Lecture 1.
Intro to ML

Today's Contents

What is Machine Learning?

AI? ML? DL?

What is Machine Learning?

AI? ML? DL?

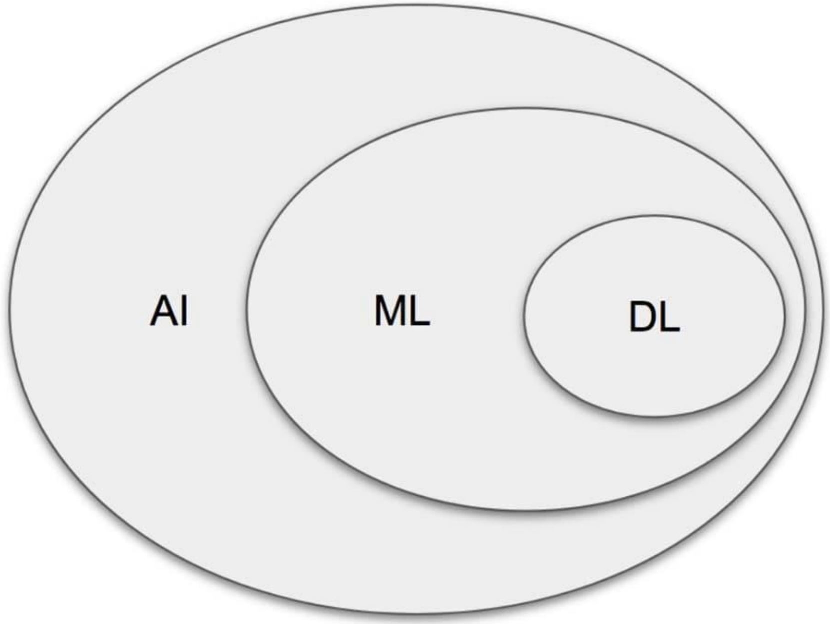
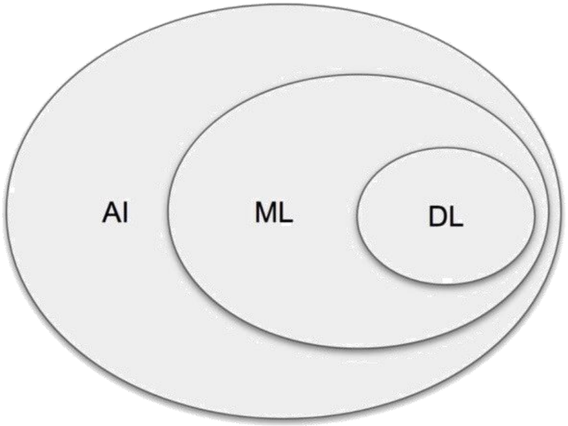


image from sonix.ai

What is Machine Learning?

AI is the intelligence demonstrated by Machines

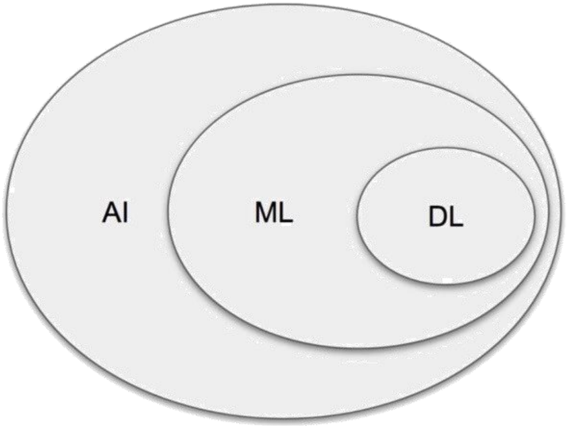


What is Machine Learning?

*"The field of **machine learning** is concerned with the question of how to construct computer programs that automatically improve with experience."*

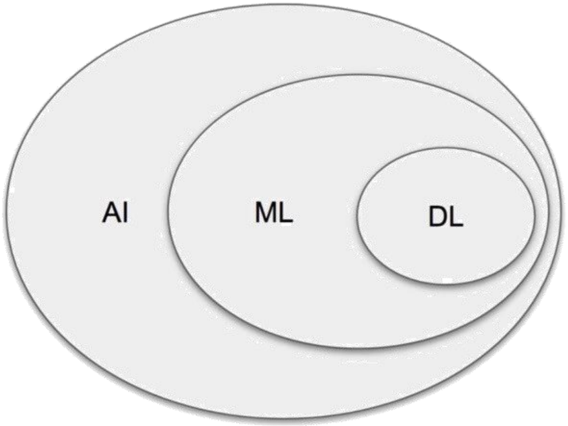
\- Tom M. Mitchell

경험을 통해 학습, 발전하는 컴퓨터 프로그램 만들기



What is Machine Learning?

