Graph-Based Semi-Supervised Learning

Jiyoon Lee

2021 . 03 . 19

Introduction

발표자 소개

❖ **이지윤 (Jiyoon Lee)**

- Data Mining & Quality Analytics Lab
- Ph.D. Candidates (2018.03 ~ Present)

❖ **Research Interest**

- **Explainable neural network** using **Attention mechanism & Bayesian neural network**
- **Graph-based semi-supervised learning** using **Label propagation**

❖ **Contact**

- Tel: +82-2-3290-3769
- E-mail: jiyoonlee@korea.ac.kr

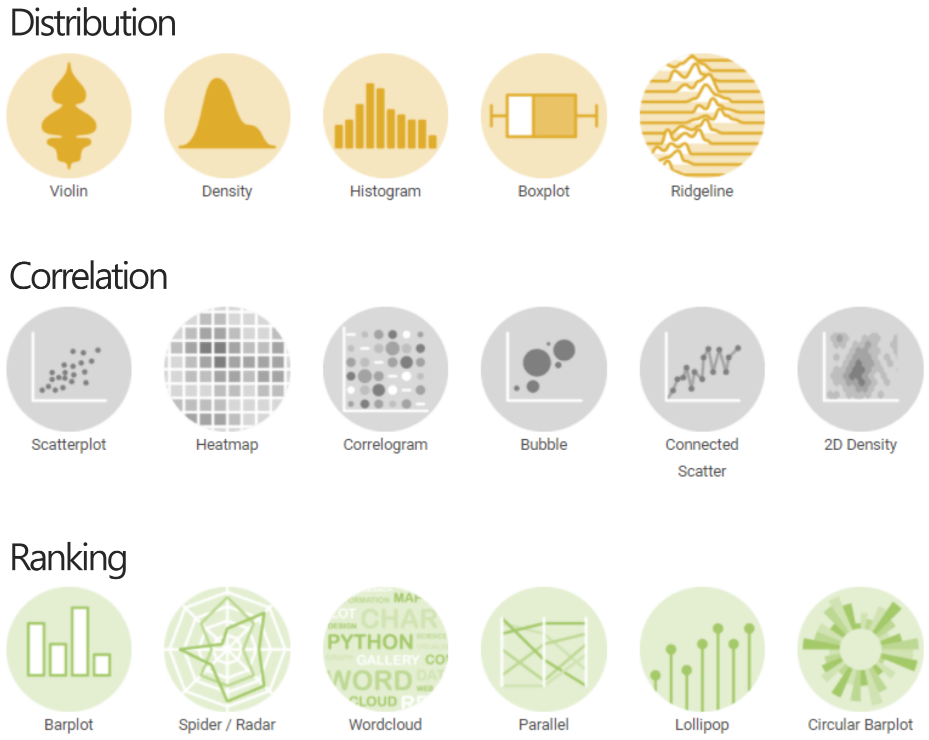
Contents

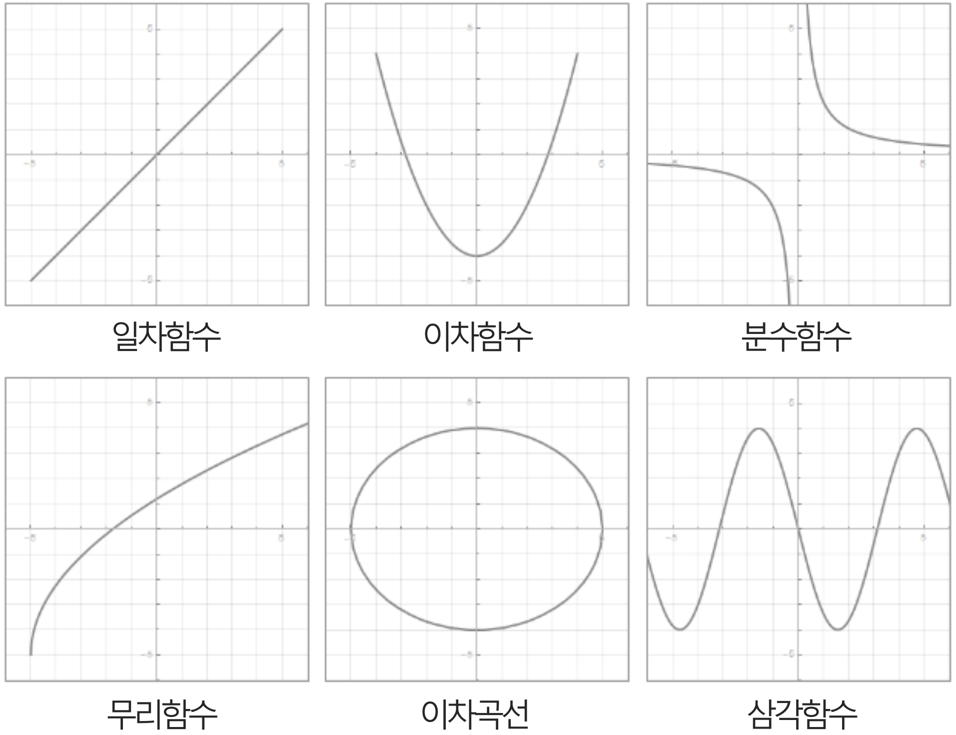
Graph-Based Semi-Supervised Learning

Graph Data

Graph definition

- ❖ 직선, 곡선, 도형 등 그래픽의 요소에 의해 **시각화 된 차트**를 의미
- ❖ 함수의 그래프는 **주어진 함수가 나타내는 직선이나 곡선**을 의미



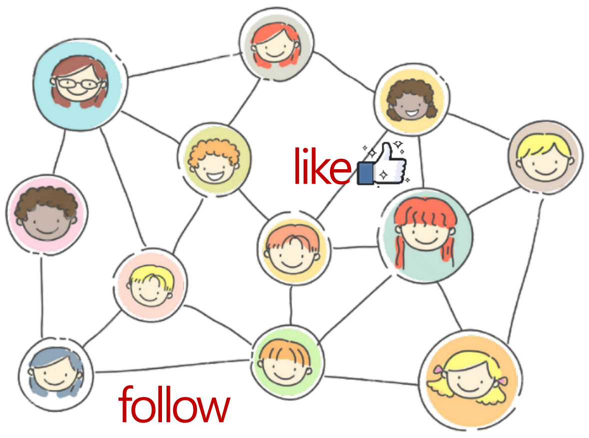


Graph Data

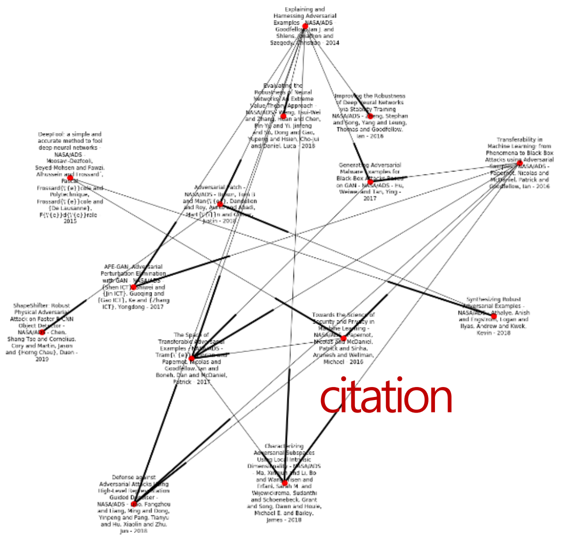
Graph for Data Analysis

❖ **데이터**와 데이터 사이의 **관계**를 모아 놓은 자료

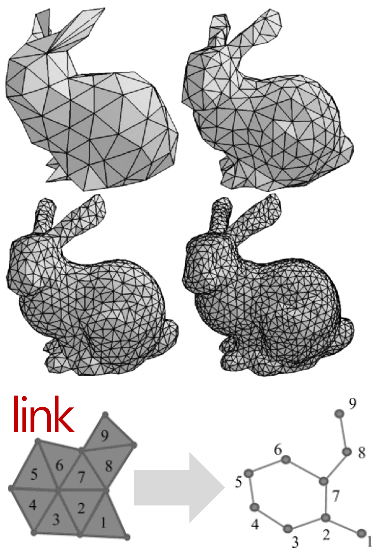
Social Graph



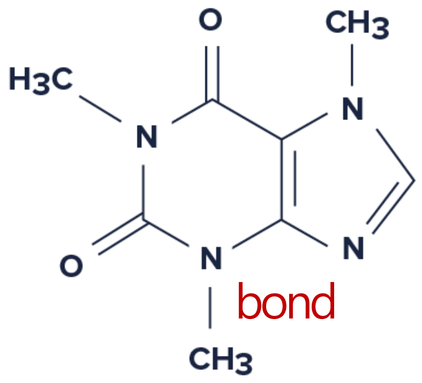
Citation Graph



3D Mesh Graph



Molecular Graph

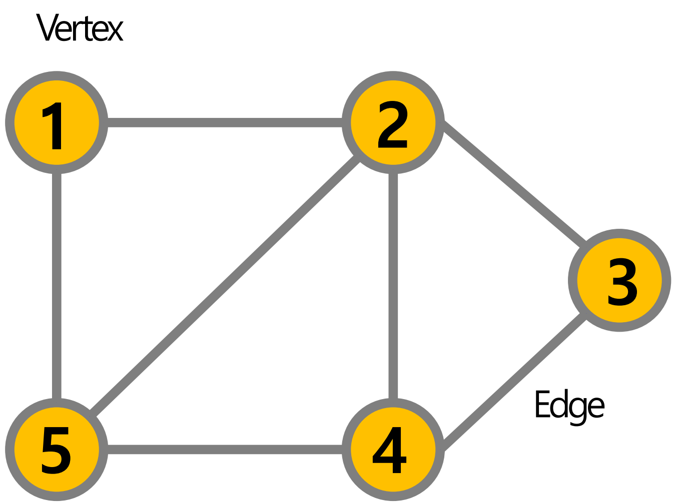


Graph Data

Graph for Data Analysis

❖ **데이터**와 데이터 사이의 **관계**를 모아 놓은 자료
Vertex (Node)　　　　Edge

❖ 그래프는 Vertices와 그사이 Edge의 집합: $\boldsymbol{G} = \{\boldsymbol{V}, \boldsymbol{E}\}$

