

Review

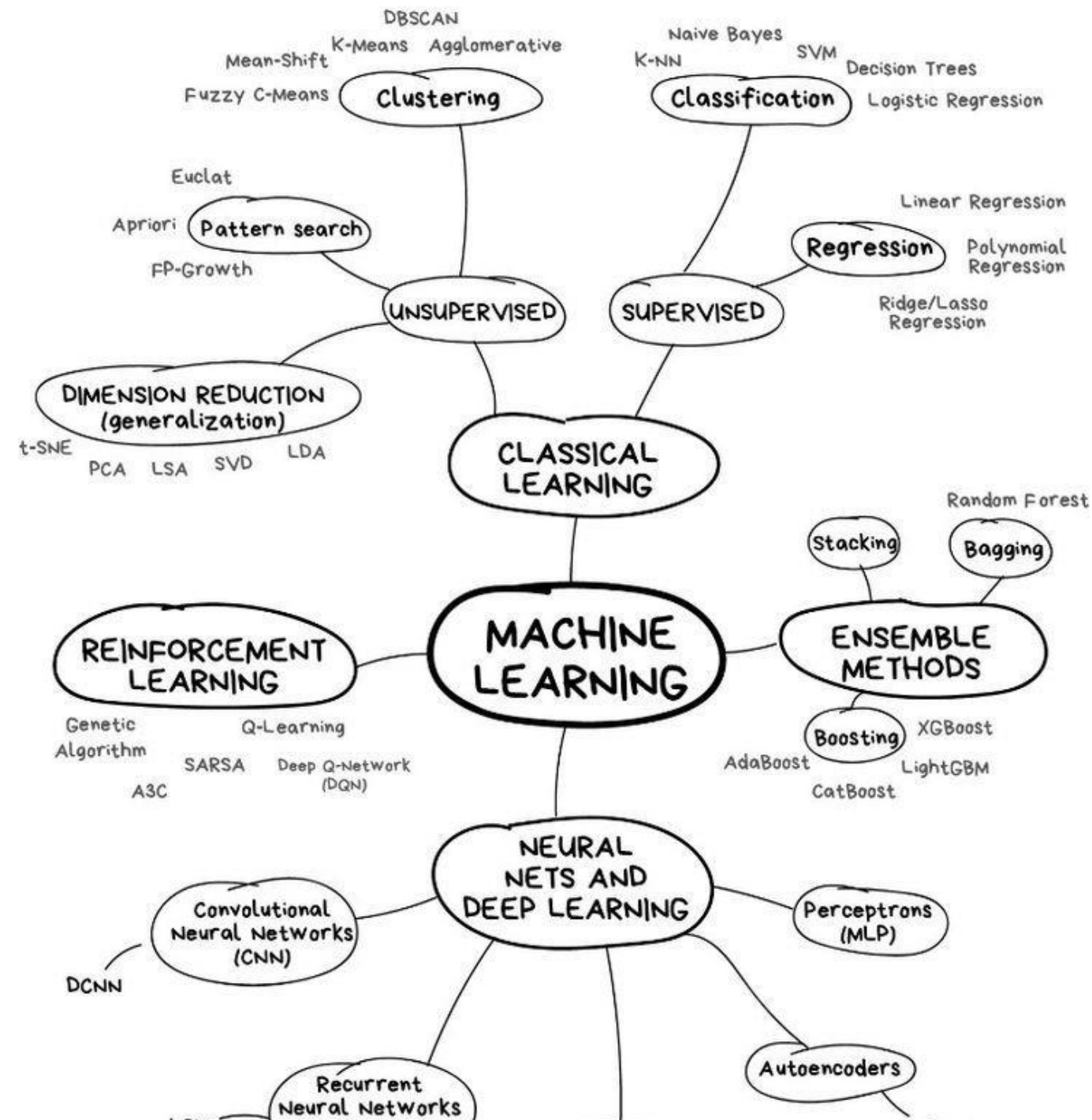


한양대학교 ERICA
소프트웨어융합대학
COLLEGE OF COMPUTING

인공지능학과
Department of
Artificial Intelligence

정 우 환 (whjung@hanyang.ac.kr)