Deep Learning



Fields of ML

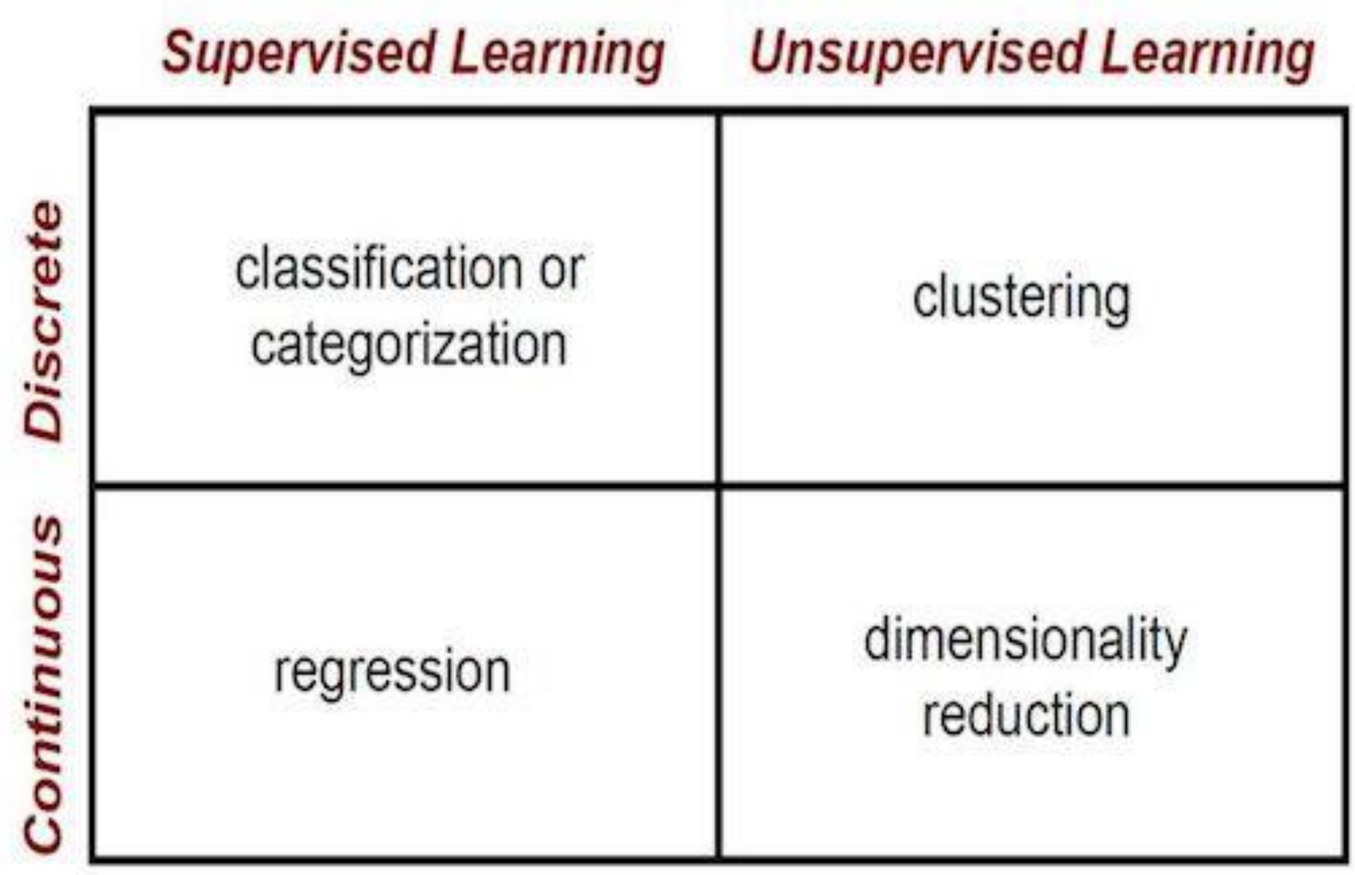
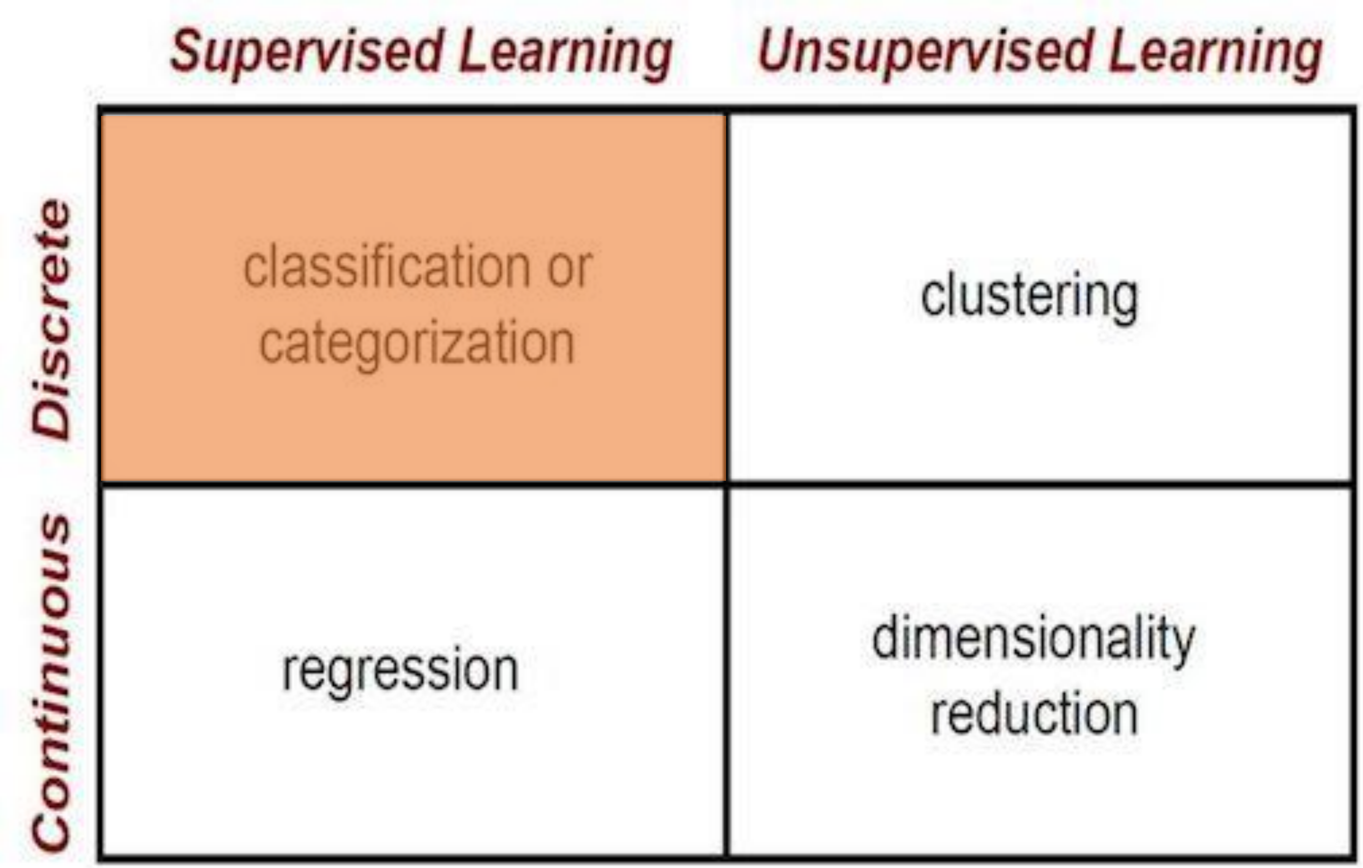