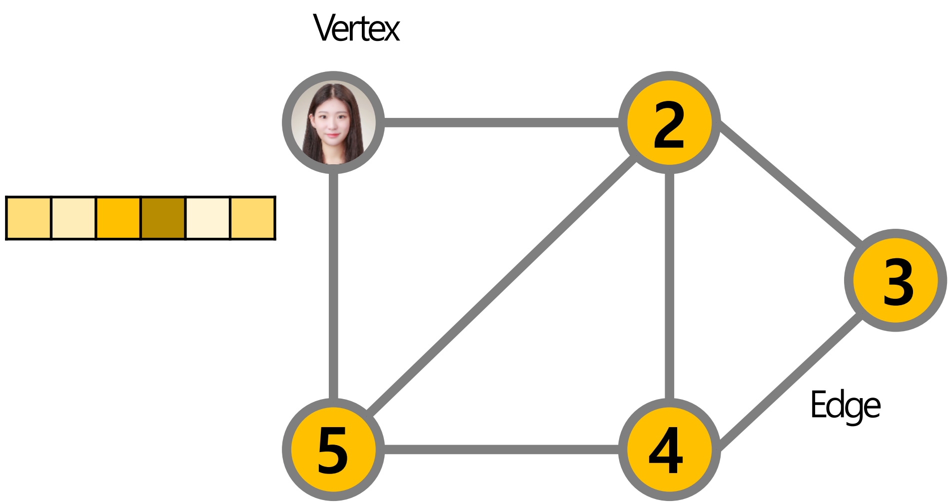


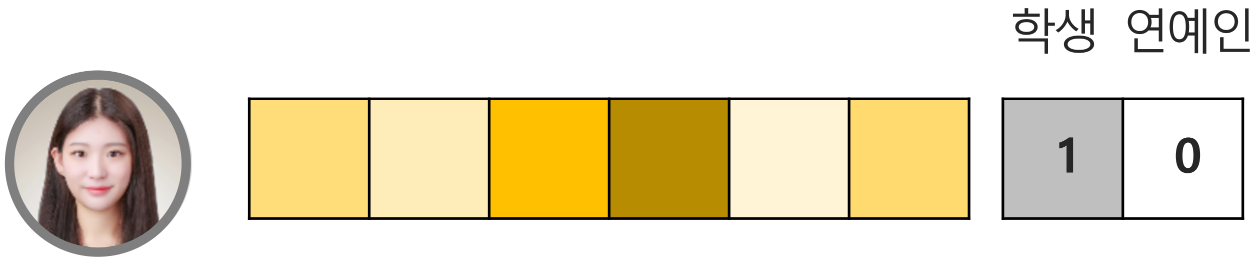
Graph Data

Graph for Data Analysis

❖ Vertex는 하나의 샘플을 의미

❖ Vertex set: $V(G) = \{v_1, v_2, \ldots, v_n\}$



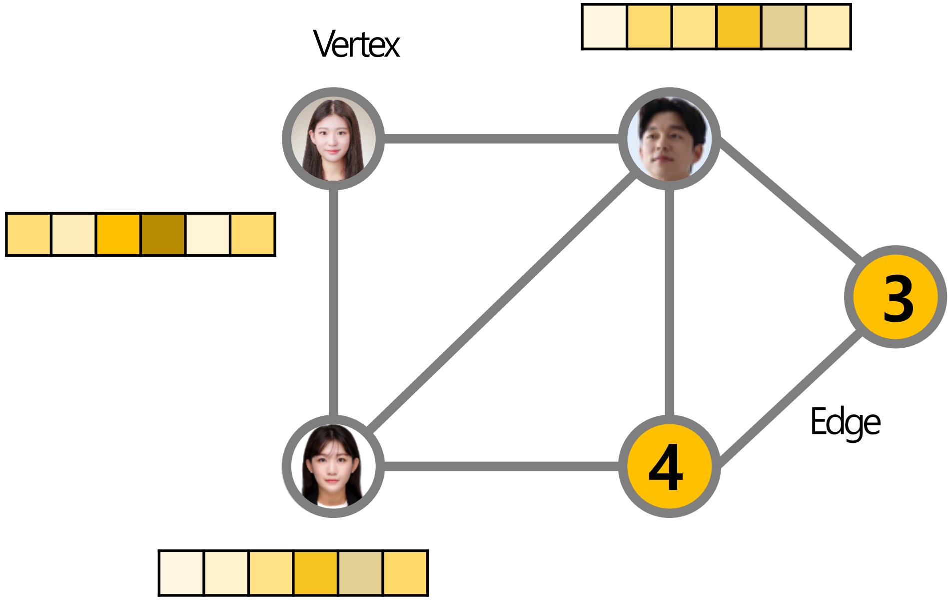


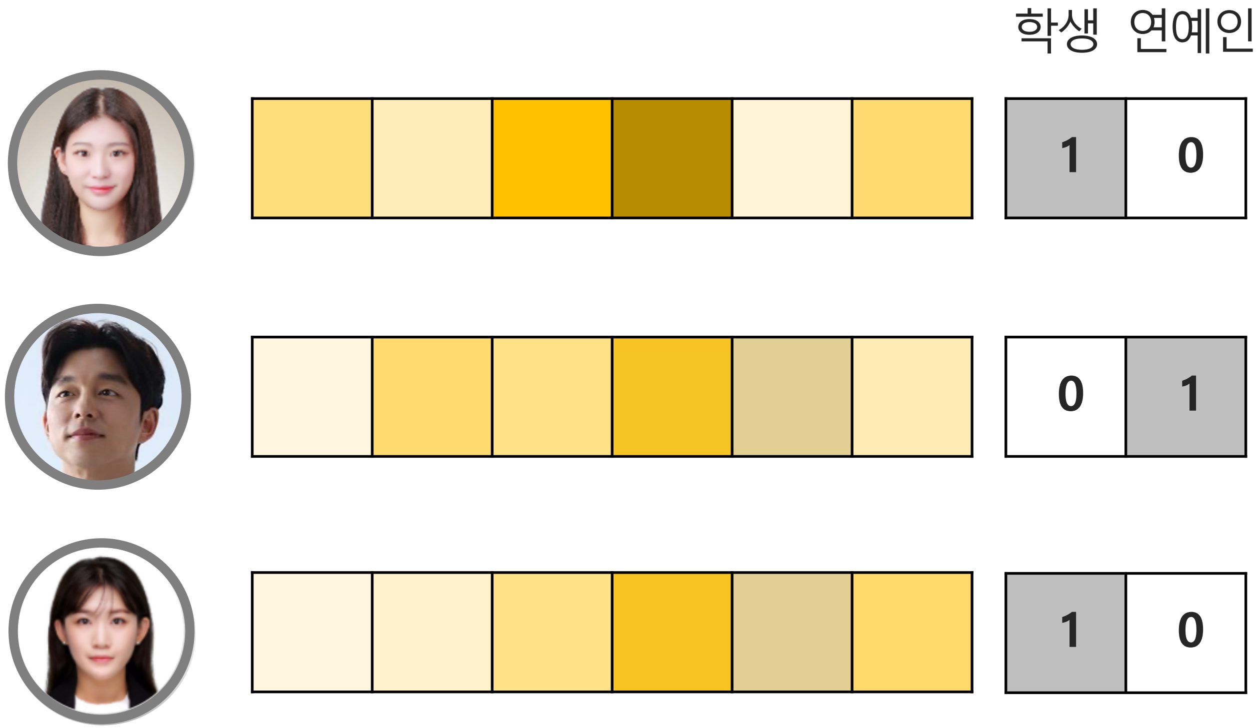
Graph Data

Graph for Data Analysis

❖ Vertex는 하나의 샘플을 의미

❖ Vertex set: $V(G) = \{v_1, v_2, \ldots, v_n\}$



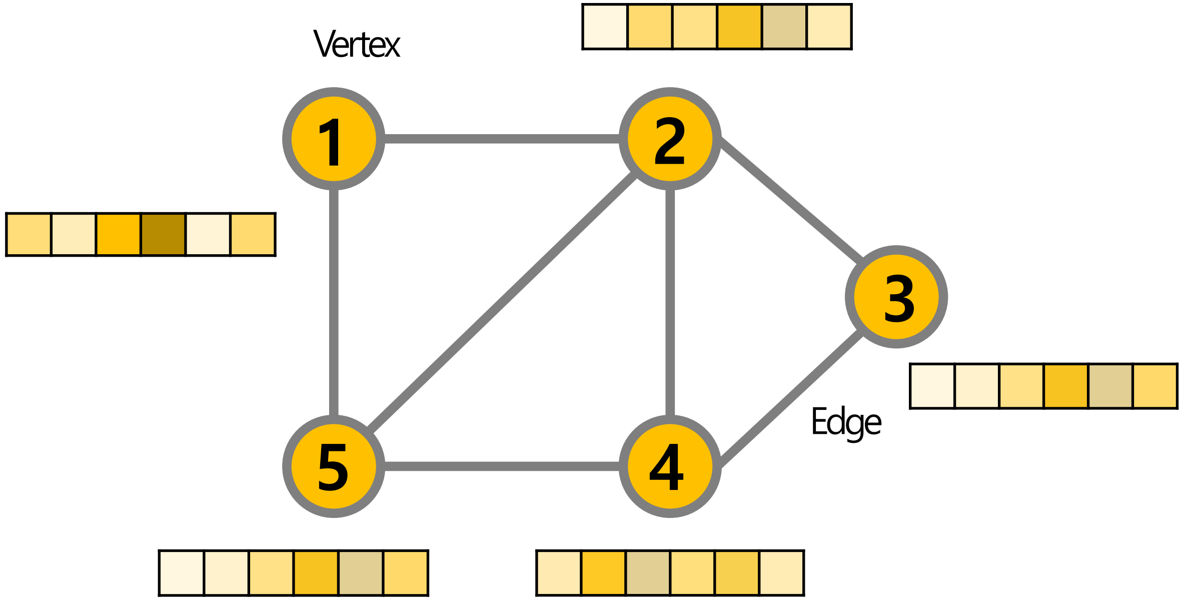


Graph Data

Graph for Data Analysis

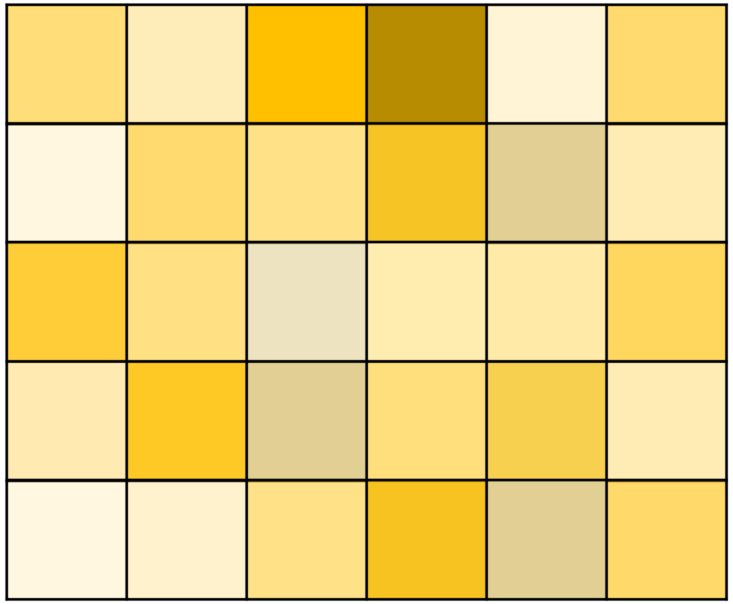
❖ Vertex는 하나의 샘플을 의미

❖ Node-feature matrix와 Node-class matrix로 표현



Node – Feature Matrix

$X \in R^{n \times f}$



Node – Class Matrix

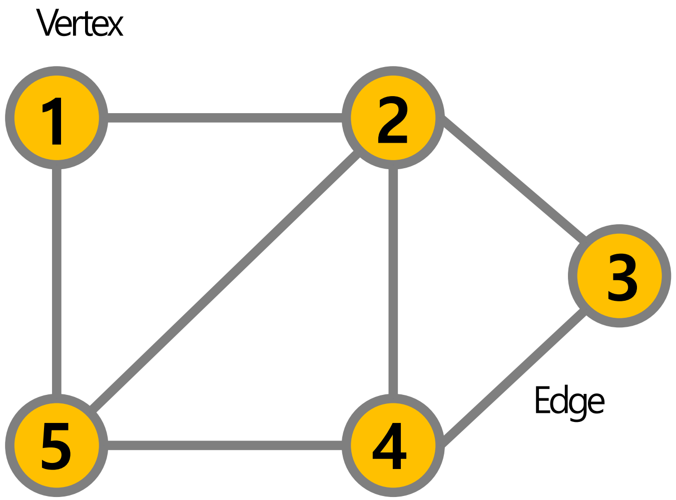
$Y \in R^{n \times c}$

1	0
0	1
0	1
1	0
1	0

Graph Data

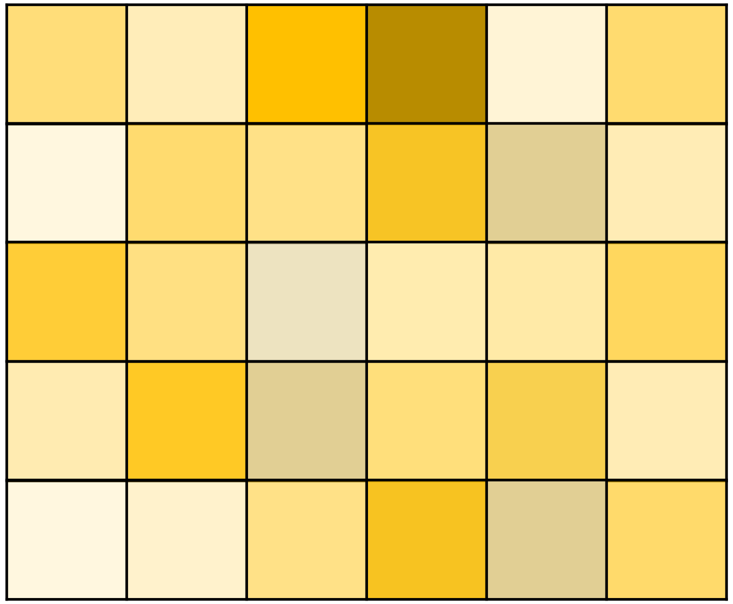
Graph for Data Analysis

- ❖ Edge는 샘플들의 연결관계를 의미
- ❖ Edge set: $\boldsymbol{E}(\boldsymbol{G}) = \{\boldsymbol{w}_{ij}\},\ \mathbf{1} \leq \boldsymbol{i},\ \boldsymbol{j} \leq \boldsymbol{N}$



Node – Feature Matrix

$X \in R^{n \times f}$



Node – Class Matrix

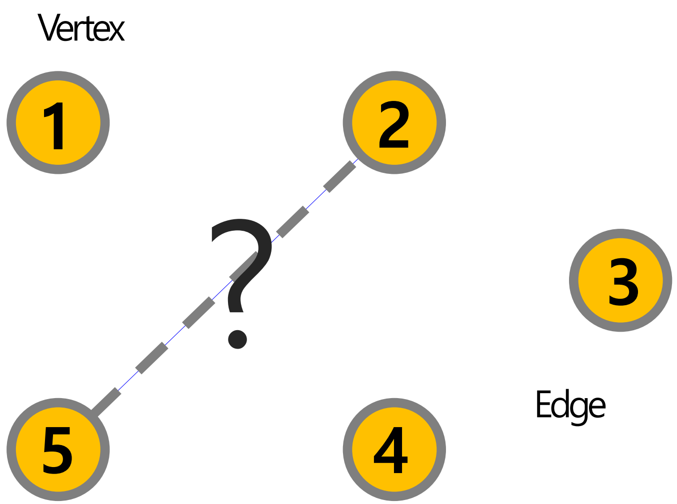
$Y \in R^{n \times c}$

1	0
0	1
0	1
1	0
1	0

Graph Data

Graph for Data Analysis

- ❖ Edge에 대한 정보가 주어지지 않는다면, vertex를 바탕으로 edge를 구성 해야함
- ❖ Edge set: $\boldsymbol{E}(\boldsymbol{G}) = \{\boldsymbol{w_{ij}}\}\ ,\ \boldsymbol{1} \le \boldsymbol{i},\ \boldsymbol{j} \le \boldsymbol{N}$



ϵ – Nearest Neighbor

$$w_{ij} = 1, \quad if\ \|x_i - x_j\| \le \epsilon$$
$$w_{ij} = 0, \quad if\ otherwise$$

K – Nearest Neighbor

$$w_{ij} = 1, \quad if\ x_i\ or\ x_j\ is\ k\ nearest\ of\ the\ other.$$
$$w_{ij} = 0, \quad if\ otherwise$$

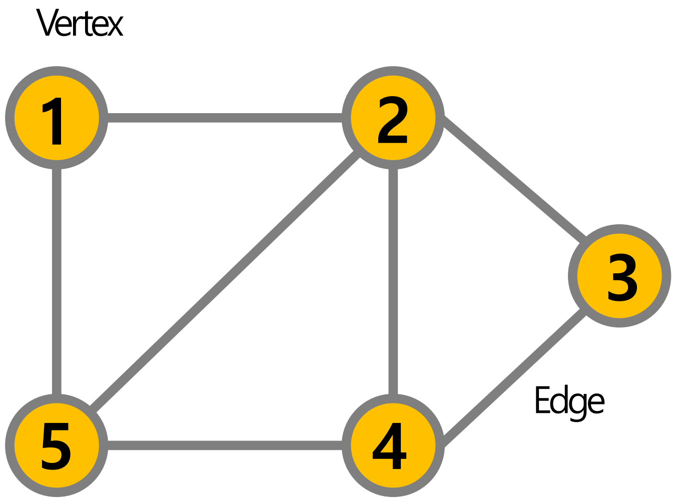
Gaussian kernel similarity function

$$w_{ij} = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}}$$

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, Laplacian matrix

- Adjacency matrix : Undirected graph → Symmetric
- Edge set: $\boldsymbol{E}(\boldsymbol{G}) = \{\mathbf{1},\ \mathbf{0}\}$



