOL

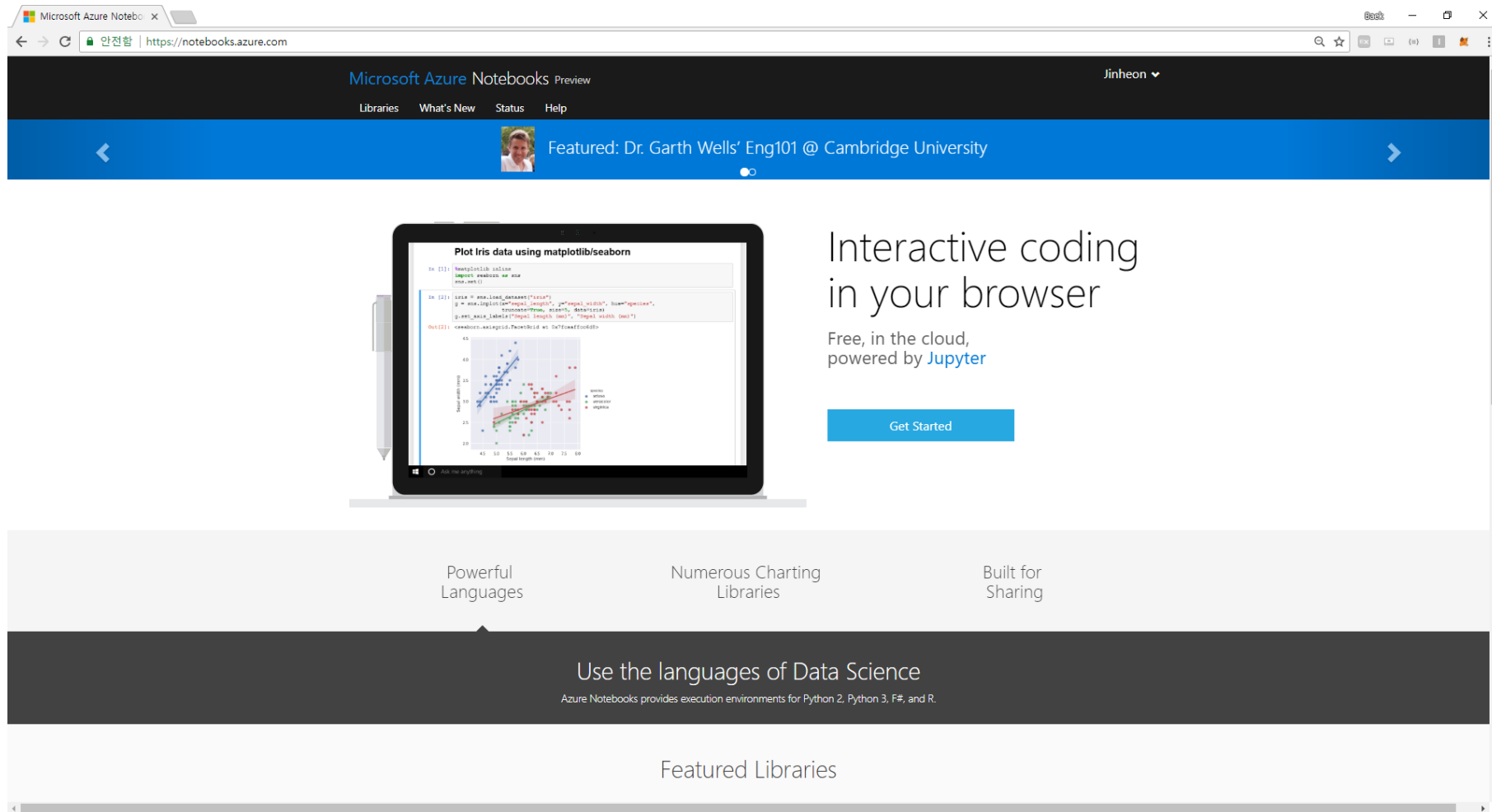
“MNIST Data Set ?”