image from towardsdatascience.com

Fields of ML



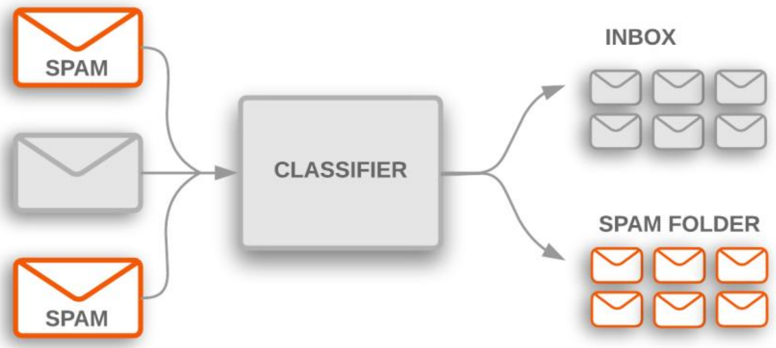


image from developers.google.com

Fields of ML

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

Fields of ML

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

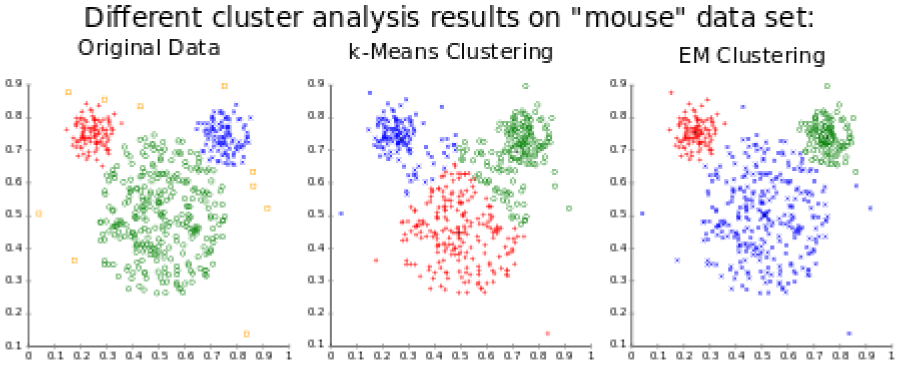


image from wikipedia.org

Fields of ML

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

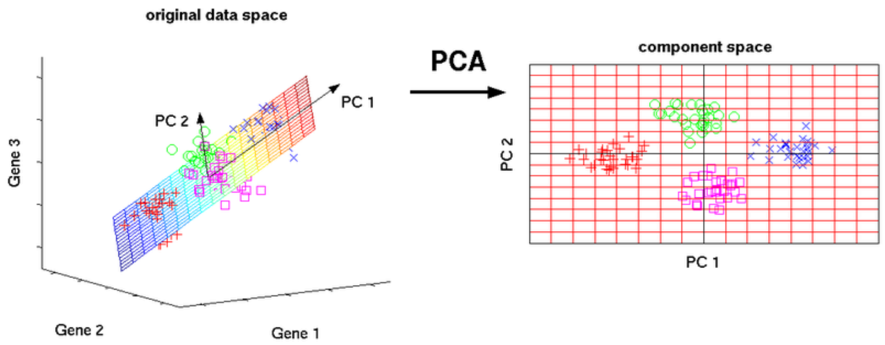
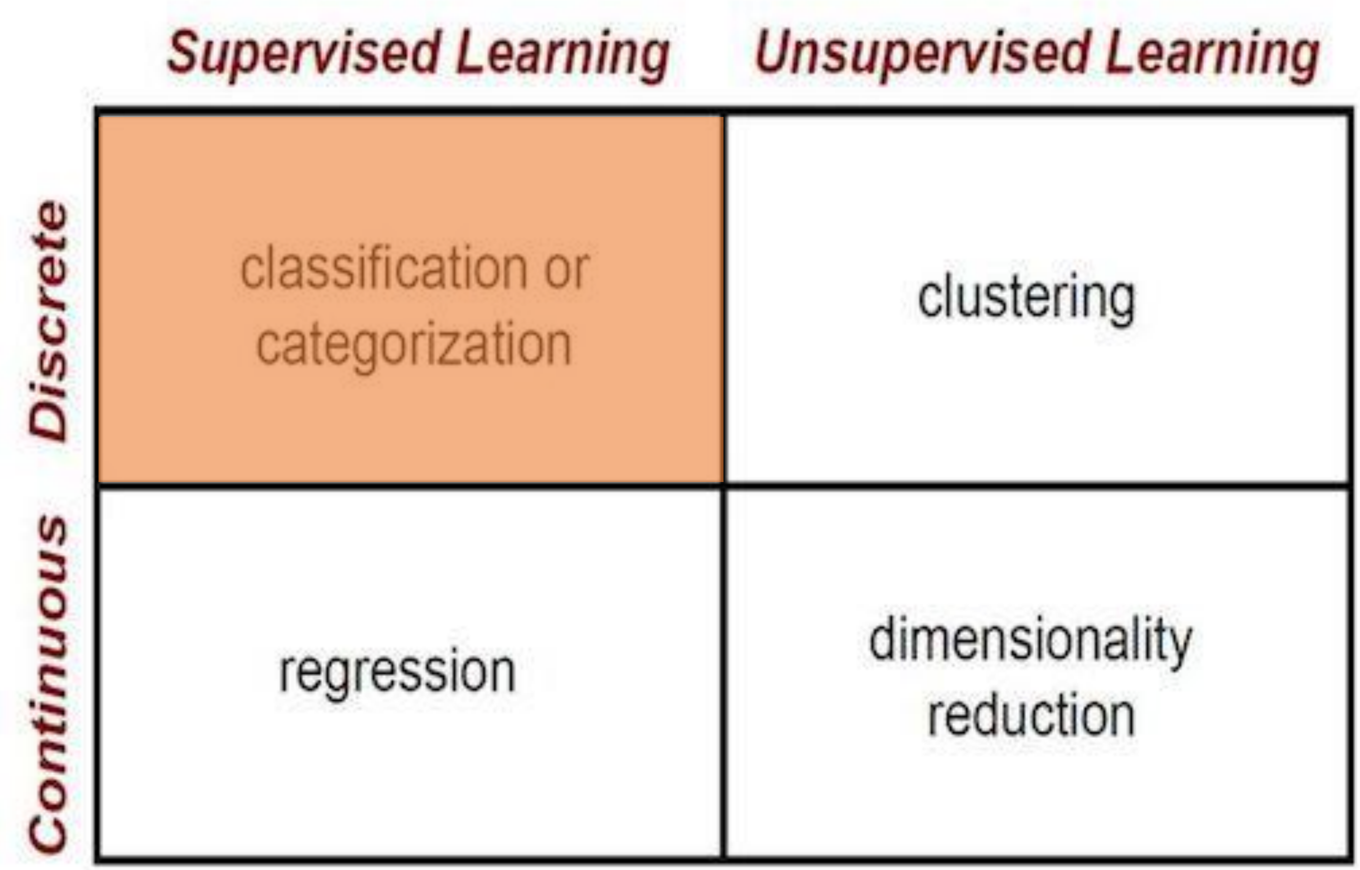


image from ratsgo.github.io

However


머신-러닝
여기 있는 사람들

Fields of ML


Supervised Learning
Unsupervised Learning
Discrete
Continuous
classification or categorization
clustering
regression
dimensionality reduction

We Will Focus On : Image Classification

** Live Classifier Running at :
http://cs231n.stanford.edu/

We Will Focus On : Image Classification

1. Train

강아지

강아지

강아지

고양이

고양이

고양이

We Will Focus On : Image Classification

2. Predict

강아지? 고양이?

... 고양이?!

We Will Focus On : Image Classification

2. Predict

강아지? 고양이?

... 고양이?!