Adjacency Matrix

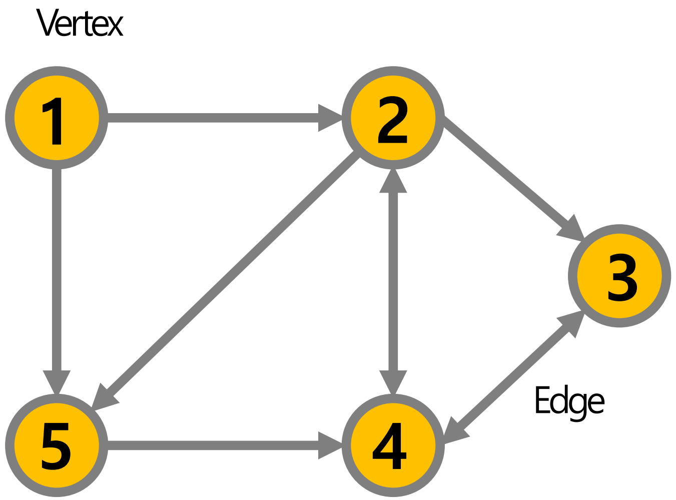
$A \in R^{n \times n}$

0	1	0	0	1
1	0	1	1	1
0	1	0	1	0
0	1	1	0	1
1	1	0	1	0

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, Laplacian matrix

- Adjacency matrix : Directed graph → Asymmetric
- Edge set: $E(G) = \{1, 0\}$



Adjacency Matrix

$A \in R^{n \times n}$

input / output

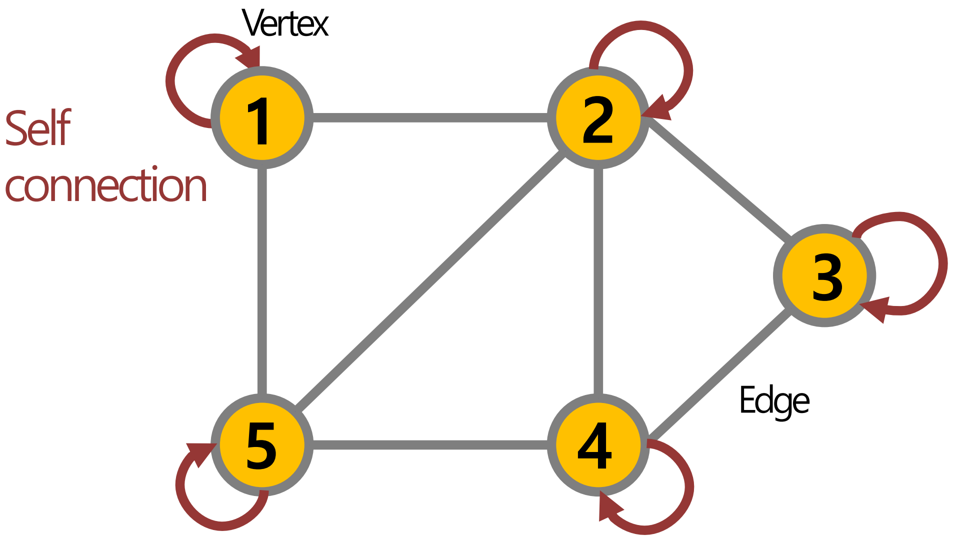
0	1	0	0	1
0	0	1	0	1
0	0	0	1	0
0	1	1	0	0
0	0	0	1	0

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, Laplacian matrix

❖ Adjacency matrix + Identity matrix

❖ Edge set: $\boldsymbol{E}(\boldsymbol{G}) = \{\mathbf{1},\ \mathbf{0}\}$



Adjacency Matrix

$\tilde{A} \in R^{n \times n}$

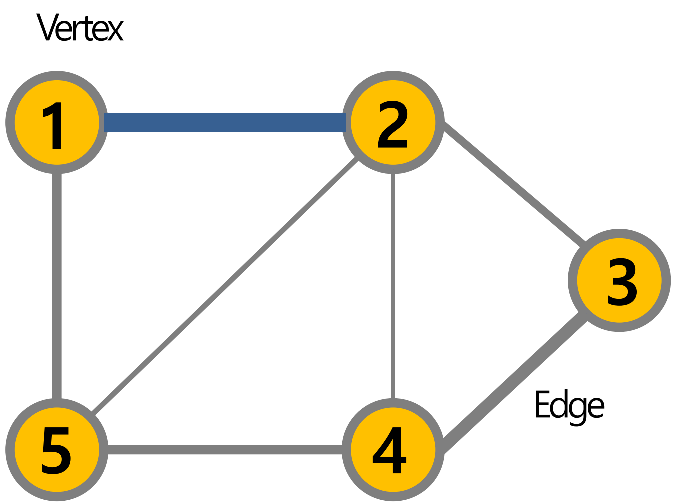
1	1	0	0	1
1	1	1	1	1
0	1	1	1	0
0	1	1	1	1
1	1	0	1	1

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, Laplacian matrix

❖ Adjacency matrix : Weighted Undirected graph

❖ Edge set: $\boldsymbol{E}(\boldsymbol{G}) = \{\boldsymbol{w}_{\boldsymbol{i},\boldsymbol{j}}\},\ \mathbf{1} \leq \boldsymbol{i},\ \boldsymbol{j} \leq \boldsymbol{N}$



Adjacency Matrix

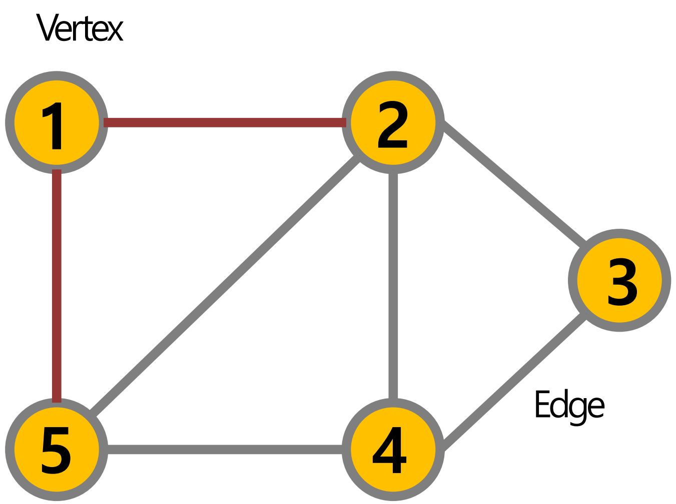
$A \in R^{n \times n}$

0	1.5	0	0	1
1.5	0	0.8	0.2	0.4
0	0.8	0	1.2	0
0	0.2	1.2	0	1
1	0.4	0	1	0

Graph Data

Graph for Data Analysis : Adjacency matrix, **Degree matrix**, Laplacian matrix

- ❖ Degree는 각 vertex와 연결된 edge의 수
- ❖ 따라서, adjacency matrix의 row sum값



Adjacency Matrix

$A \in R^{n \times n}$

0	1	0	0	1
1	0	1	1	1
0	1	0	1	0
0	1	1	0	1
1	1	0	1	0

Degree

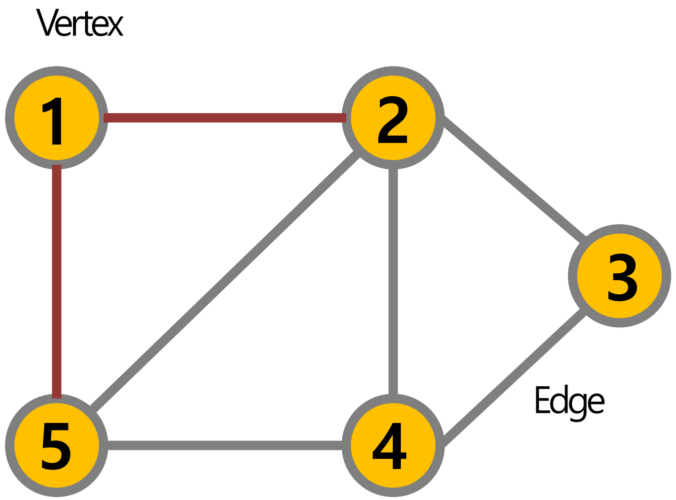
$\sum_j A_{i,j}$

2
4
2
3
3

Graph Data

Graph for Data Analysis : Adjacency matrix, **Degree matrix**, Laplacian matrix

❖ Degree matrix : 대각행렬에 degree값으로 구성되며, 나머지는 0을 지님



Degree Matrix

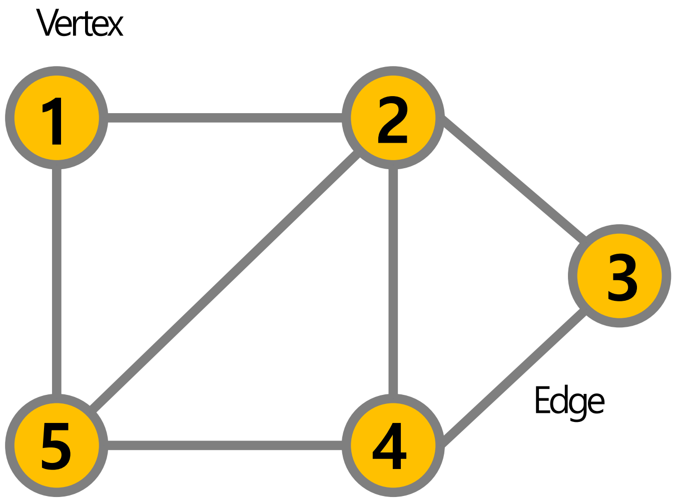
$$D \in R^{n \times n},\ D_{i,i} = \sum_j A_{i,j}$$

2	0	0	0	0
0	4	0	0	0
0	0	2	0	0
0	0	0	3	0
0	0	0	0	3

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, **Laplacian matrix**

❖ Laplacian matrix : Degree matrix – Adjacency matrix



Degree Matrix

$D \in R^{n \times n},\ D_{i,i} = \sum_j A_{i,j}$

2	0	0	0	0
0	4	0	0	0
0	0	2	0	0
0	0	0	3	0
0	0	0	0	3

−

Adjacency Matrix

$A \in R^{n \times n}$

0	1	0	0	1
1	0	1	1	1
0	1	0	1	0
0	1	1	0	1
1	1	0	1	×0

=

Laplacian Matrix

$L \in R^{n \times n},\ L = D - A$

2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3

The elements of L are given by

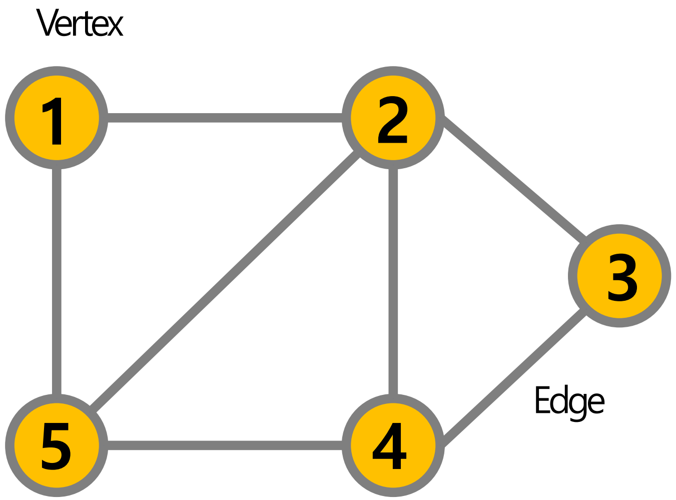
$$L_{i,j} = \begin{cases} deg(v_i) & if\ i = j \\ -1 & if\ i \neq j \\ 0 & otherwise \end{cases}$$

where $deg(v_i)$ is the degree of the vertex i

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, **Laplacian matrix**

❖ Laplacian matrix : Degree matrix – Adjacency matrix



Laplacian Matrix

$$L \in R^{N \times N},\ L = D - A$$

2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3

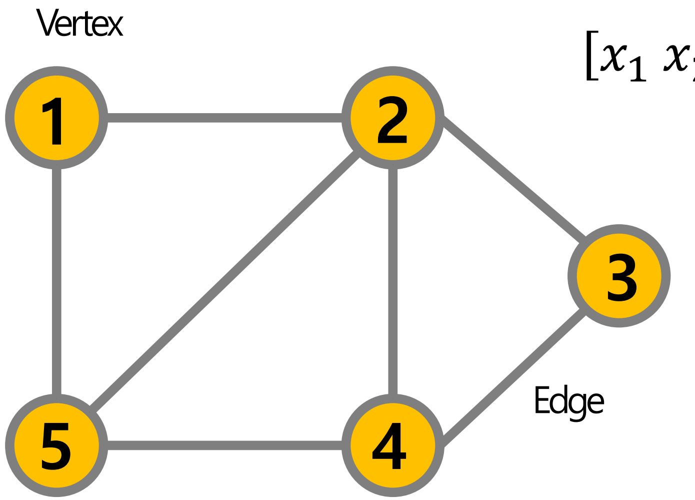
이웃 vertex와의 관계 정보

Degree 정보

Graph Data

Graph for Data Analysis : Adjacency matrix, Degree matrix, **Laplacian matrix**

❖ Laplacian matrix : Degree matrix – Adjacency matrix



Laplacian Matrix

$$L \in R^{N \times N},\ L = D - A$$

$$[x_1\ x_2\ x_3\ x_4\ x_5] \times$$

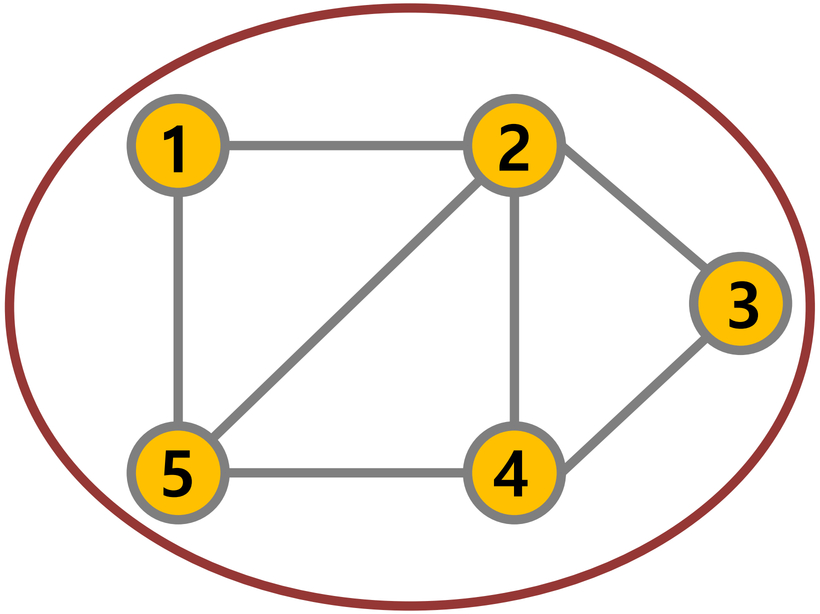
2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3

$$-x_1 + 4x_2 - x_3 - x_4 - x_5$$
$$= (x_2 - x_1) + (x_2 - x_3)$$
$$+ (x_2 - x_4) + (x_2 - x_5)$$