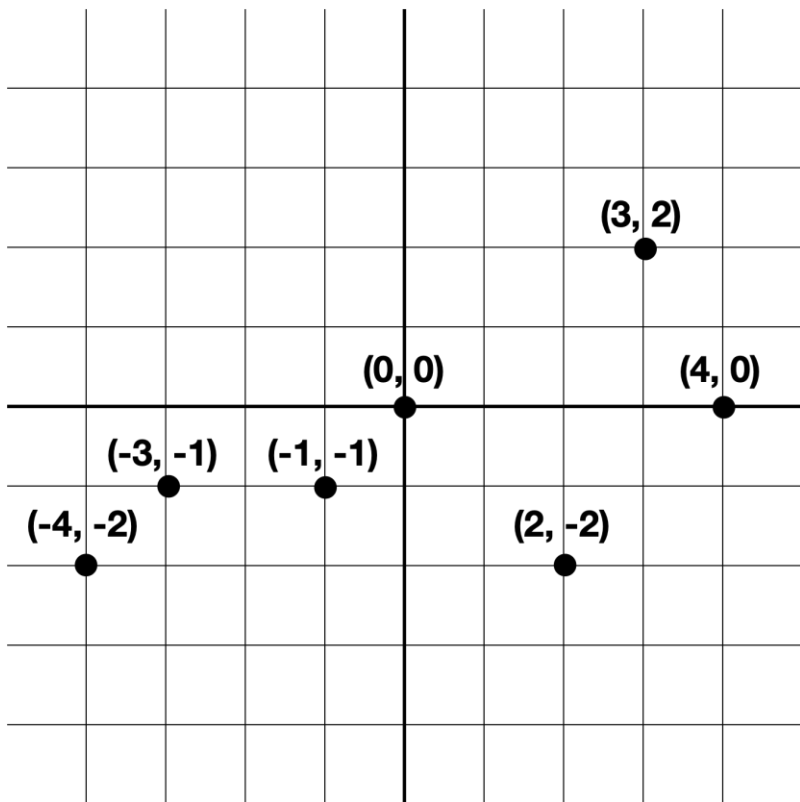
Fall 2021



- Linear Regression
- Logistic Regression
- K-NN
- Naïve Bayes
- Decision Trees
- SVM
- MLP
- K-means
- DBSCAN
- Apriori
- PCA
- t-SNE
- Q-Learning

- Machine learning basics
- Pytorch

Clustering



- K-means clustering (K = 3) 알고리즘을 사용 했을때의 클러스터링 과정과 결과를 제시하시오

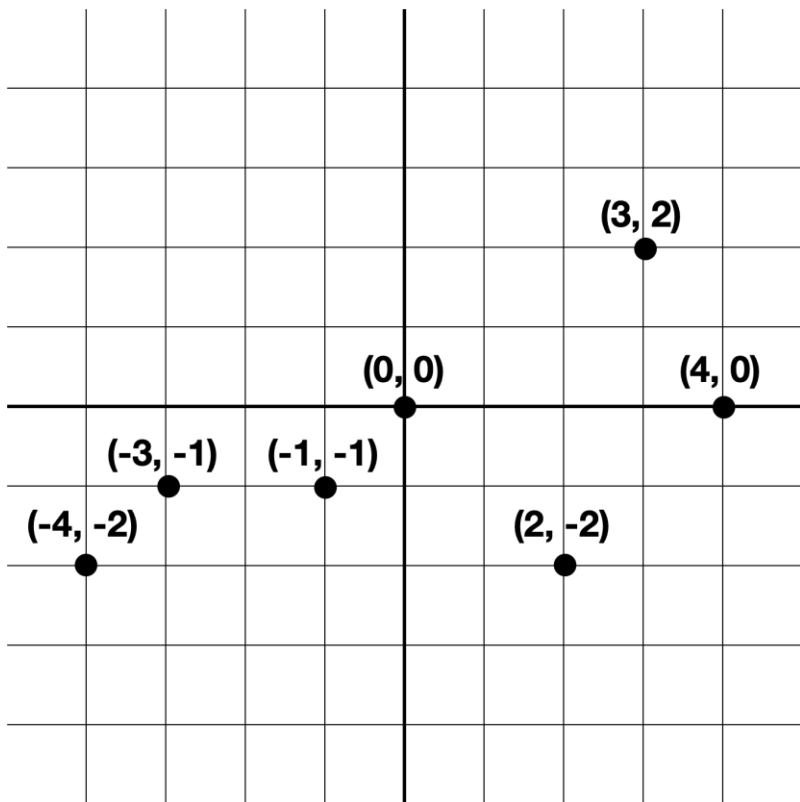
초기 클러스터

A: (-1, -1), (-4, -2)