Model

Making a Model

Model

Narrow Down to : Iris Classification

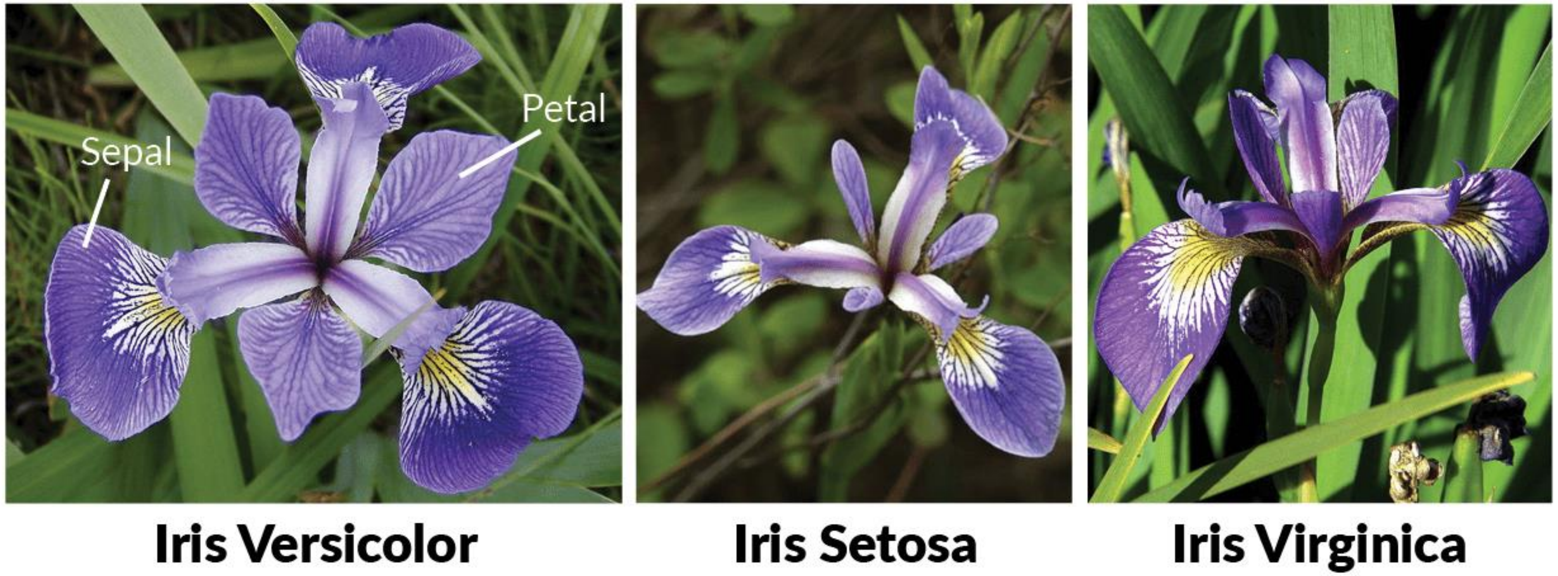
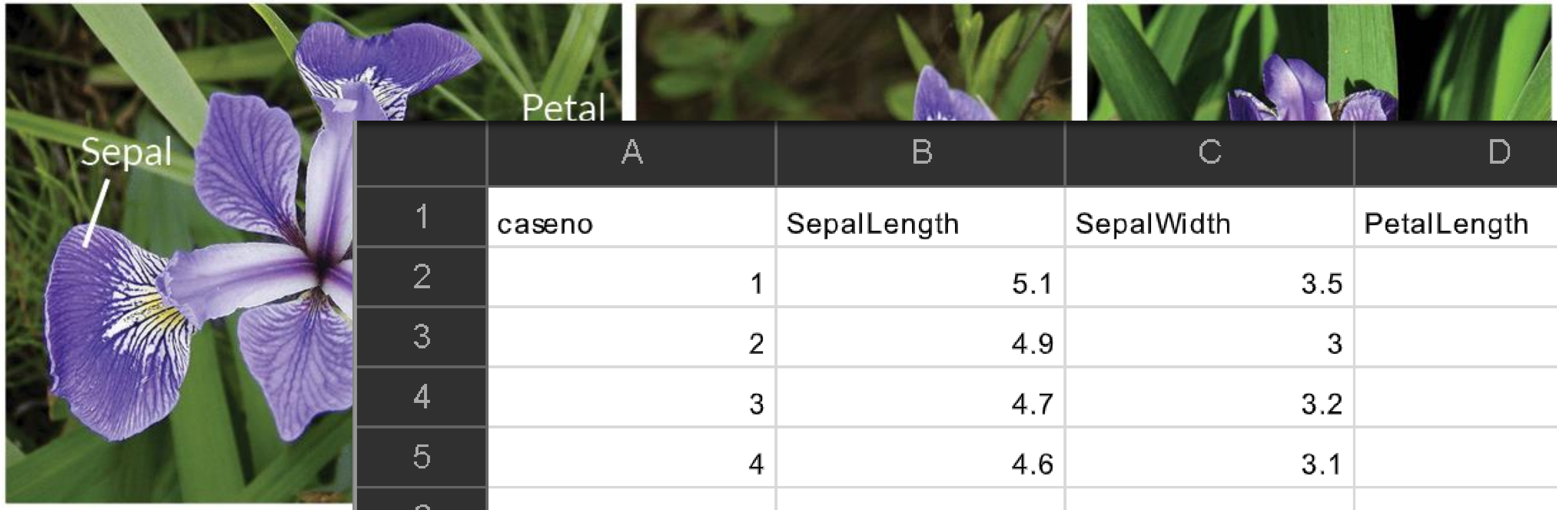


image from medium.com

Narrow Down to : Iris Classification



	A	B	C	D	E	F
1	caseno	SepalLength	SepalWidth	PetalLength	PetalWidth	Species
2	1	5.1	3.5	1.4	0.2	setosa
3	2	4.9	3	1.4	0.2	setosa
4	3	4.7	3.2	1.3	0.2	setosa
5	4	4.6	3.1	1.5	0.2	setosa
6	5	5	3.6	1.4	0.2	setosa
7	6	5.4	3.9	1.7	0.4	setosa
8	7	4.6	3.4	1.4	0.3	setosa
9	8	5	3.4	1.5	0.2	setosa
10	9	4.4	2.9	1.4	0.2	setosa
11	10	4.9	3.1	1.5	0.1	setosa

Iris Versi

Narrow Down to : Iris Classification

주어진 정보는, **Training Data : (Feature , Label)** 뿐!

목표는, **새로운 input feature에 대해 label을 predict하는 것!**

How…?

Making a Model : Data-Driven Approach

현재 가지고 있는 data를 통한 분류

Train : Memorize all training data : (feature, label)

Predict : Predict the label of the most similar training data

How do we define "similar"?

Nearest Neighbor

"similar" 판단 기준 = **distance btw feature vectors**

L1 distance) $d_1(x_1, x_2) = \sum_i |x_1^i - x_2^i|$

L2 distance) $d_2(x_1, x_2) = \sqrt{\sum_i (x_1^i - x_2^i)^2}$

Nearest Neighbor

"similar" 판단 기준 = **distance btw feature vectors**

L1 distance) $d_1(x_1, x_2) = \sum_i |x_1^i - x_2^i|$

L2 distance) $d_2(x_1, x_2) = \sqrt{\sum_i (x_1^i - x_2^i)^2}$

전체 dataset에 대해, distance가 가장 작은 label = prediction.

K-Nearest Neighbor

전체 dataset 중에서,

distance 작은 순서대로 K개 data 뽑은 다음,

majority를 차지하는 label로 predict

** Live K-NN running at : http://vision.stanford.edu/teaching/cs231n-demos/knn/

Testing a Model

Model을 열심히 만들었다.

그럼 이 Model을 Prediction에 바로 투입?!?!?

Testing a Model

Model을 열심히 만들었다.

그럼 이 Model을 Prediction에 바로 투입?!?!?

이 Model은 얼마나 정확한 Model인가?

이 Model의 Accuracy를 어떻게 신뢰할 것인가?

Testing a Model

1. Training set으로 model fit

2. Test set에서 performance 계산

Testing a Model

Model은 1개만 만들 수 있을까?

Testing a Model

Model은 1개만 만들 수 있을까?

No!

L1 distance / L2 distance

K = 1 / K = 2 / K = 3 / …

Testing a Model : Hyperparameter

choices about the algorithm that we set rather than learn

- L1 distance or L2 distance?
- K를 얼마로 설정할 것인가?
- ……

Model의 performance를 maximize하는 선택 하고 싶음. How?

Testing a Model

IDEA 1.

1. 여러가지 hyperparameter set으로 model들을 만들어 train시킴
2. Test set에서의 performance가 가장 높은 model을 선택

Testing a Model

IDEA 1.

1. 여러가지 hyperparameter set으로 model들을 만들어 train시킴
2. Test set에서의 performance가 가장 높은 model을 선택

However,

2에서의 performance를

unseen data에 대한 performance라 할 수는 없다.

Testing a Model

Test set은 model을 만드는 과정에 포함되어서는 안된다.

Test set의 존재 의의는, unseen data에 대한 performance를 얻기 위함.

model 1 → model 2

↓ ↓

test test

성능 별로네?

Testing a Model

IDEA 2.

1. 여러가지 hyperparameter set으로 model들을 만들어 train시킴
2. Validation set에서의 performance가 가장 높은 model을 선택
3. Test set로 unseen data에 대한 performance 계산

Testing a Model : Cross Validation

전체 dataset의 size가 작으면,

전과 같이 dataset을 나눴을 때 training set의 크기가 너무 작아진다.