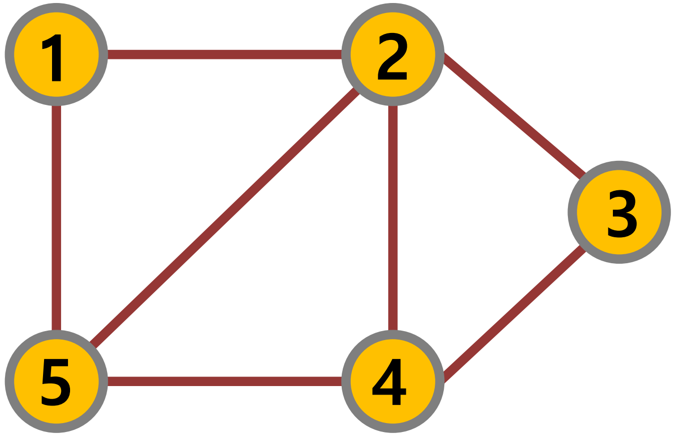
중심 vertex와 이웃 vertex
사이의 관계 정보

Graph Data

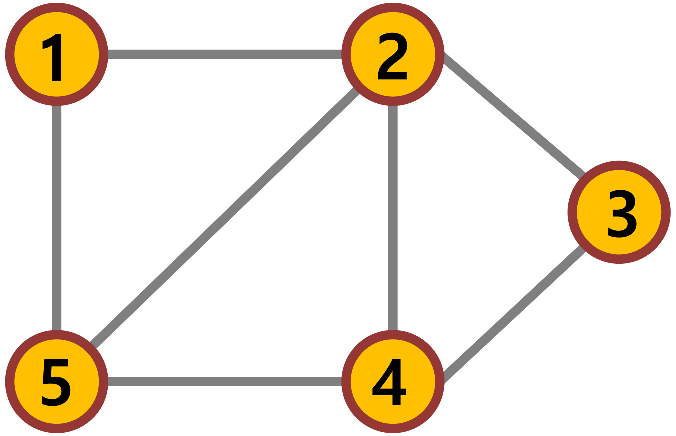
Graph Tasks : Graph prediction, Edge prediction, Node prediction