28 x 28, 필기체 숫자 ( 0 ~ 9 )

# Neural Network

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



The screenshot shows the Microsoft Azure Notebooks web interface. At the top, there's a navigation bar with 'Libraries', 'What's New', 'Status', and 'Help'. Below this is a featured banner for 'Dr. Garth Wells' Eng101 @ Cambridge University'. The main content area features a central image of a laptop displaying a scatter plot titled 'Plot Iris data using matplotlib/seaborn'. To the right of the laptop, the text reads 'Interactive coding in your browser' and 'Free, in the cloud, powered by Jupyter', with a 'Get Started' button. Below this, three key features are listed: 'Powerful Languages', 'Numerous Charting Libraries', and 'Built for Sharing'. A dark grey bar contains the text 'Use the languages of Data Science' and 'Azure Notebooks provides execution environments for Python 2, Python 3, F#, and R.' At the bottom, there's a section for 'Featured Libraries'.

Microsoft Azure Notebooks Preview Jinheon

Libraries What's New Status Help

Featured: Dr. Garth Wells' Eng101 @ Cambridge University

Plot Iris data using matplotlib/seaborn

```
In [1]: %matplotlib inline
import pandas as pd
from sklearn import datasets

In [2]: iris = datasets.load_iris()
g = pd.DataFrame(iris.data, columns=['sepal_length', 'petal_length', 'sepal_width', 'petal_width'], index=range(iris.data.shape[0]))
g['species'] = iris.target
g['species'] = g['species'].map({'0': 'setosa', '1': 'versicolour', '2': 'virginica'})

Out[2]:
```

Interactive coding in your browser

Free, in the cloud, powered by Jupyter

Get Started

Powerful Languages Numerous Charting Libraries Built for Sharing

Use the languages of Data Science

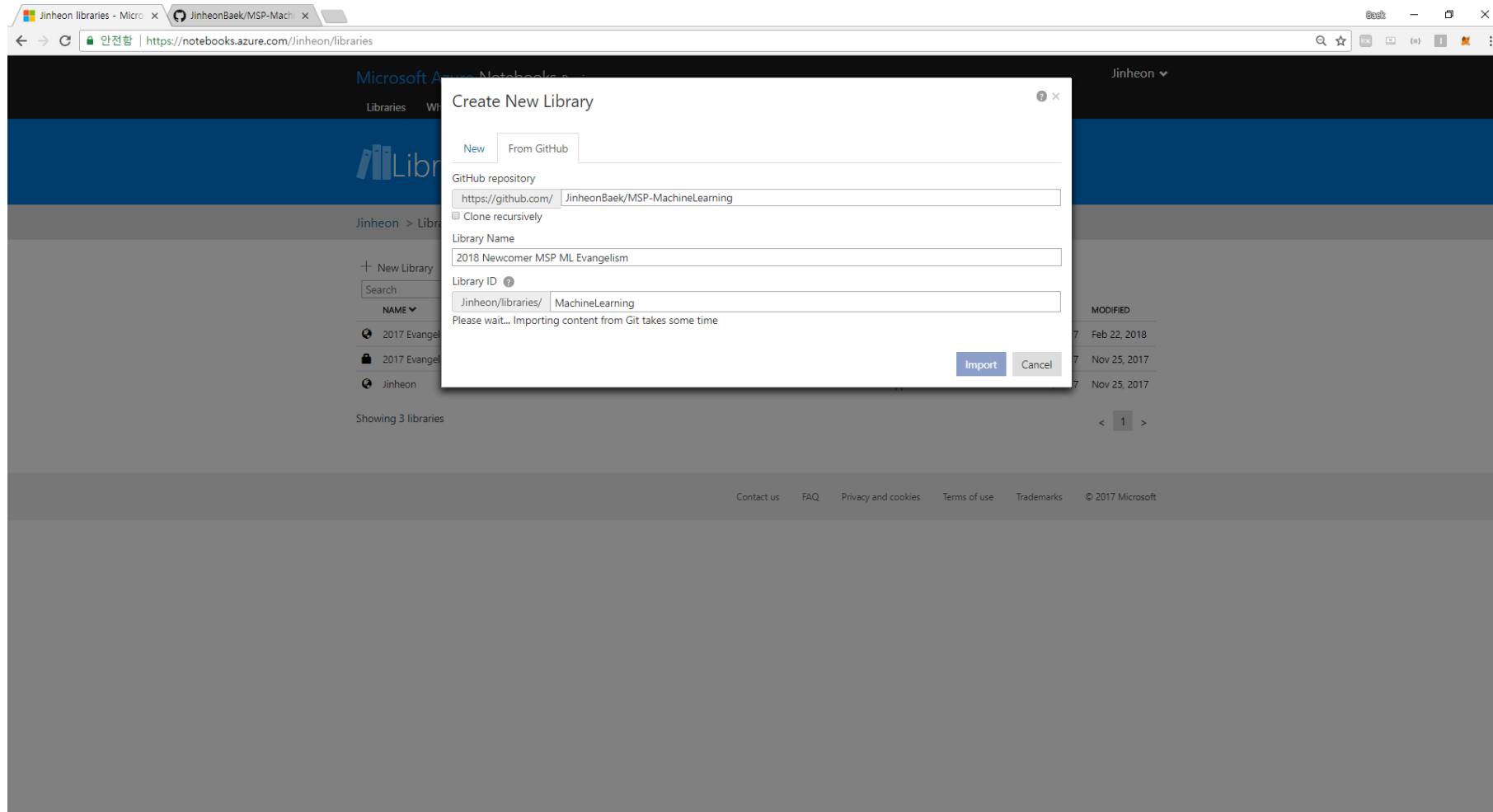
Azure Notebooks provides execution environments for Python 2, Python 3, F#, and R.