Thus, use **Cross Validation**

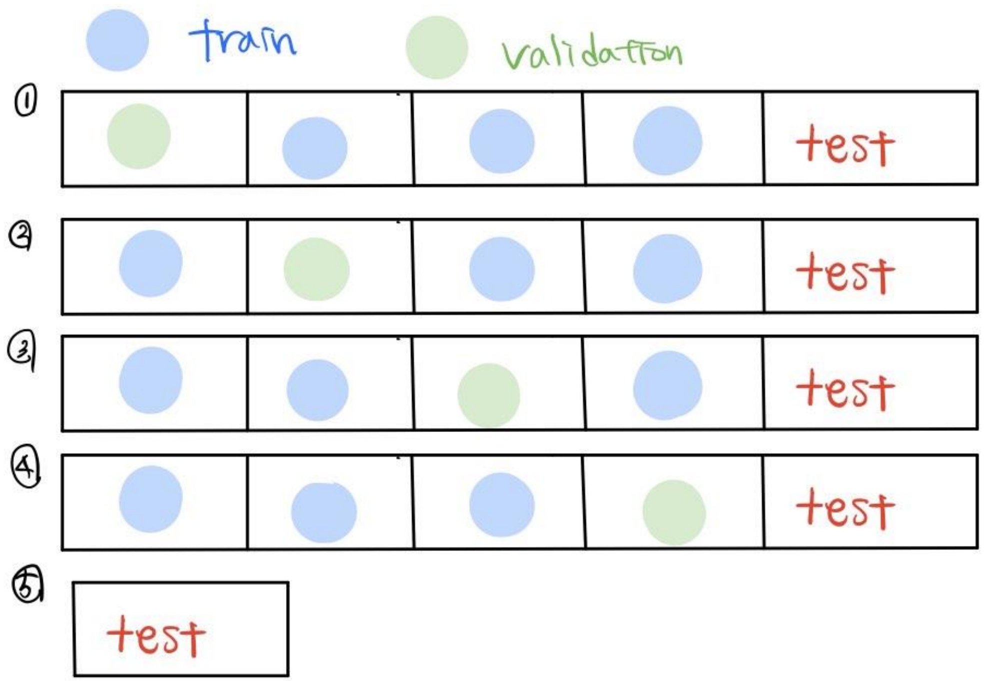
Testing a Model : Cross Validation

K-Fold Cross Validation

Test set을 제외한 나머지 부분을
k 조각(fold)로 나눈다.

한 iteration에서
조각 1개를 validation set으로 사용해,

총 k번의 iteration 후,
performance의 평균으로 hyper parameter 채택



Back to K-NN

전체 dataset 중에서,

distance 작은 순서대로 K개 data 뽑은 다음,

majority를 차지하는 label로 predict

Limitations of K-NN

1. Poor classification performance

image from Stanford cs231n

Limitations of K-NN

2. Poor prediction efficiency

K-NN은 train O(1), predict O(N)

in real world problems, we want train O(N), predict O(1)

Making a Model

Data-Driven Approach에서 벗어나,

꽃잎, 꽃받침의 길이 정보에서 꽃의 종의 특징 뽑아내기

How…?

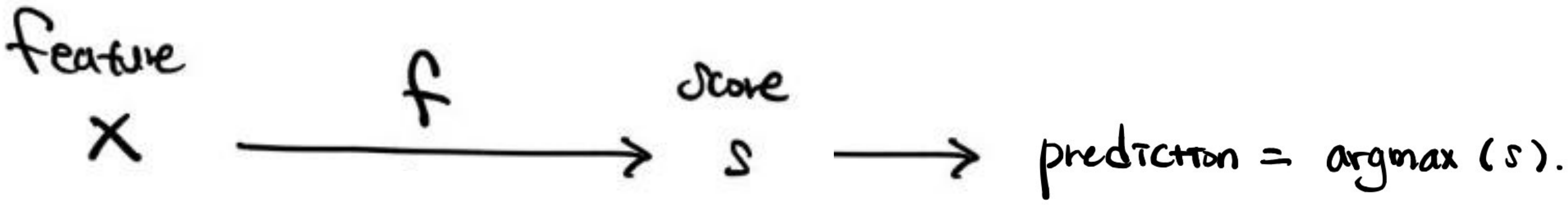
Making a Model : Parametric Approach

train data에서 뽑아낸 label별 특징들을 **parameter**에 저장

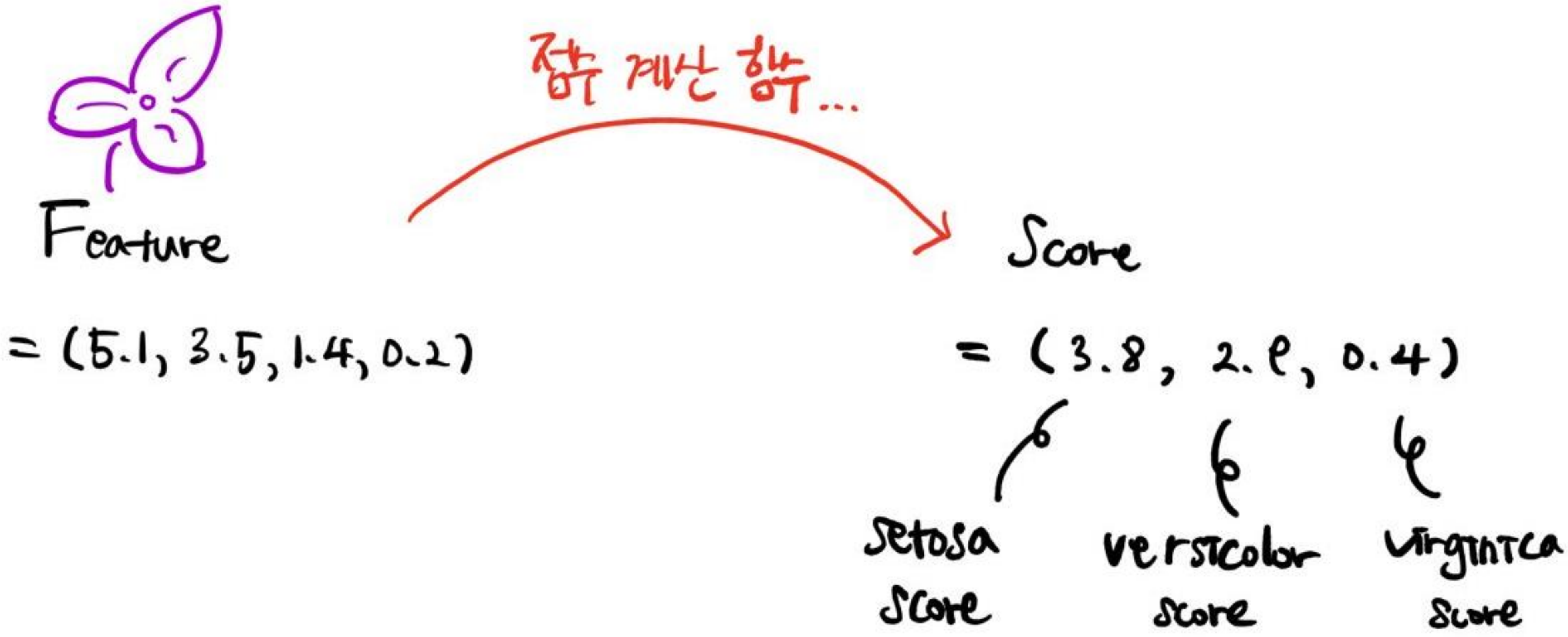
then, prediction에서는 **parameter**만 사용해 predict!