Graph prediction



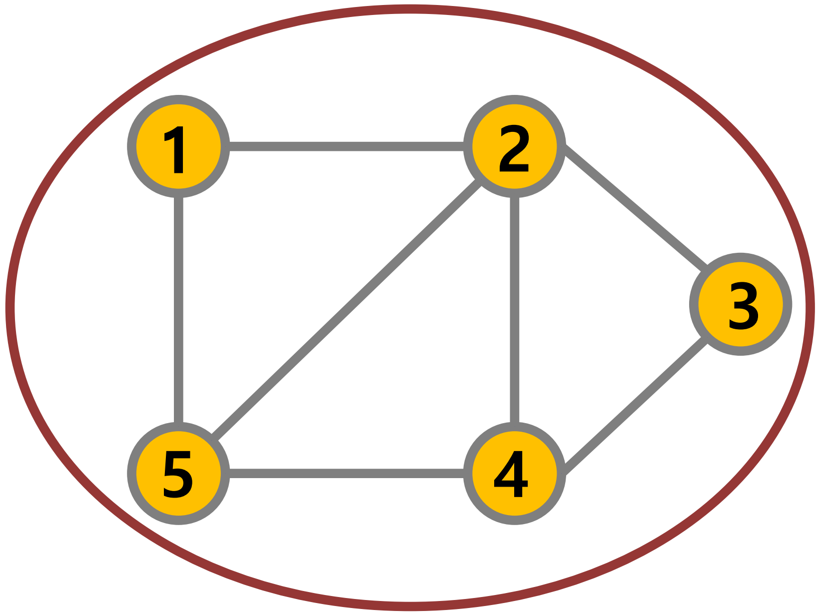
Edge prediction



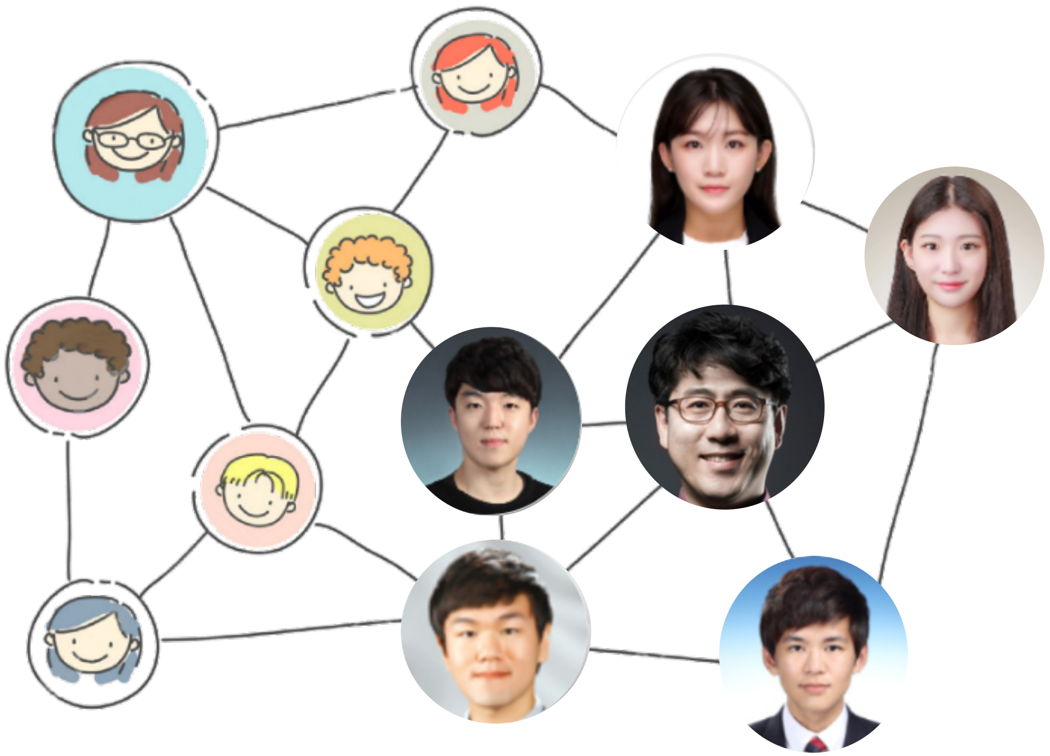
Node prediction

Graph Data

Graph Tasks : **Graph prediction**, Edge prediction, Node prediction



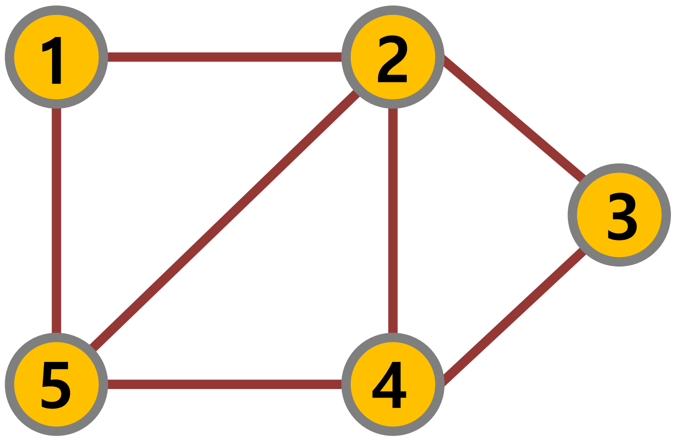
Graph prediction



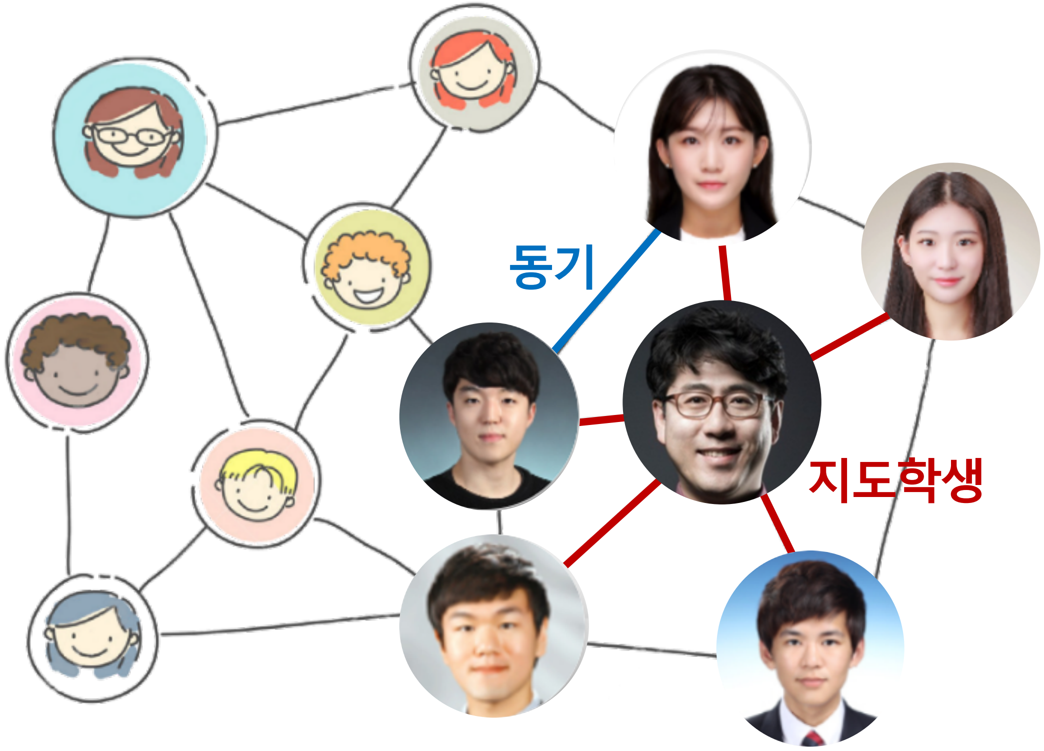
DMQA

Graph Data

Graph Tasks : Graph prediction, **Edge prediction**, Node prediction



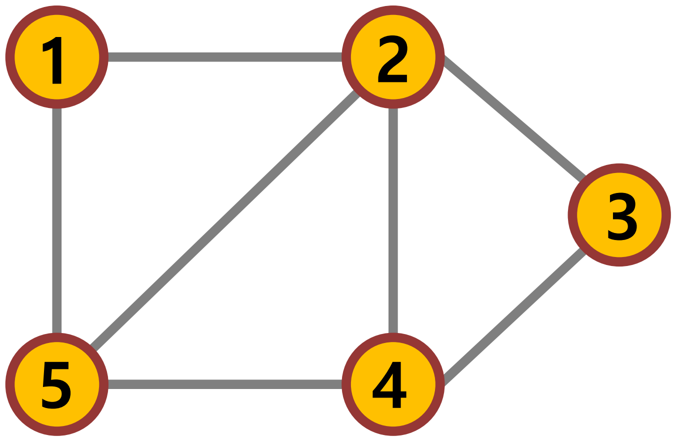
Edge prediction



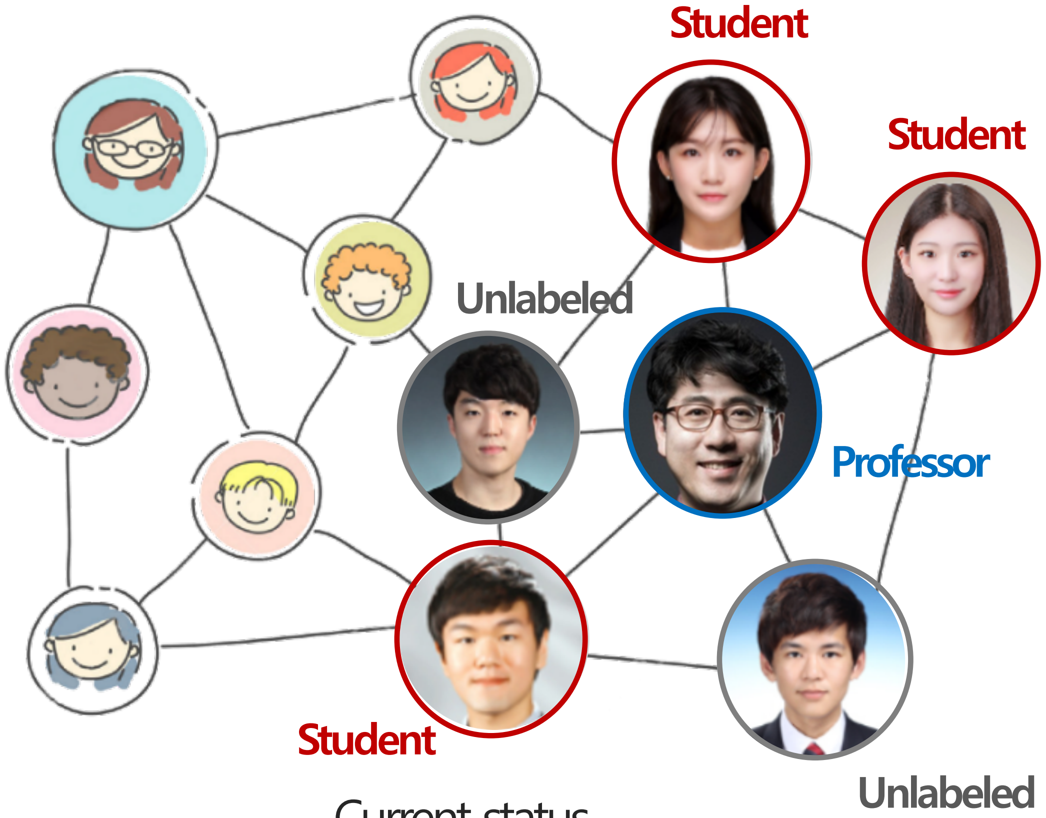
Relationship

Graph Data

Graph Tasks : Graph prediction, Edge prediction, **Node prediction**



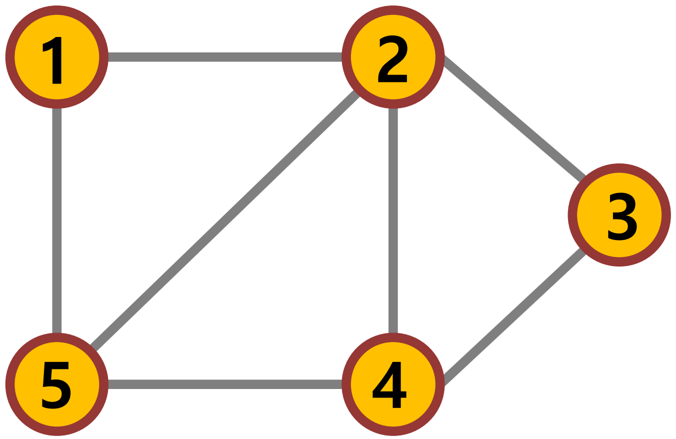
Node prediction



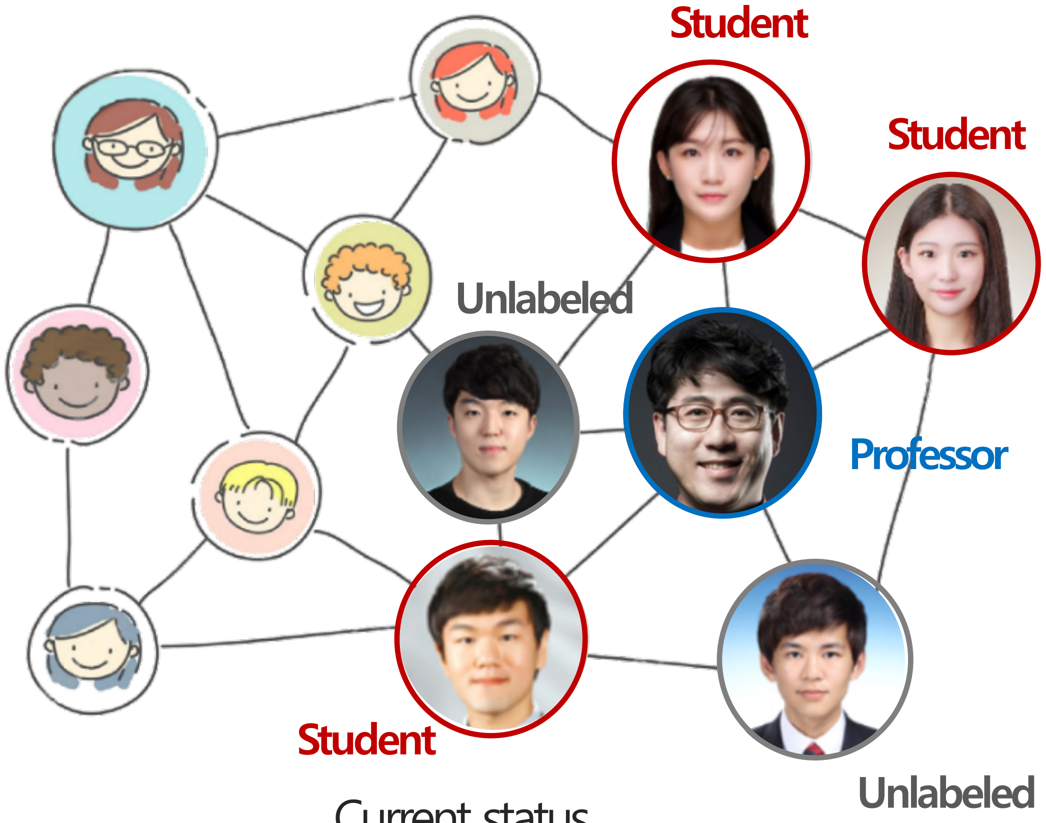
Current status

Graph Data

Graph Tasks : Graph prediction, Edge prediction, **Node prediction**



Node prediction



Current status

Semi-supervised learning

Graph-Based

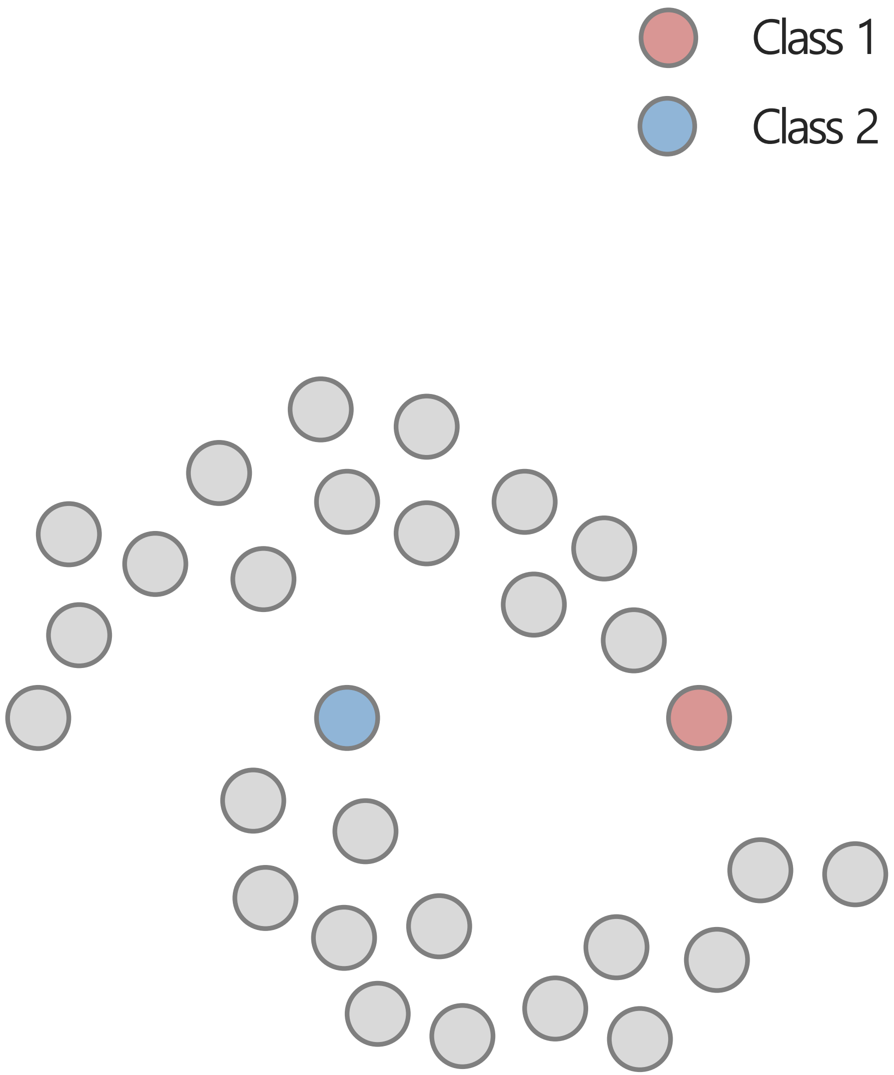
Semi-Supervised Learning

Semi-Supervised Learning

Limitations of Supervised Learning

❖ 데이터 레이블링에는 많은 시간과 비용이 발생

$X_{labeled}$	y
$X_{unlabeled}$	?

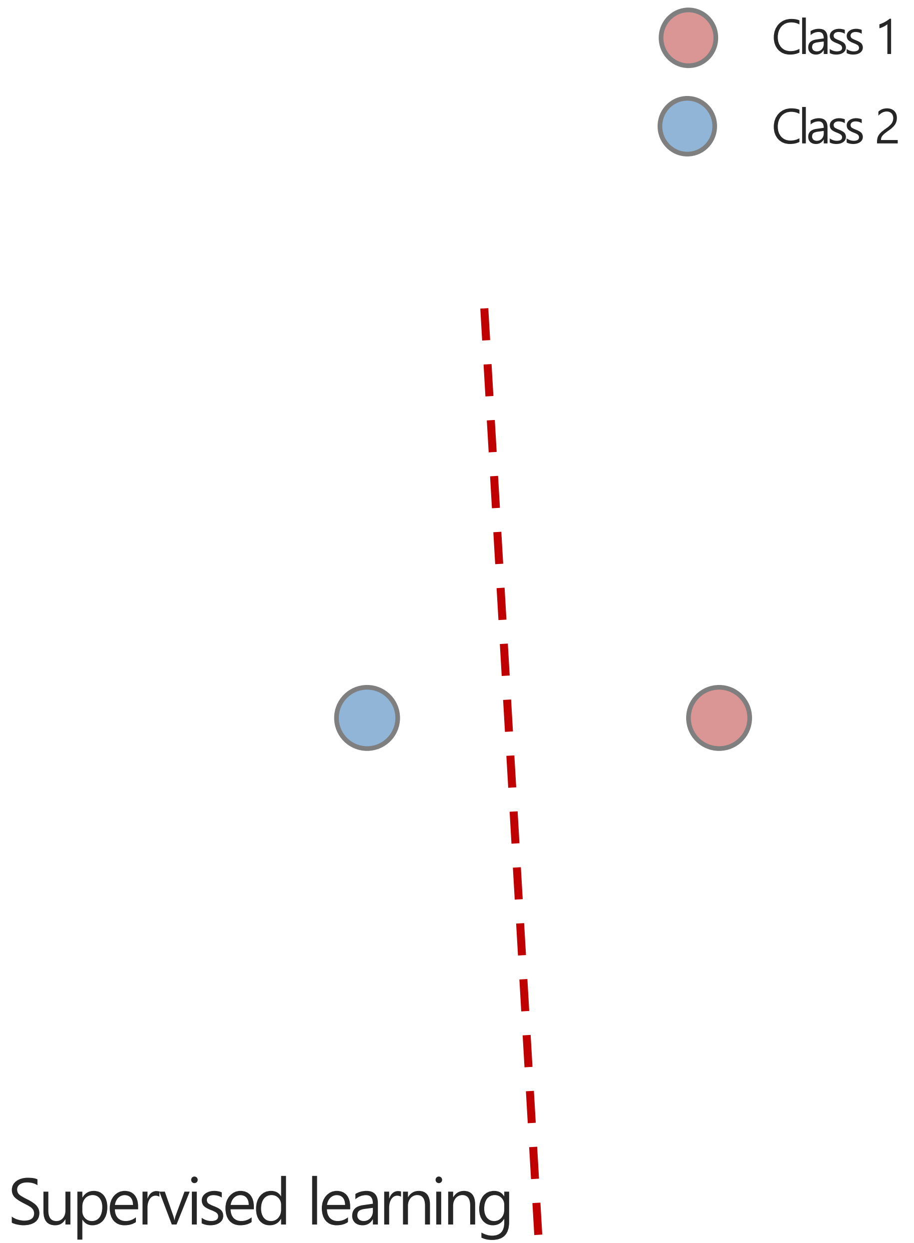


Semi-Supervised Learning

Limitations of Supervised Learning

❖ 데이터 레이블링에는 많은 시간과 비용이 발생

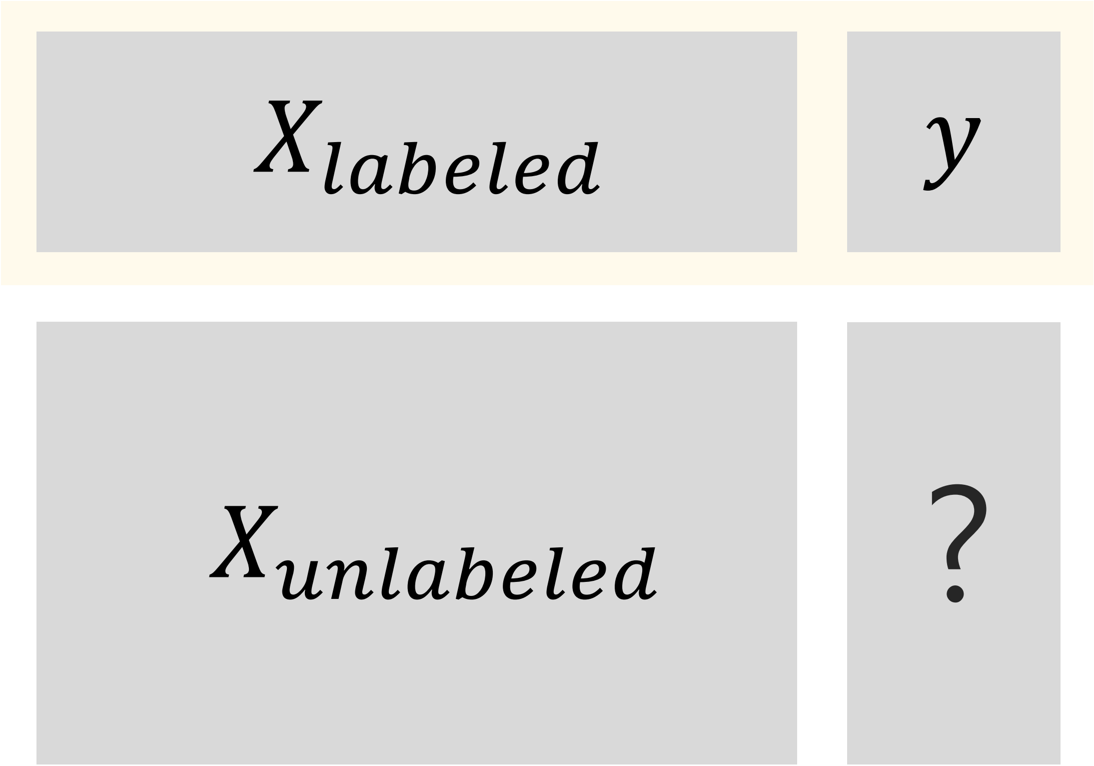
$X_{labeled}$	y
$X_{unlabeled}$	?