Featured Libraries

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

1. Logic
2. Libraries 메뉴
3. New Library

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



✓HOL – Challenge

**“Testing score is above 0.98”**



✓HOL – Challenge

**“Testing score is above 0.98”**

# Fitting

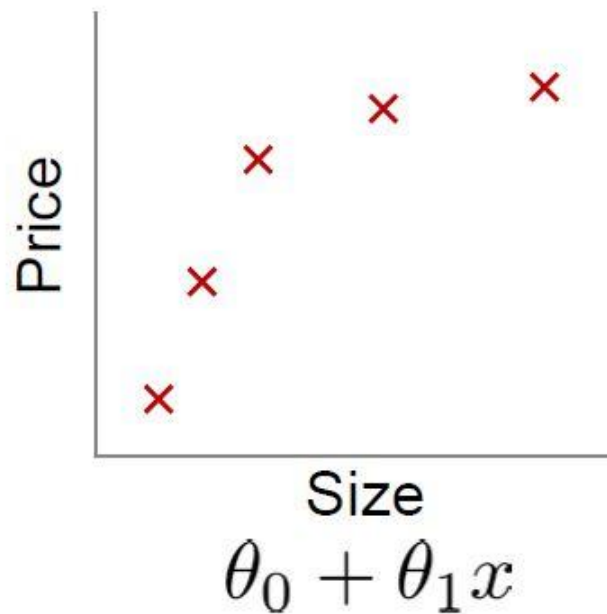