Linear Classifier : Idea

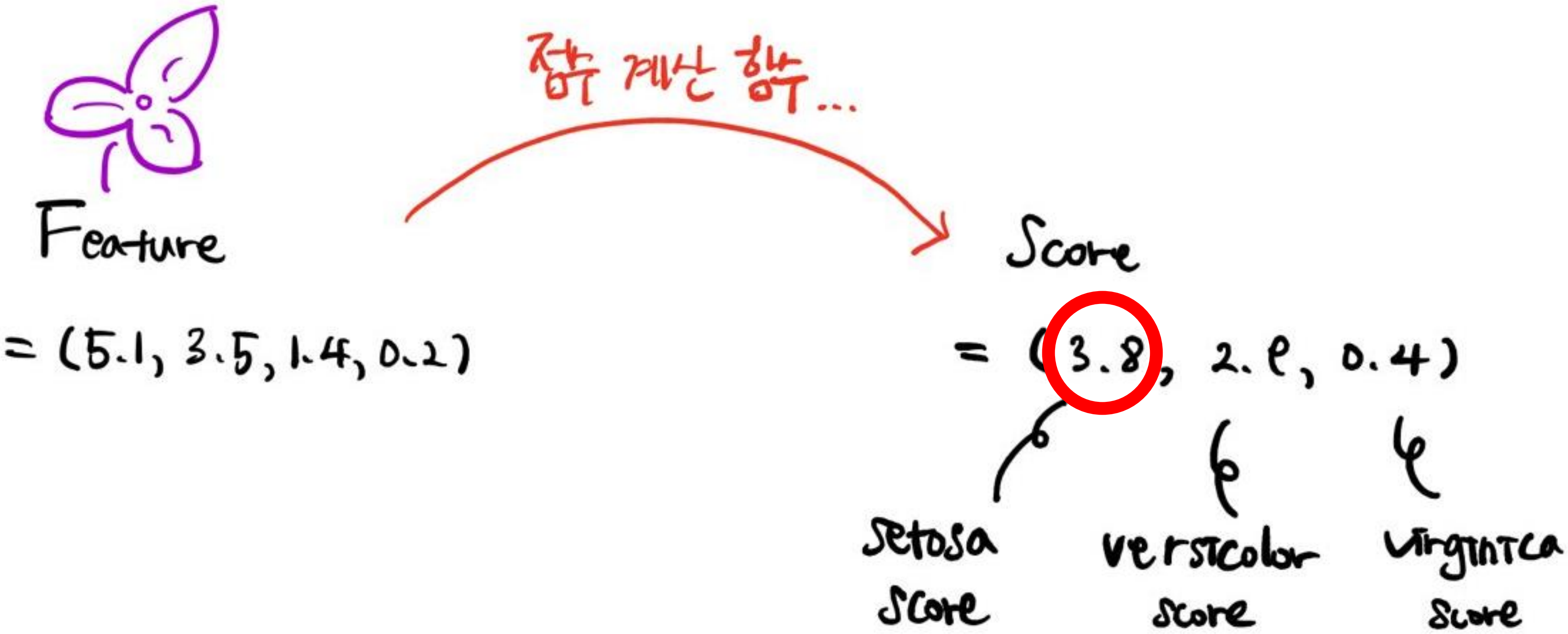
Feature를 각 Label에 대한 점수로 Mapping하자!



Linear Classifier : Score Function



Linear Classifier : Score Function



Linear Classifier : Idea

feature $x \xrightarrow{f}$ score $s \longrightarrow$ prediction = argmax(s).

점수 계산 함수 f는,

각 feature의 중요도를 반영하여 score 계산

** Setosa의 sepal width가 다른 종에 비해 길다면?

Linear Classifier : Idea

점수 계산 방법

: 각각의 Feature에 **가중치**를 곱해서 더하기!

feature : (x_0. x_1. x_2. x_3)

1. Setosa 에 대한 가중치 : (W_{00}. W_{01}. W_{02}. W_{03})

   ↓

   setosa score : $W_{00}x_0 + W_{01}x_1 + W_{02}x_2 + W_{03}x_3$

2. versicolor 에 대한 가중치 : (W_{10}. W_{11}. W_{12}. W_{13})

   ↓

   versicolor score : $W_{10}x_0 + W_{11}x_1 + W_{12}x_2 + W_{13}x_3$

3. virginica도 비슷한 방법으로.

Linear Classifier : Algebraic

feature $\quad$ score

$$x \xrightarrow[\varphi]{f} s$$

$$f(x) = Wx.$$

$(W : 3 \times 4,\ x : 4 \times 1).$

Linear Classifier : Algebraic

feature $\quad\quad\quad$ score

$$x \xrightarrow[\varphi]{f} s$$

$$f(x) = Wx + b$$

$(W : 3\times4,\ x : 4\times1)$.

Linear Classifier : Algebraic

feature		score

$$x \xrightarrow[\varphi]{f} s$$

$$f(x) = Wx + b$$

Bias

$(W : 3 \times 4,\ x : 4 \times 1)$.

Weight

Linear Classifier : Algebraic

	A	B	C	D	E
1	caseno	SepalLength	SepalWidth	PetalLength	PetalWidth
2	1	4.8	3	1.4	0.3

$$\begin{pmatrix} 1.063 & 2.179 & -1.557 & -0.611 \\ 0.952 & 0.259 & 0.532 & -0.182 \\ -0.851 & -1.12 & 2.865 & 2.254 \end{pmatrix} \begin{pmatrix} 4.8 \\ 3 \\ 1.4 \\ 0.3 \end{pmatrix} + \begin{pmatrix} 0.262 \\ 0.495 \\ 0.294 \end{pmatrix} = \begin{pmatrix} \boxed{9.5383} \\ 6.5318 \\ -2.4636 \end{pmatrix}$$

Linear Classifier : Algebraic

	A	B	C	D	E	F
1	caseno	SepalLength	SepalWidth	PetalLength	PetalWidth	Species
2	1	4.8	3	1.4	0.3	setosa

$$\begin{pmatrix} 1.063 & 2.179 & -1.557 & -0.611 \\ 0.952 & 0.259 & 0.532 & -0.182 \\ -0.851 & -1.12 & 2.865 & 2.254 \end{pmatrix} \begin{pmatrix} 4.8 \\ 3 \\ 1.4 \\ 0.3 \end{pmatrix} + \begin{pmatrix} 0.262 \\ 0.495 \\ 0.294 \end{pmatrix} = \begin{pmatrix} \boxed{9.5383} \\ 6.5318 \\ -2.4636 \end{pmatrix}$$

Linear Classifier : Geometric

Why is it a **"Linear"** Classifier?

Linear Classifier : Geometric

$$\begin{pmatrix} 1.063 & 2.179 & -1.557 & -0.611 \\ 0.952 & 0.259 & 0.532 & -0.182 \\ -0.851 & -1.12 & 2.865 & 2.254 \end{pmatrix} \begin{pmatrix} x0 \\ x1 \\ x2 \\ x3 \end{pmatrix} + \begin{pmatrix} 0.262 \\ 0.495 \\ 0.294 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