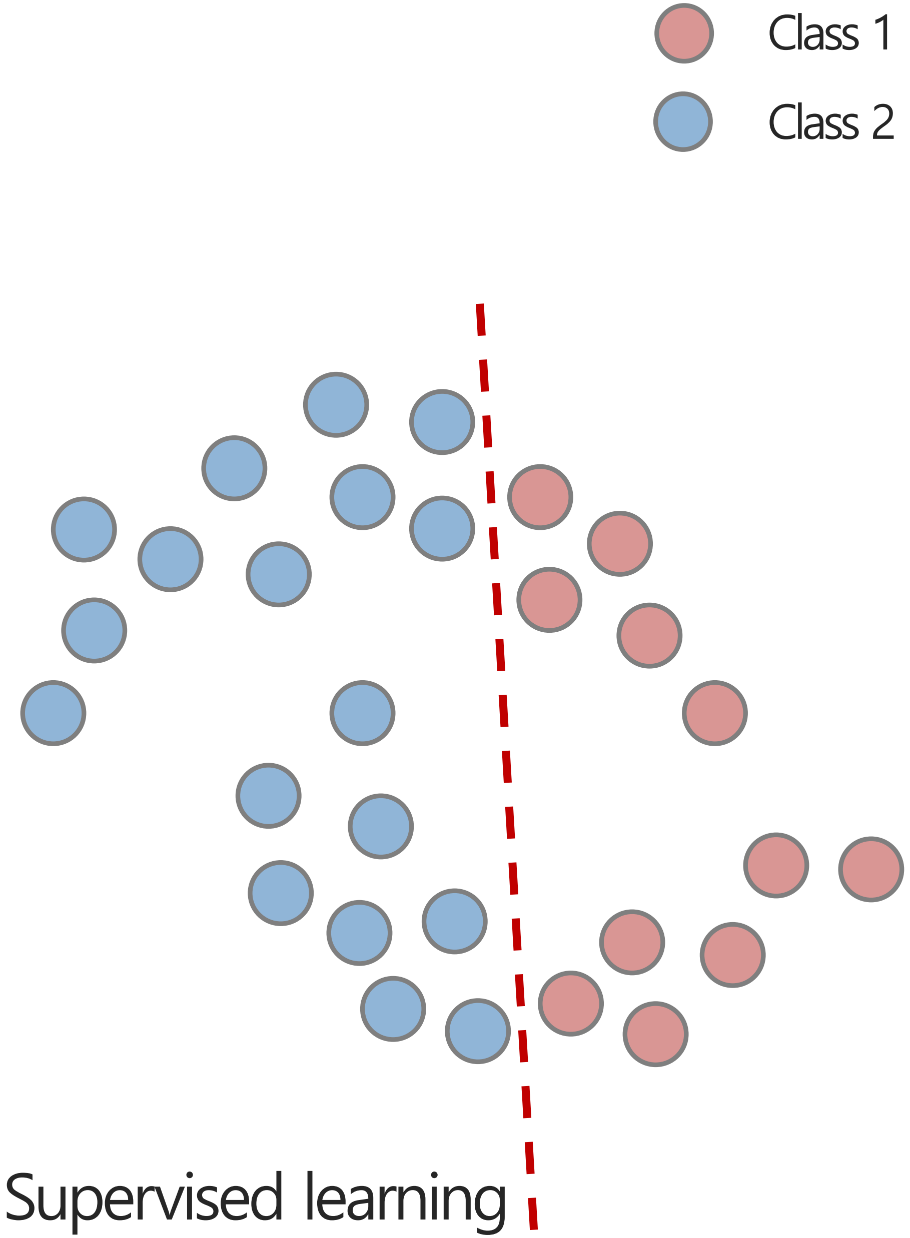


Semi-Supervised Learning

Limitations of Supervised Learning

❖ 실제 label distribution $p(y|x)$를 학습하기에는 labeled data가 부족

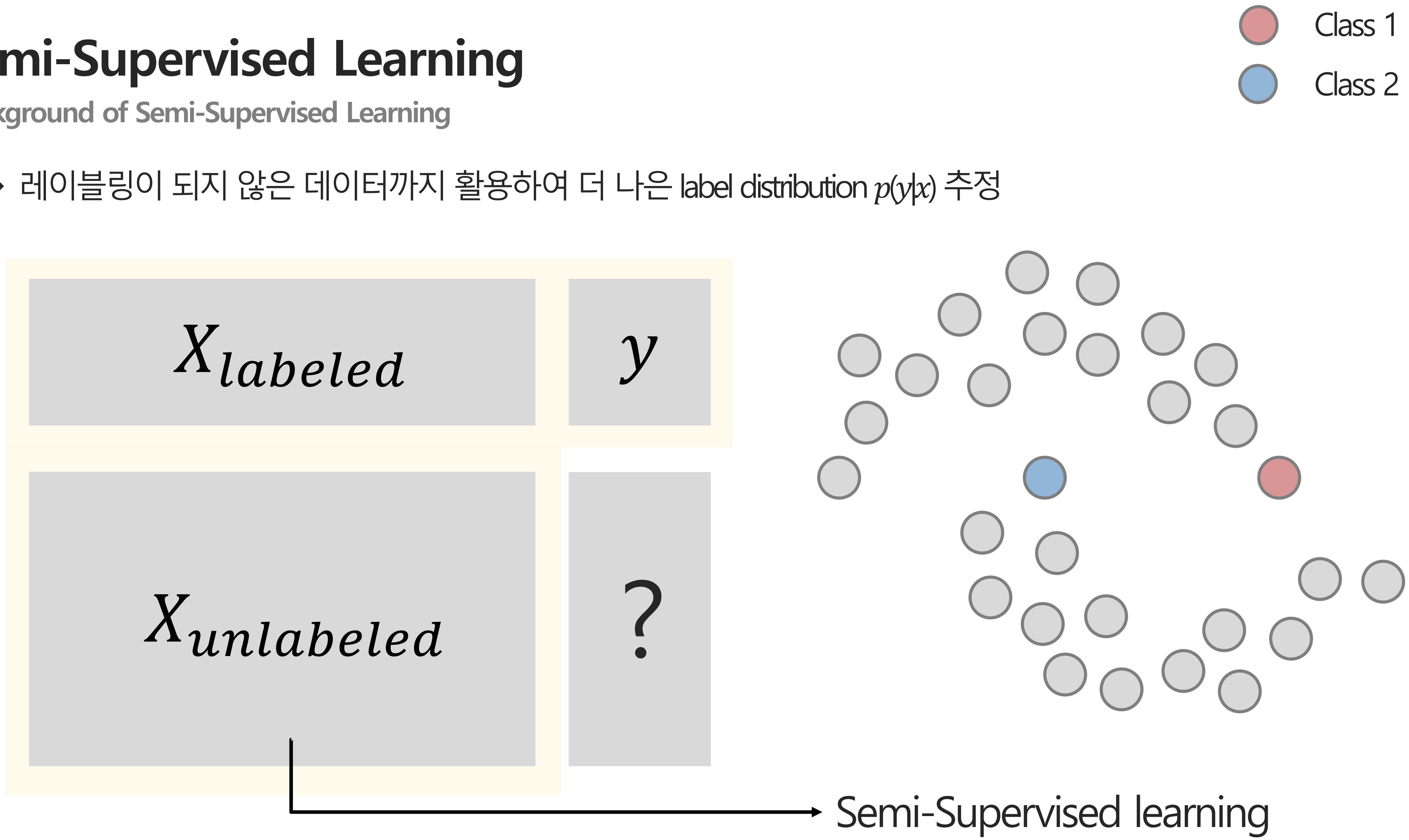




Semi-Supervised Learning

Background of Semi-Supervised Learning

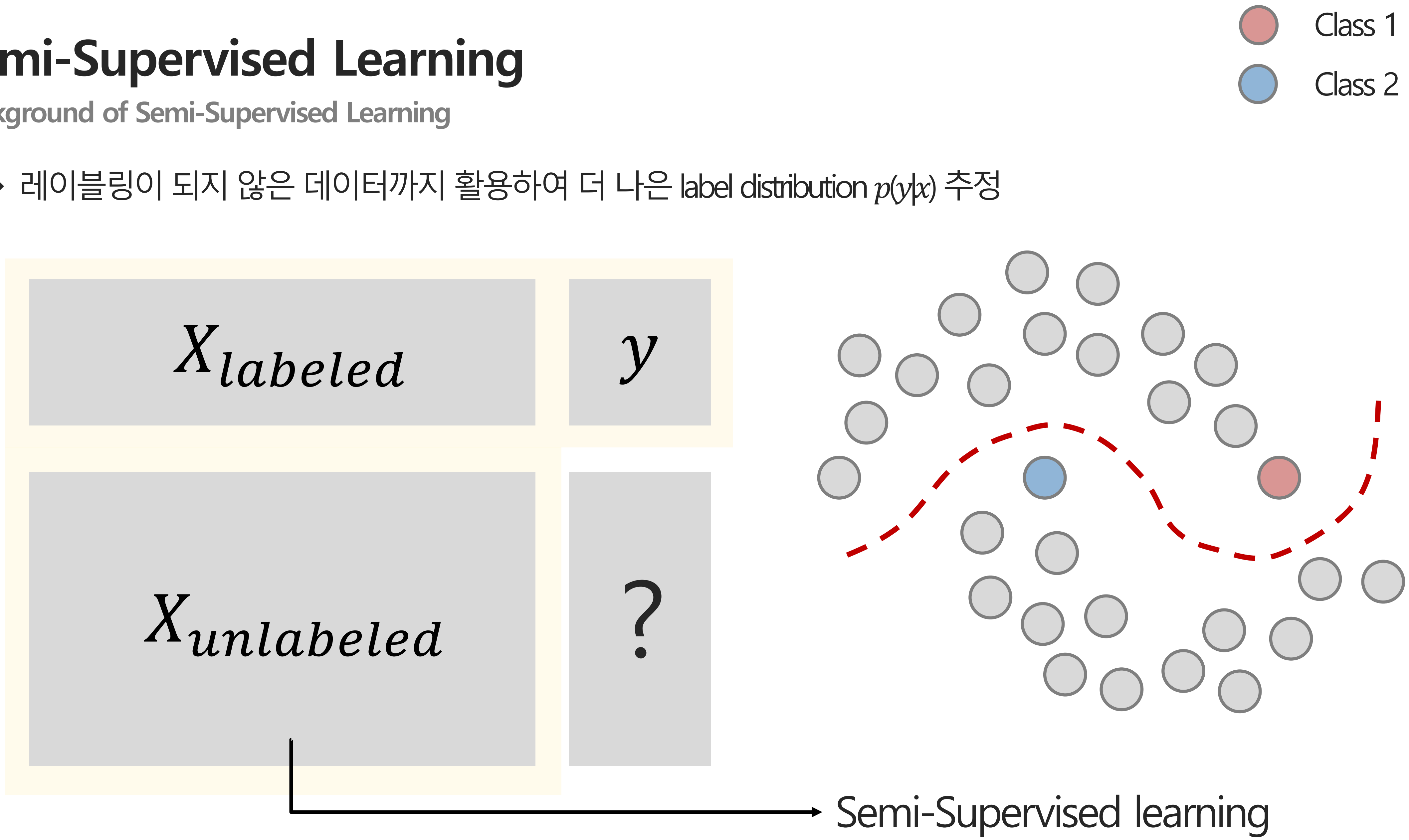
❖ 레이블링이 되지 않은 데이터까지 활용하여 더 나은 label distribution $p(y|x)$ 추정