## Machine Learning

Microsoft Student Partner

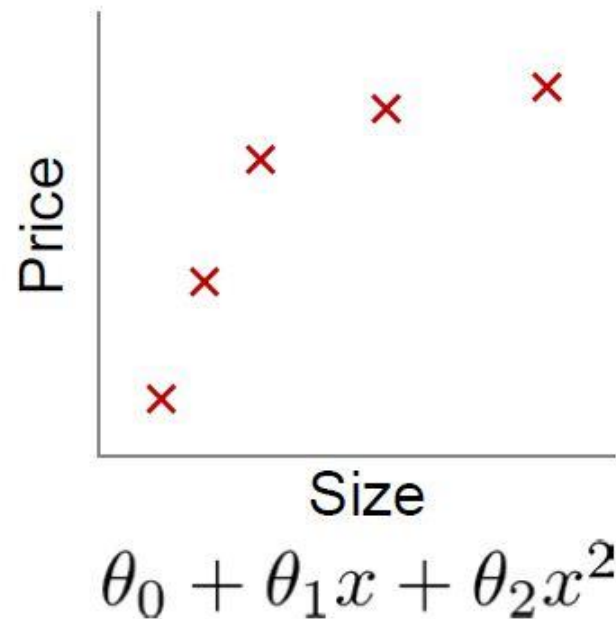
백진헌

## Fitting

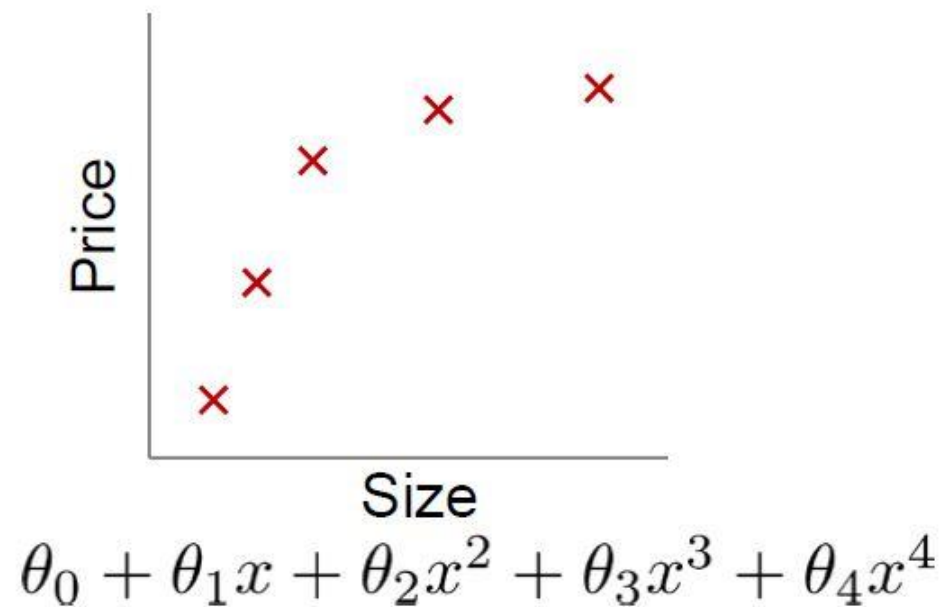
✓ Fitting



High bias  
(underfit)



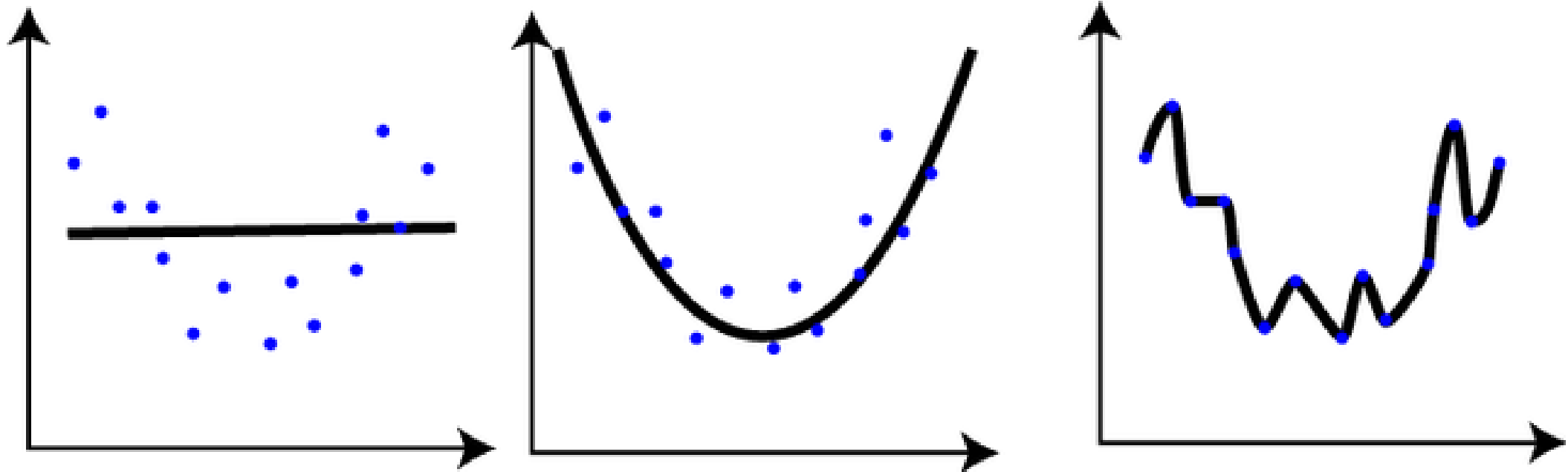
“Just right”



High variance  
(overfit)