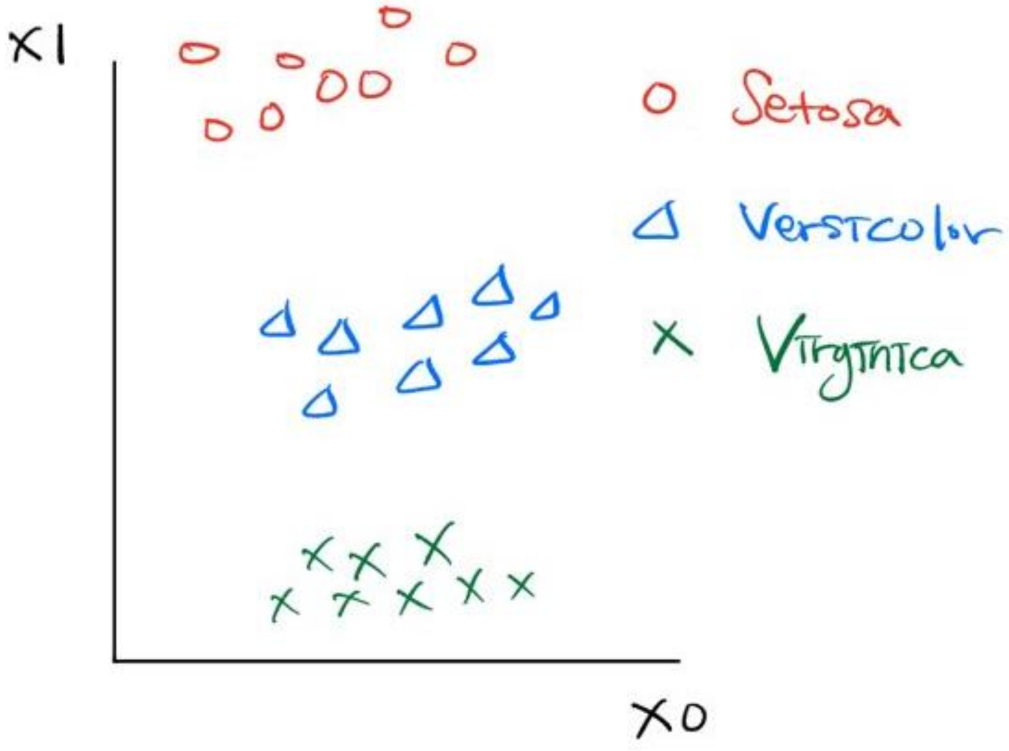
Linear Classifier : Geometric

$$\begin{pmatrix} 1.063 & 2.179 & -1.557 & -0.611 \\ 0.952 & 0.259 & 0.532 & -0.182 \\ -0.851 & -1.12 & 2.865 & 2.254 \end{pmatrix} \begin{pmatrix} x0 \\ x1 \\ x2 \\ x3 \end{pmatrix} + \begin{pmatrix} 0.262 \\ 0.495 \\ 0.294 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

1.063 * x0 + 2.179 * x1 – 1.577 * x2 – 0.611 * x3 + 0.262 = 0

는, Setosa를 표현하는 **Hyperplane**

Linear Classifier : Geometric



각각의 data는 feature가 4개이므로,

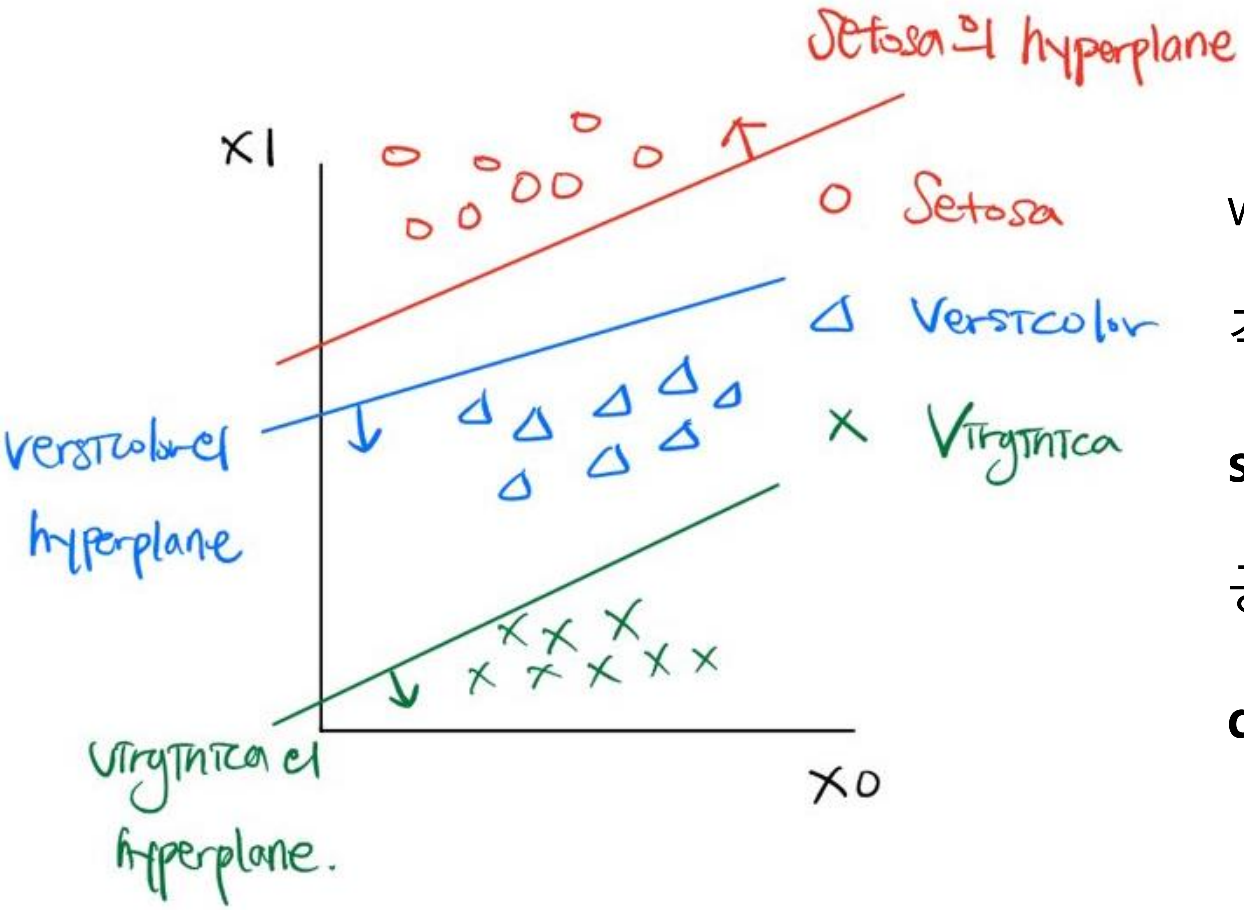
4차원 공간상의 한 점이다.

또, 같은 종의 꽃들은

서로 비슷한 곳에 위치할 것이다.

(4차원 공간을 visualize할 수 없으므로,
왼쪽에서는 2차원으로 줄여서 표현함)