Semi-Supervised Learning

Background of Semi-Supervised Learning

❖ 레이블링이 되지 않은 데이터까지 활용하여 더 나은 label distribution $p(y|x)$ 추정



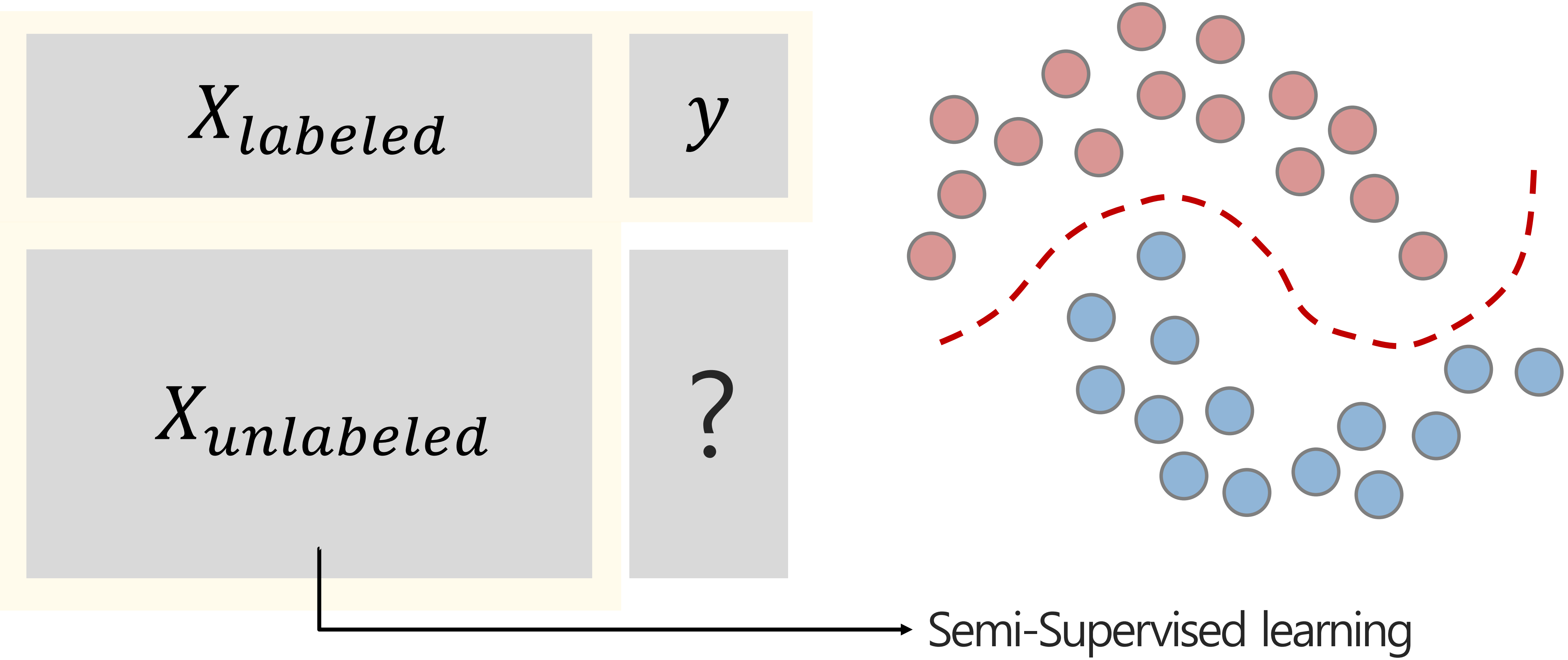
Semi-Supervised Learning

Background of Semi-Supervised Learning

Class 1

Class 2

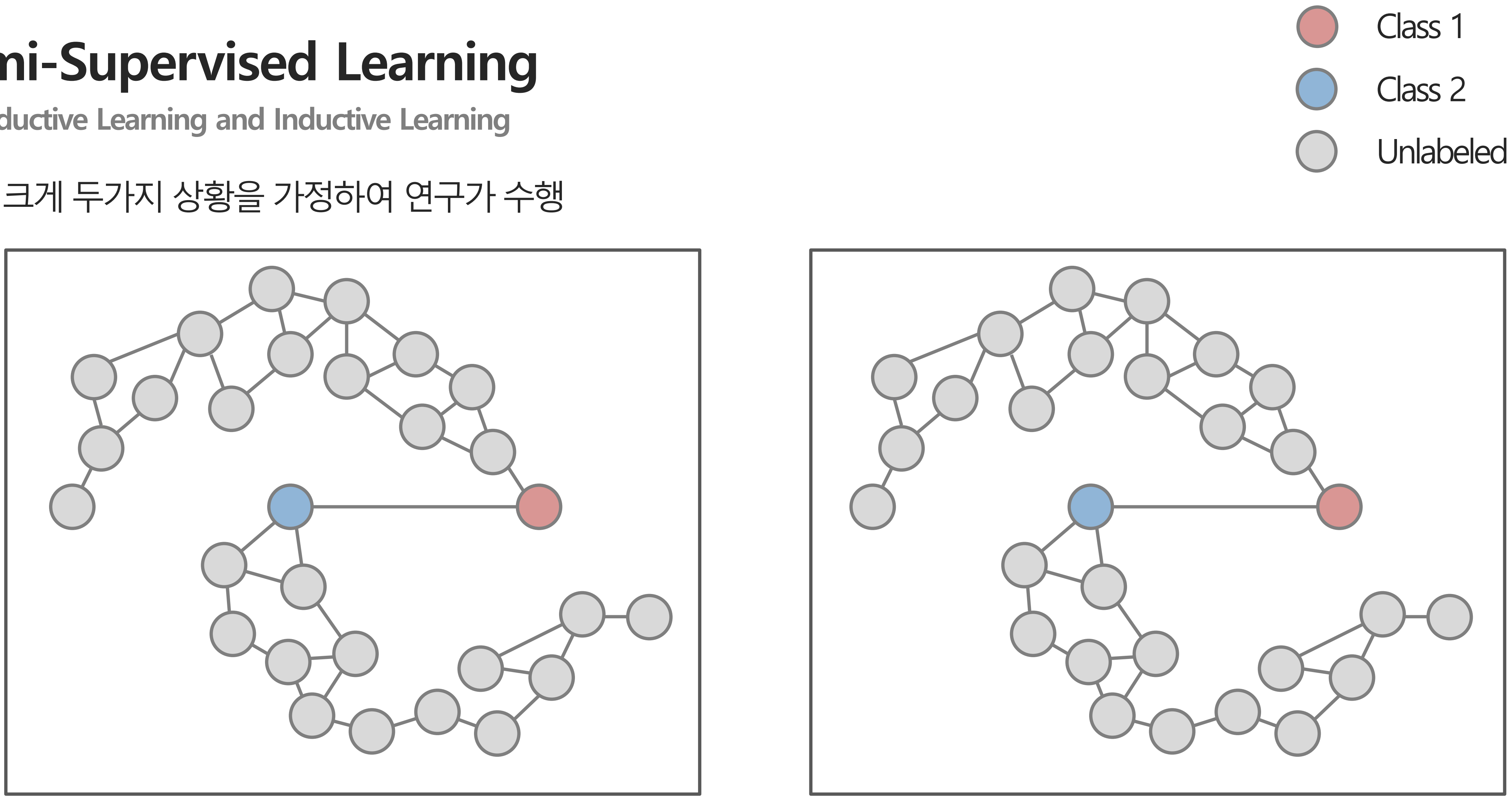
❖ 레이블링이 되지 않은 데이터까지 활용하여 더 나은 label distribution $p(y|x)$ 추정



Semi-Supervised Learning

Transductive Learning and Inductive Learning

❖ 크게 두가지 상황을 가정하여 연구가 수행



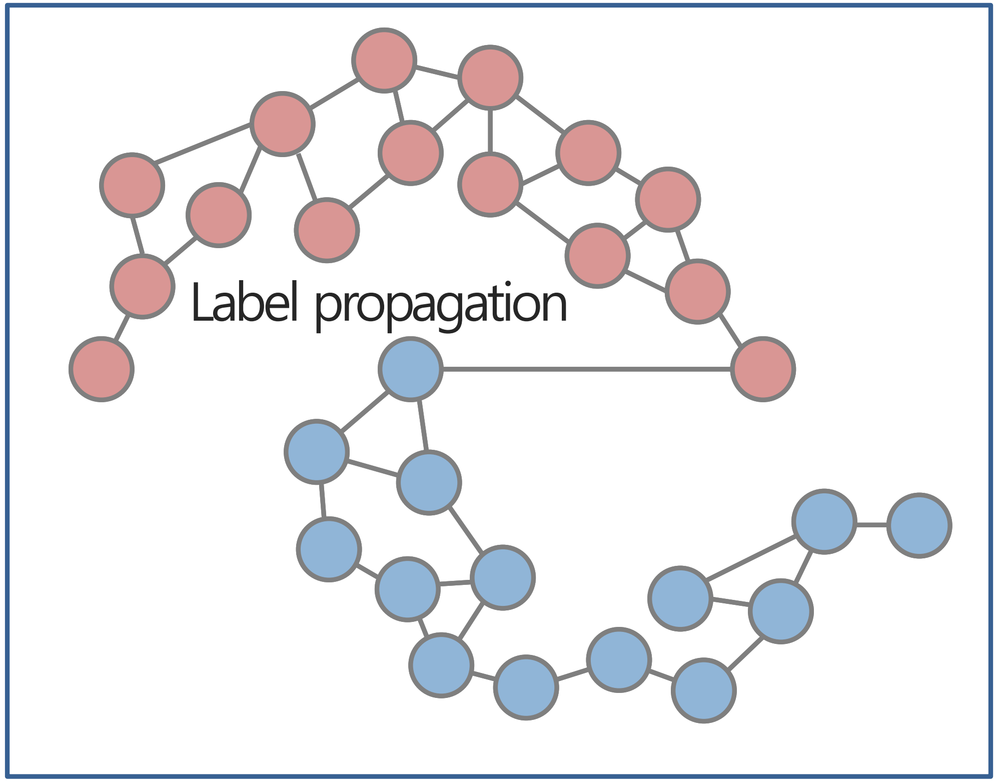
Transductive learning

Inductive learning

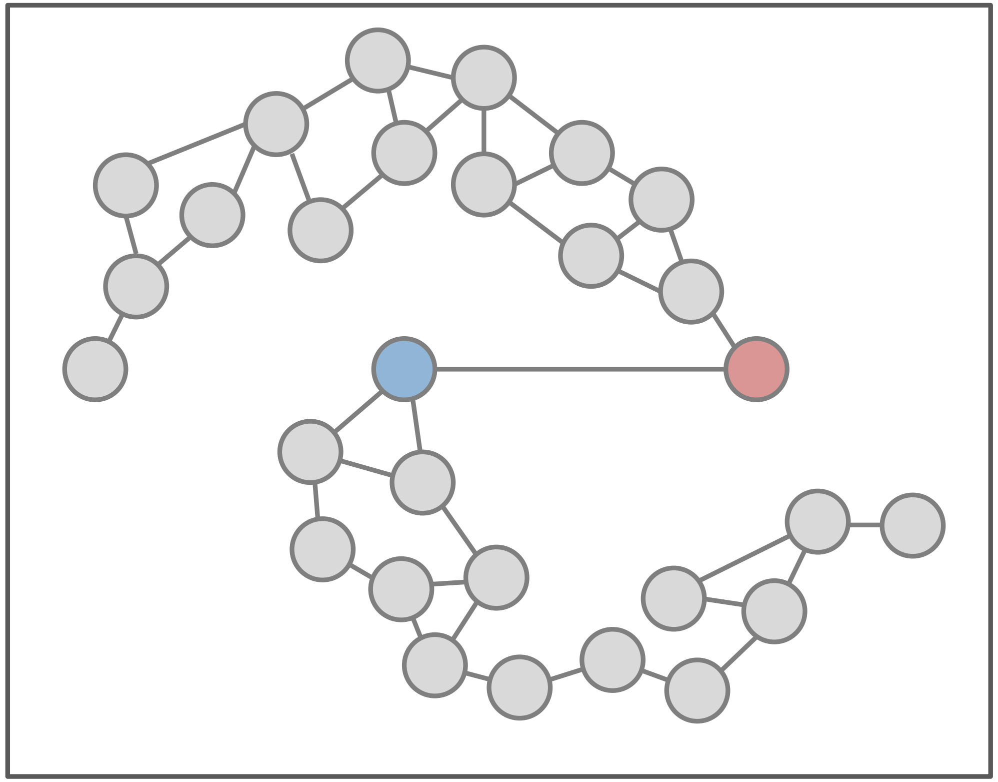
Semi-Supervised Learning

Transductive Learning and Inductive Learning

❖ 크게 두가지 상황을 가정하여 연구가 수행



Transductive learning

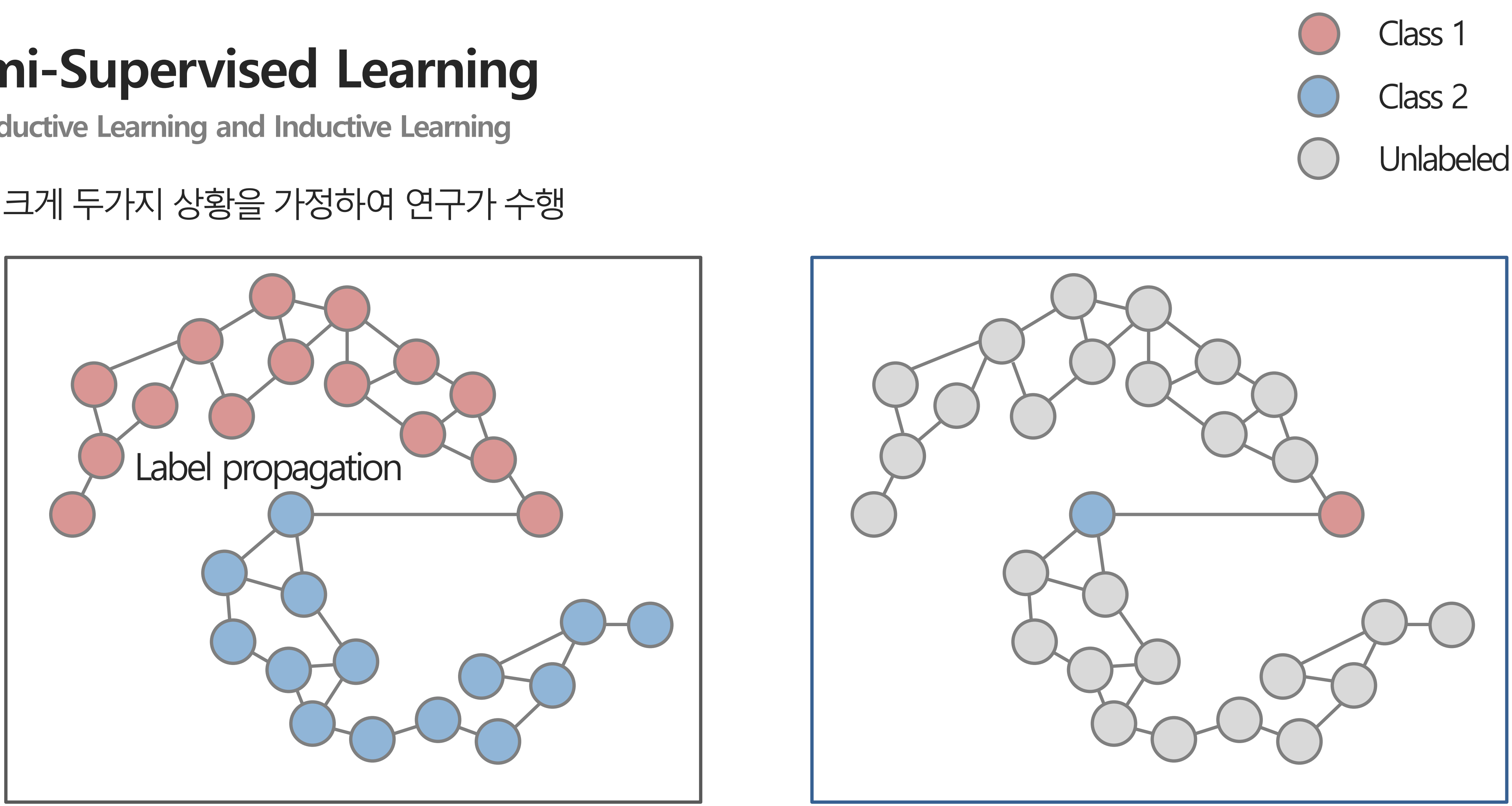


Inductive learning

Semi-Supervised Learning

Transductive Learning and Inductive Learning

❖ 크게 두가지 상황을 가정하여 연구가 수행



Transductive learning

Inductive learning

Semi-Supervised Learning

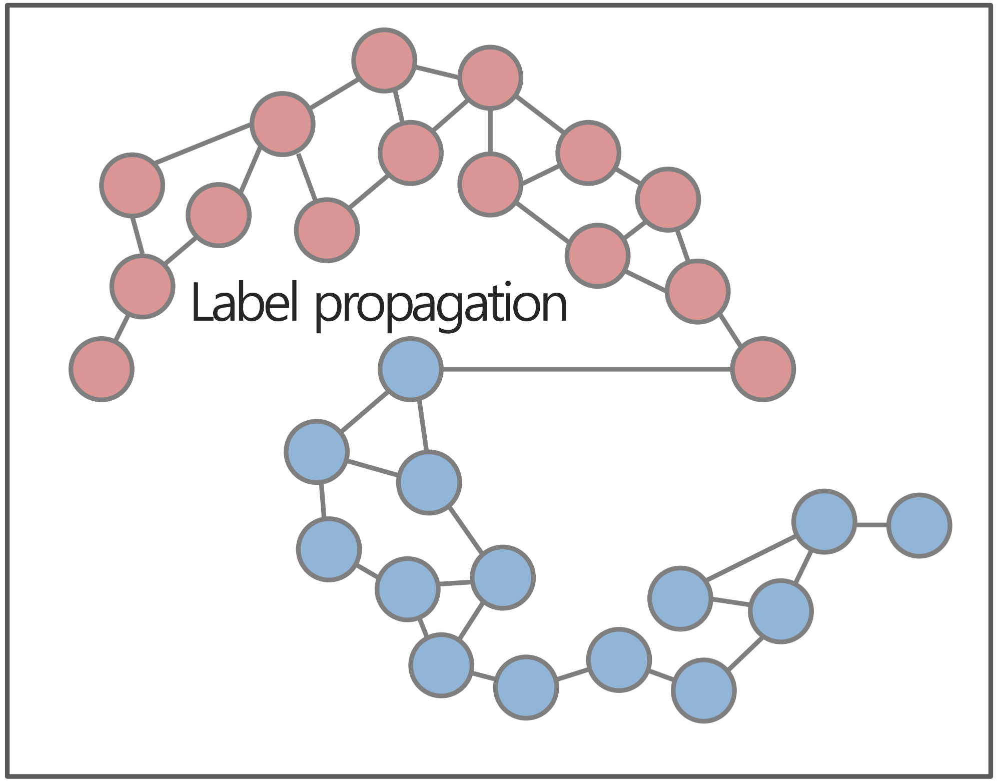
Transductive Learning and Inductive Learning