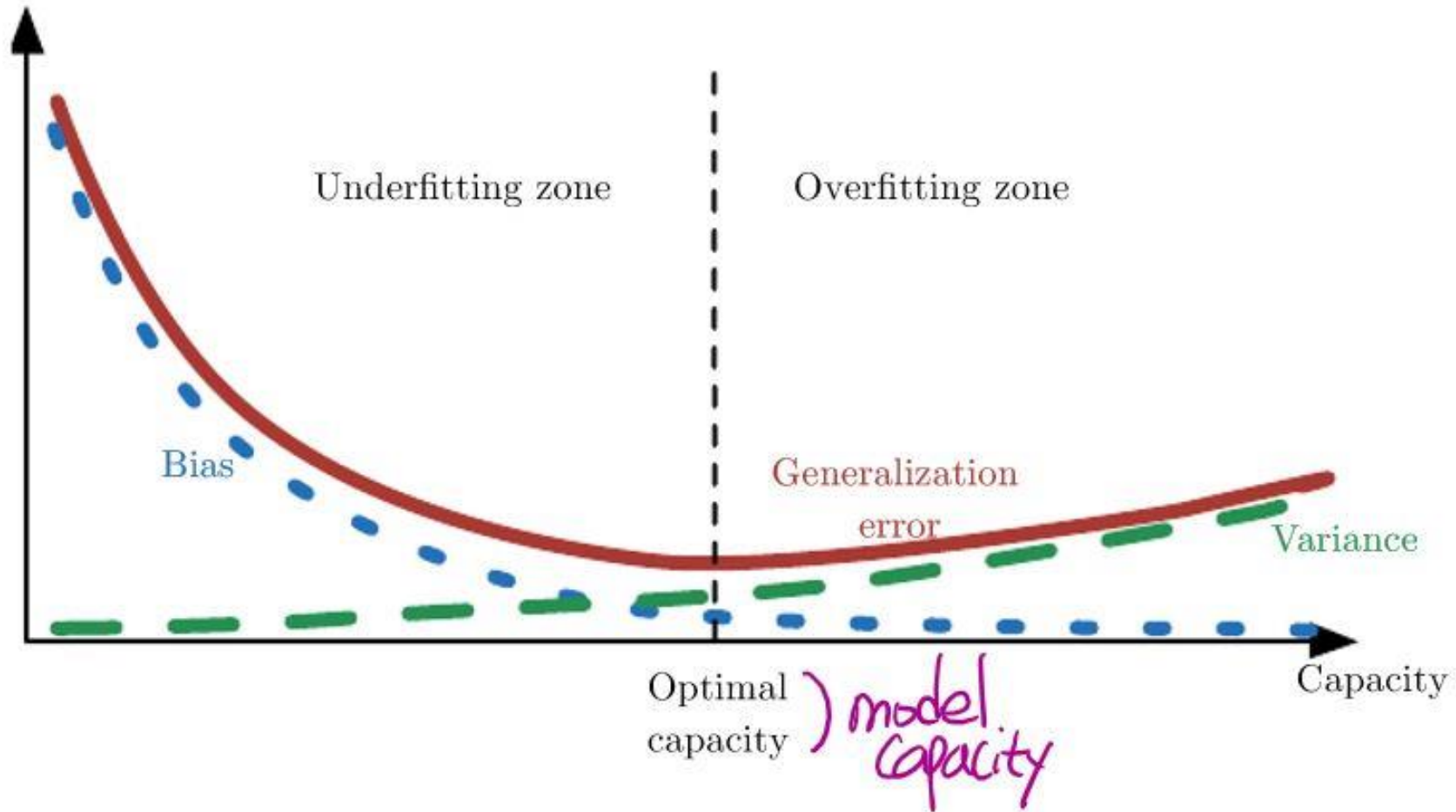
# Fitting

## ✓ Fitting



# Fitting

✓ Fitting



# Model

Machine Learning

Microsoft Student Partner