Linear Classifier : Geometric



weight matrix **W**의 각 row는

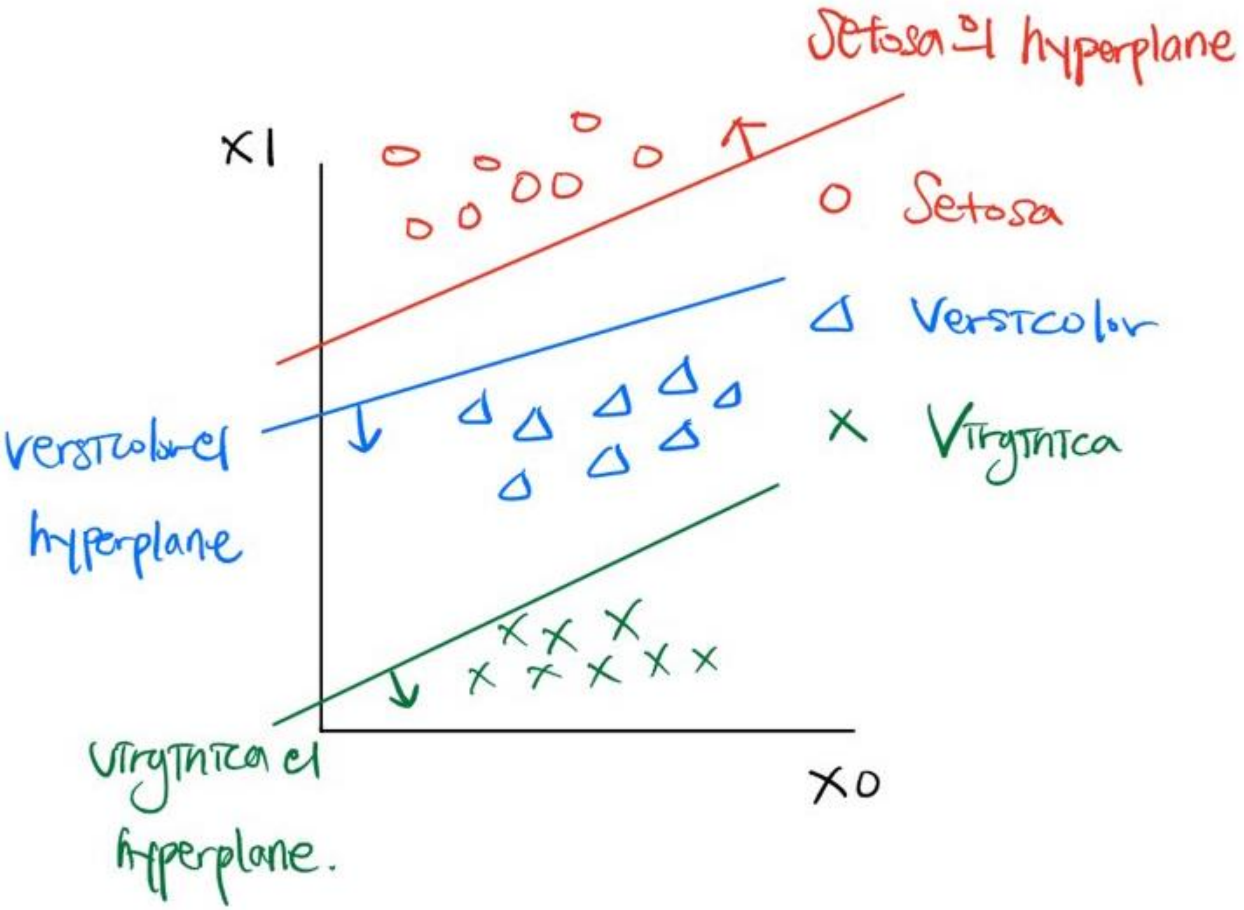
각 label의 **score function**이다.

score function = 0 으로 만든 **hyperplane**은

공간상의 label들을 classify하는

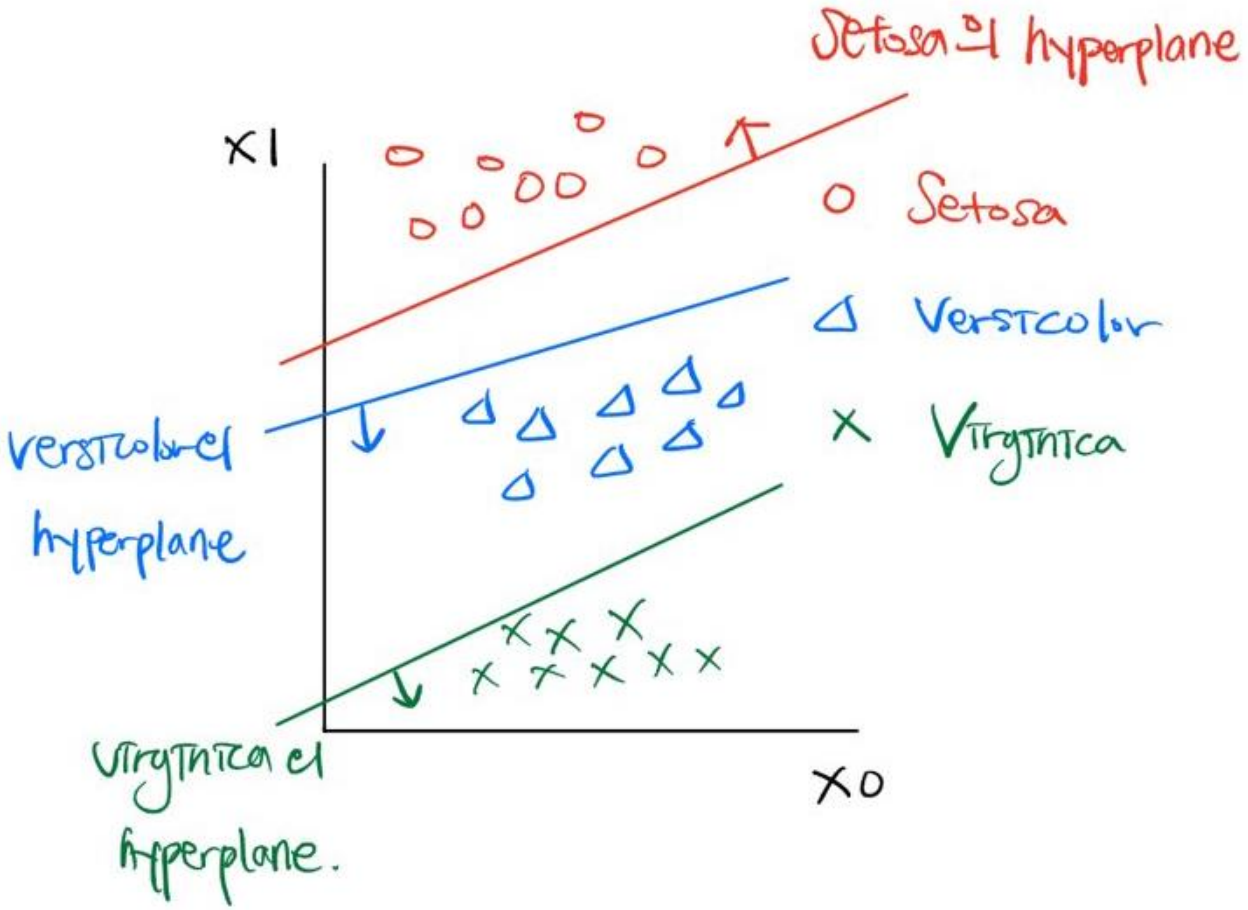
decision boundary가 된다.

Linear Classifier : Geometric



then, bias가 왜 필요할까?

Linear Classifier : Geometric



then, bias가 왜 필요할까?

bias **B**가 없었다면,

모든 hyperplane이 원점을 지나야 함.

따라서, classify가 잘 안됨!

However,

In reality, the correct parameter W is not given from the start

따라서, **(current)** incorrect param. → **(objective)** correct param.

How should we optimize the model?

** Live Linear Classifier Running at :

http://vision.stanford.edu/teaching/cs231n-demos/linear-classify/

Questions on Model Optimization

- **Train이 잘 되었는지 판단할 수치적 척도 필요**

: define a **Loss Function** that quantifies our unhappiness with the scores across the training data

- **Parameter를 update하는 algorithm 필요**

: come up with a way of efficiently finding the parameters that minimize the **Loss Function**

Review

1. **What is Machine Learning?**
    - Definition
    - Fields of ML
    - Narrow down to Image Classification
2. **Making a Model I**
    - Narrow down to Iris Classification
    - Data-Driven Approach
        - NN, K-NN Algorithms
3. **How to Test a Model**
    - Hyper parameter
    - Cross Validation
4. **Making a Model II**
    - Score Function
    - Linear Classifier : Algebraic & Geometric

Preview on Next Lecture

Questions

- **Train이 잘 되었는지 판단할 수치적 척도 필요**

: define a **Loss Function** that quantifies our unhappiness with the scores across the training data

- **Parameter를 update하는 algorithm 필요**

: come up with a way of efficiently finding the parameters that minimize the **Loss Function**

More on Linear Classifier

How to calculate the Loss

21