B: (-3, -1), (2, -2)

C: (0, 0), (3, 2), (4, 0)

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$



초기 클러스터

A: (-1, -1), (-4, -2)

B: (-3, -1), (2, -2)

C: (0, 0), (3, 2), (4, 0)

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

■ K-means clustering (K = 3) 알고리즘을 사용 했을때의 클러스터링 과정과 결과를 제시하시오

- 초기 클러스터의 센터는 각각 $(-2.5, -1.5)$, $(-0.5, -1.5)$, $(7/3, 2/3)$
- 센터에 가까운 클러스터에 데이터를 재할당하면 다음과 같은 새로운 그룹이 만들어진다.
 - A : $(-4, -2)$, $(-3, -1)$, B : $(-1, -1)$, $(0, 0)$, $(2, -2)$, C: $(3, 2)$, $(4, 0)$
- 다시 클러스터의 센터를 구하면 각각 $(-3.5, -1.5)$, $(1/3, -1.0)$, $(3.5, 1.0)$
- 센터에 가까운 클러스터에 데이터를 재할당
 - A : $(-4, -2)$, $(-3, -1)$, B : $(-1, -1)$, $(0, 0)$, $(2, -2)$, C: $(3, 2)$, $(4, 0)$
- 클러스터에 변화가 없었기 때문에 클러스터링을 종료한다.

Machine learning basics

Overfitting

- Overfitting 이란?
 - 학습 데이터에 대해 과하게 학습되어 학습 데이터에 대해선 성능이 좋지만, 학습 데이터 이외의 데이터에 대해서는 성능이 좋지 못한 상황
- 방지법
 - Decision tree
 - Pruning
 - Neural nets
 - Regularization
 - Iterative algorithms
 - Early stopping
 - ...

Precision, Recall, F1 score

		Predicted class	
		True	False
Actual Class	True	TP	FN
	False	FP	TN

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

PyTorch

■ Question

- Pytorch Tensor X는 아래와 같다.
 - X : tensor([[2, 0], [3, 1]]) Size : [2, 2]
- 다음 명령어를 수행했을때 out1~out6에 저장된 Tensor를 쓰시오

```
out1 = X.T  
out2 = X.view(1, -1)
```

■ Answer

```
Out1: [[2, 3], [0, 1]]  
Out2: [[2, 0, 3, 1]]
```

Decision Tree

$$\log_2 3 \approx 1.58$$

Decision Tree

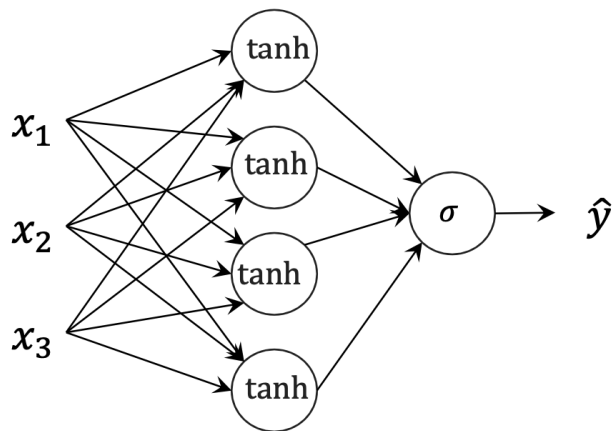
- Question) $\text{Info}_{\text{price}}(D) = ?$

price	maintenance	safety	decision
low	high	low	negative
high	med	low	negative
low	low	high	positive
low	high	high	positive

$$\text{Info}_{\text{price}}(D) = \frac{1}{4}I(0,1) + \frac{3}{4}I(2,1) = \frac{3}{4} * \left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right) = \frac{3}{4} * \left(-\frac{2}{3} + \log_2 3\right) \approx 0.69$$