백진헌

### ✓ 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

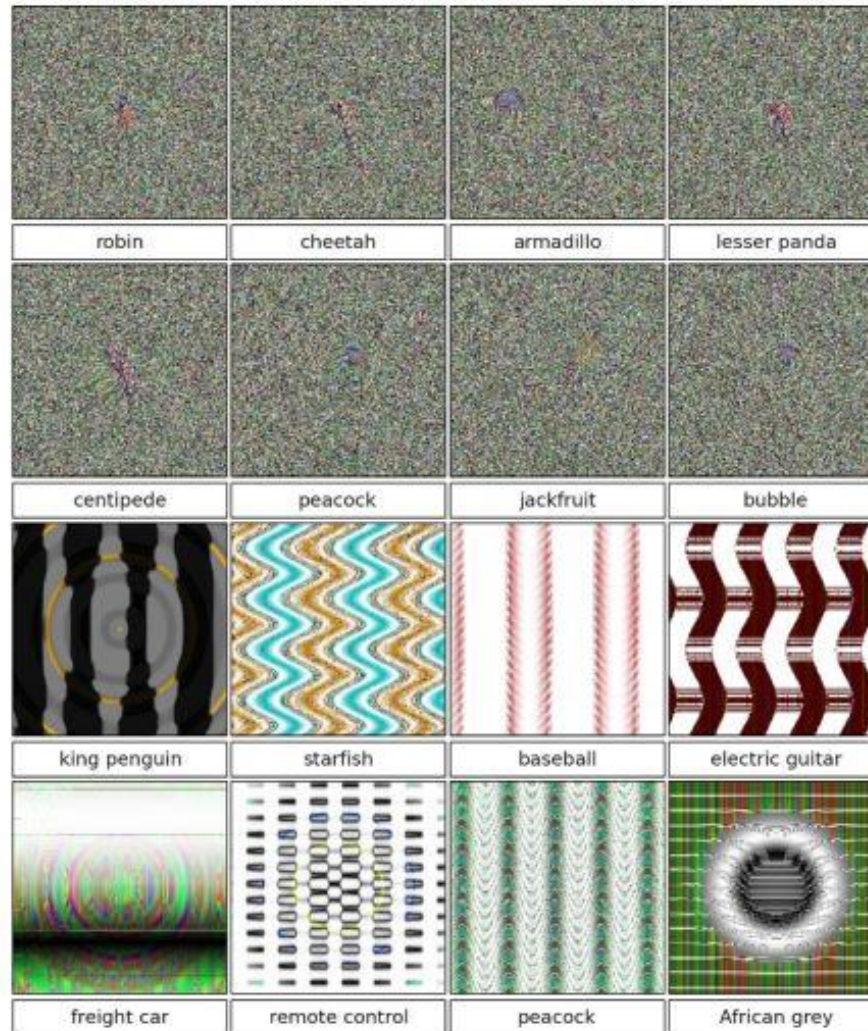
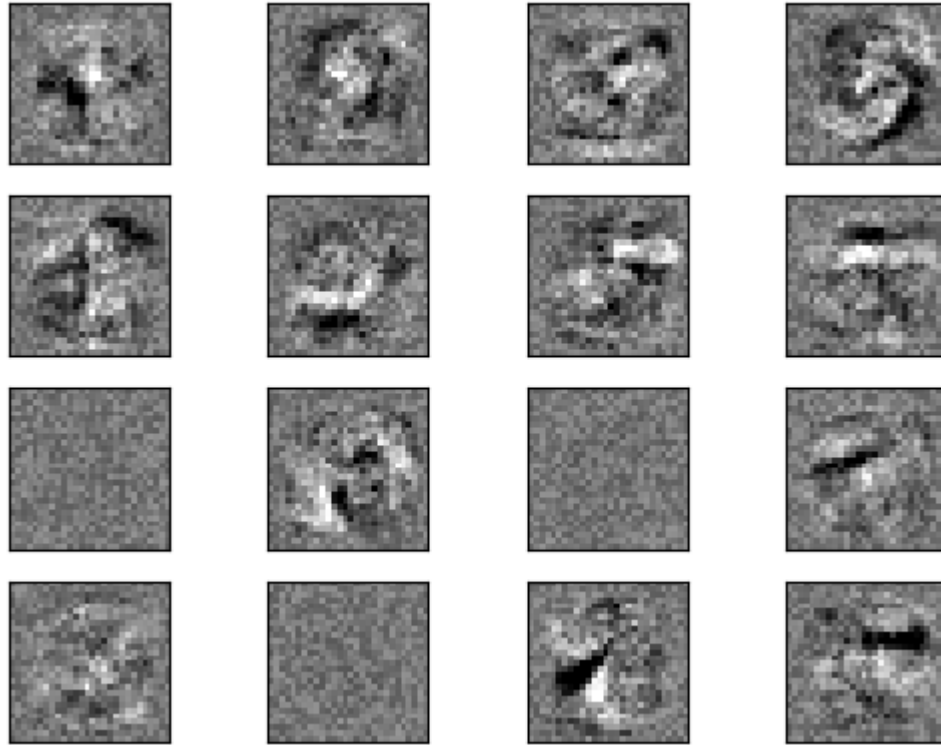


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with  $\geq 99.6\%$  certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly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



# Models



### ✓ First-principle vs Data-driven

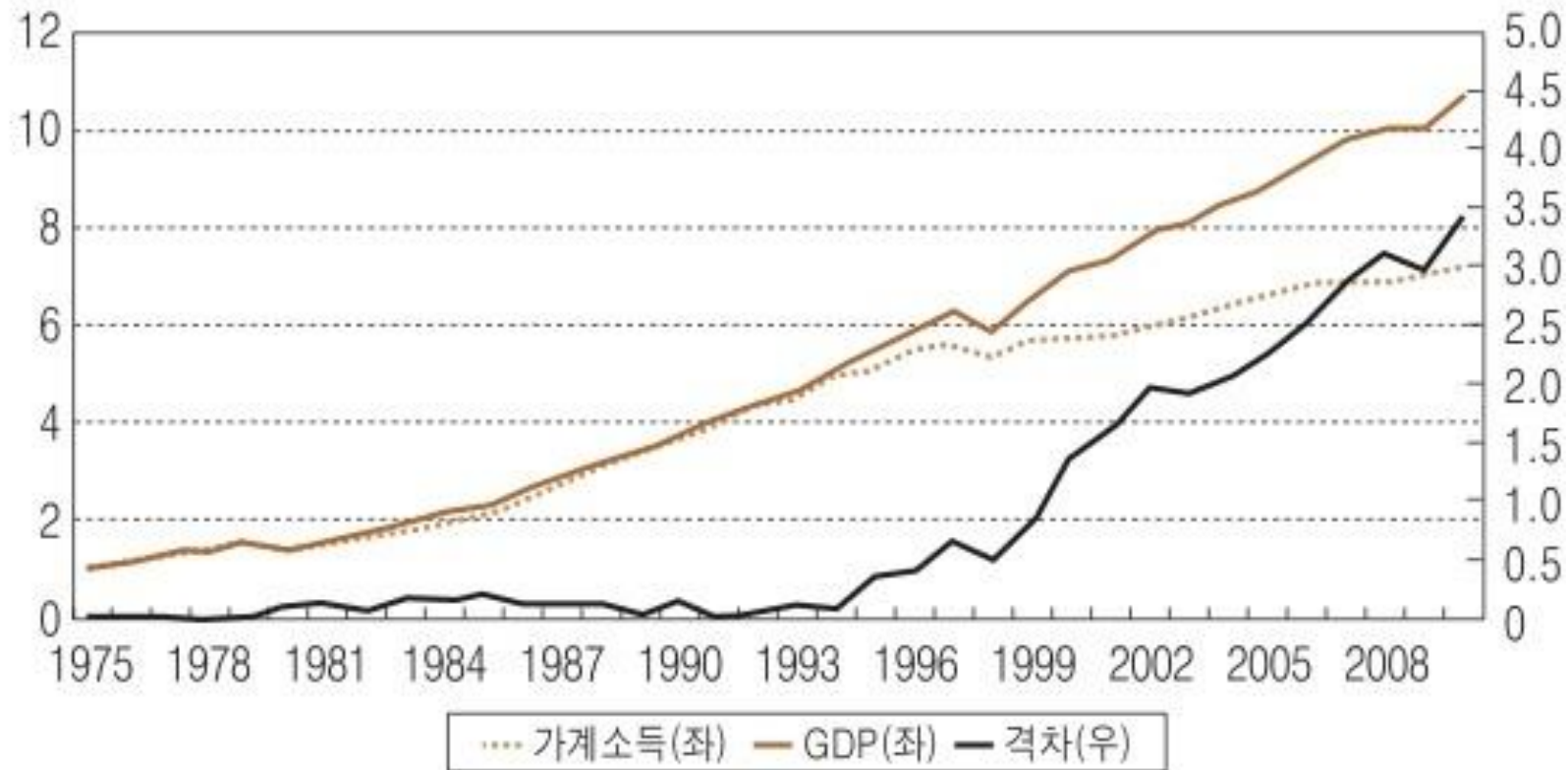
#### First-principle

: Models based on a theoretical explanation.

#### Data-driven

: Models based on observed data correlations.

〈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