Neural Network

Shallow Neural Network



Parameters : $\theta = \{W^{[1]}, b^{[1]}, w^{[2]}, b^{[2]}\}$

Loss : $L(\hat{y}, y)$

$$z^{[1]} = W^{[1]}x + b^{[1]}$$

$$a^{[1]} = \tanh z^{[1]}$$

$$z^{[2]} = w^{[2]T} a^{[1]} + b^{[2]}$$

$$\hat{y} = a^{[2]} = \sigma(z^{[2]})$$

■ Question: $\frac{\partial L(a^{[2]}, y)}{\partial b^{[2]}} = ?$

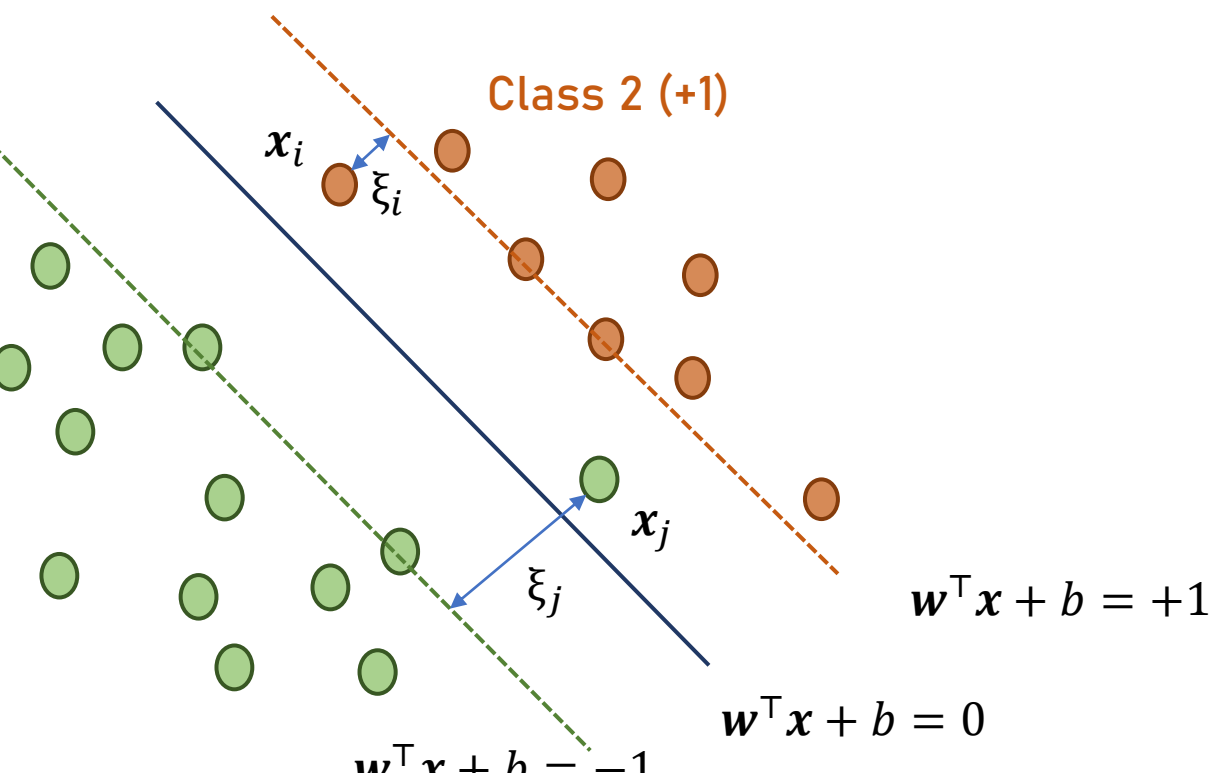
■ 참고: $\frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$, $\frac{d \tanh x}{dx} = 1 - \tanh^2 x$

SVM

- Linearly separable한 문제에 대한 SVM formulation이 아래와 같다. 이때, soft margin을 추가한 버전에 대한 문제정의를 작성하시오

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|_2^2$$

$$\text{subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, n$$



$$\text{minimize}_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n \xi_i$$

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i \quad \text{For } i = 1, 2, \dots, n$$

$$\xi_i \geq 0 \quad \text{For } i = 1, 2, \dots, n$$