❖ 크게 두가지 상황을 가정하여 연구가 수행

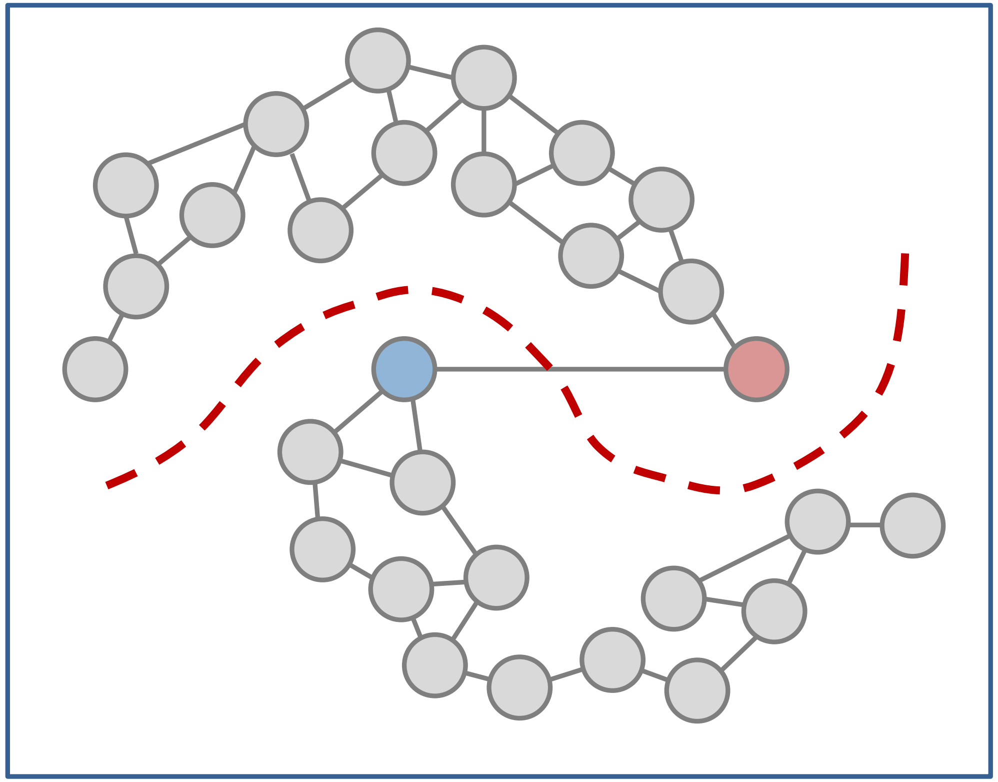
Class 1

Class 2

Unlabeled



Transductive learning



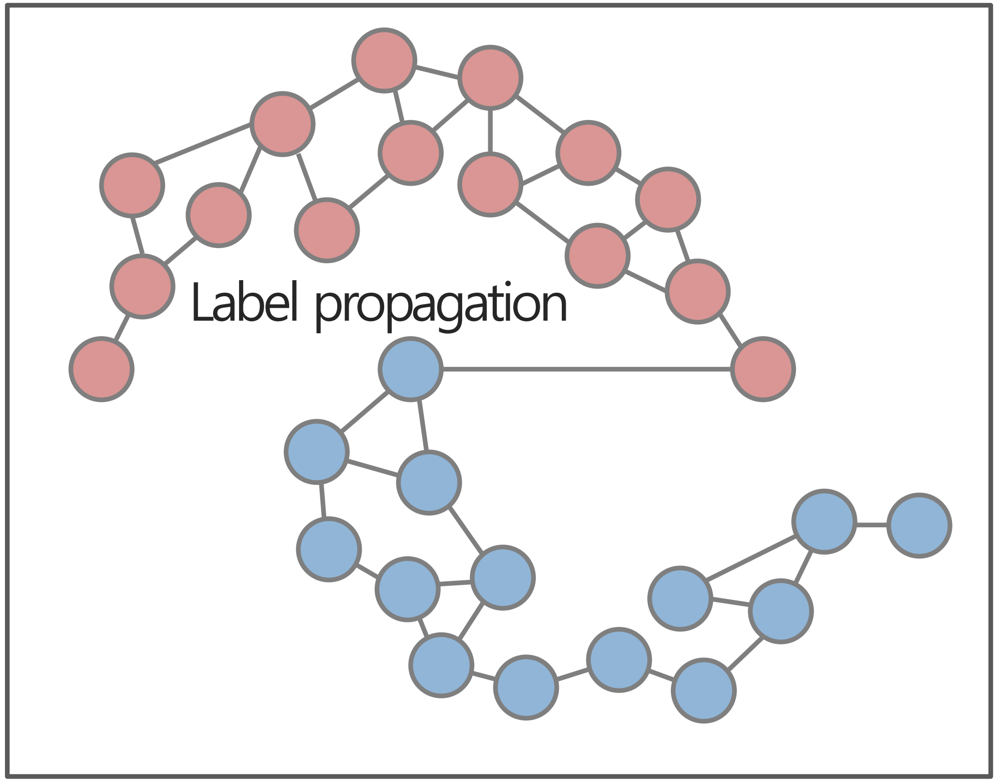
Inductive learning

Semi-Supervised Learning

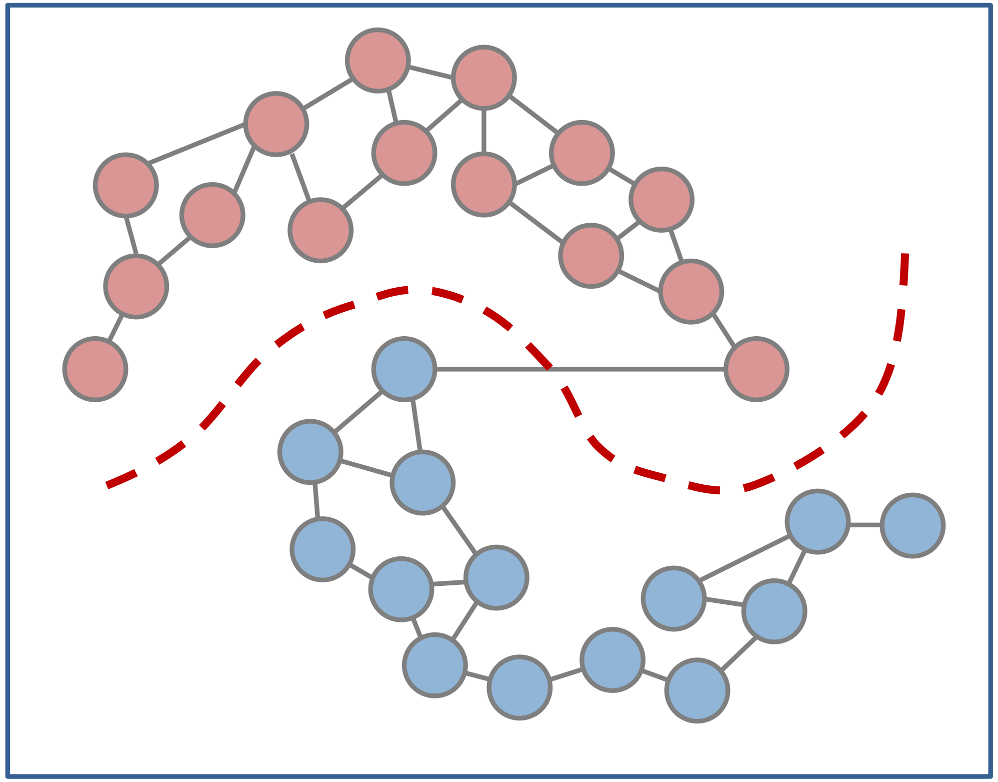
Transductive Learning and Inductive Learning

- Class 1
- Class 2
- Unlabeled

❖ 크게 두가지 상황을 가정하여 연구가 수행



Transductive learning

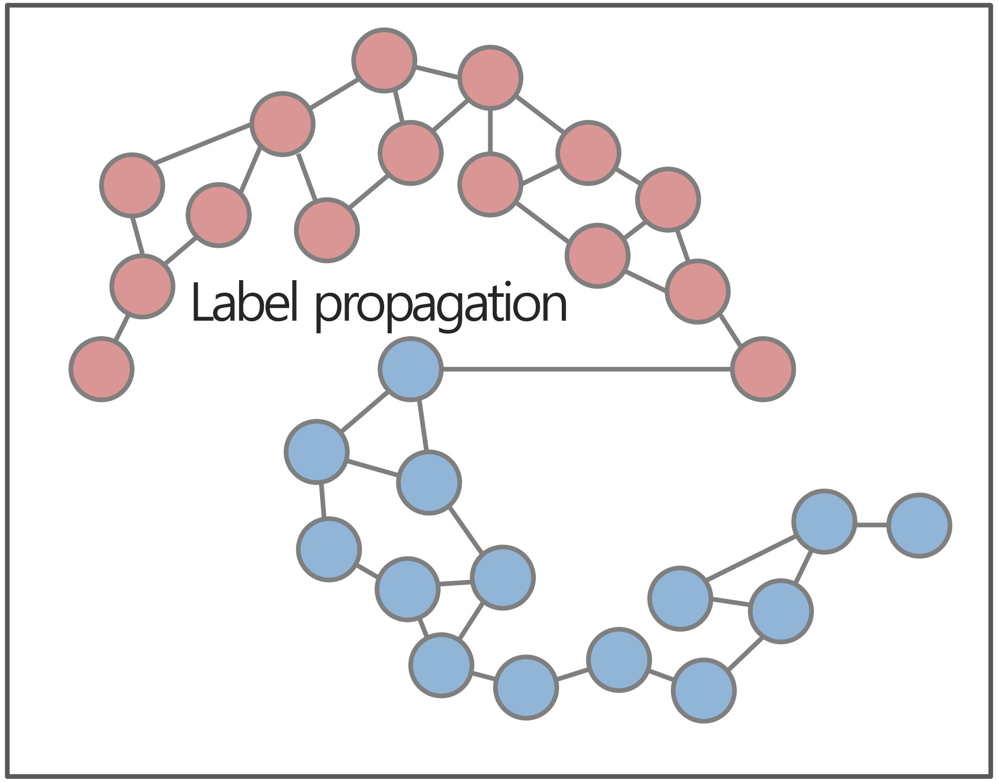


Inductive learning

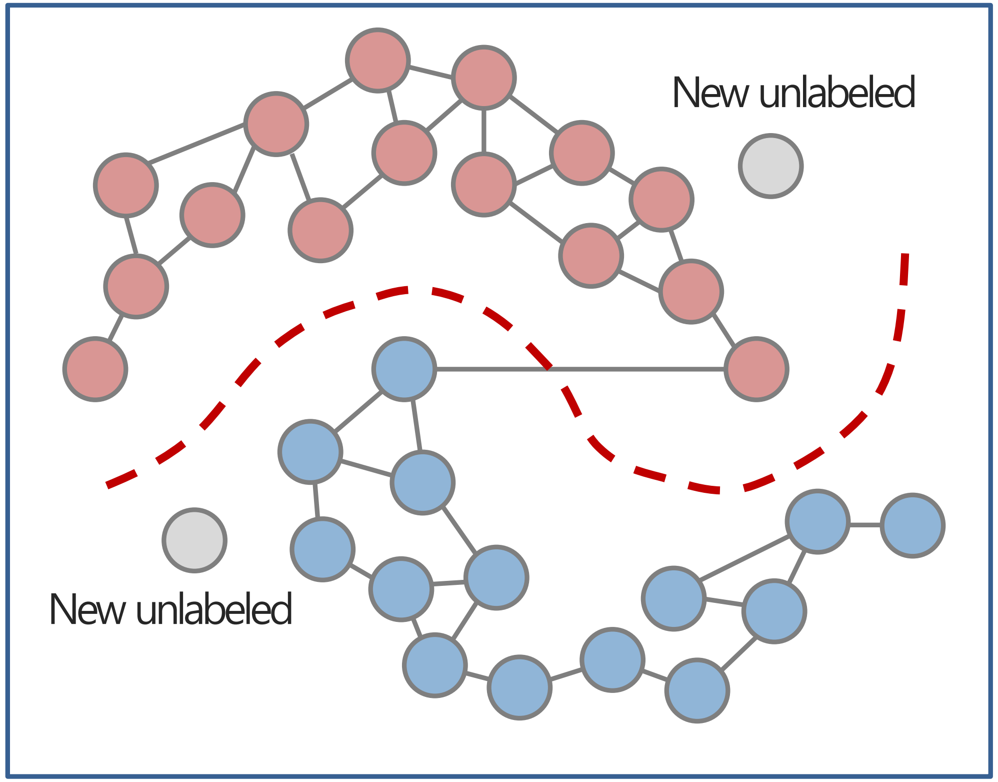
Semi-Supervised Learning

Transductive Learning and Inductive Learning

❖ 크게 두가지 상황을 가정하여 연구가 수행



Transductive learning

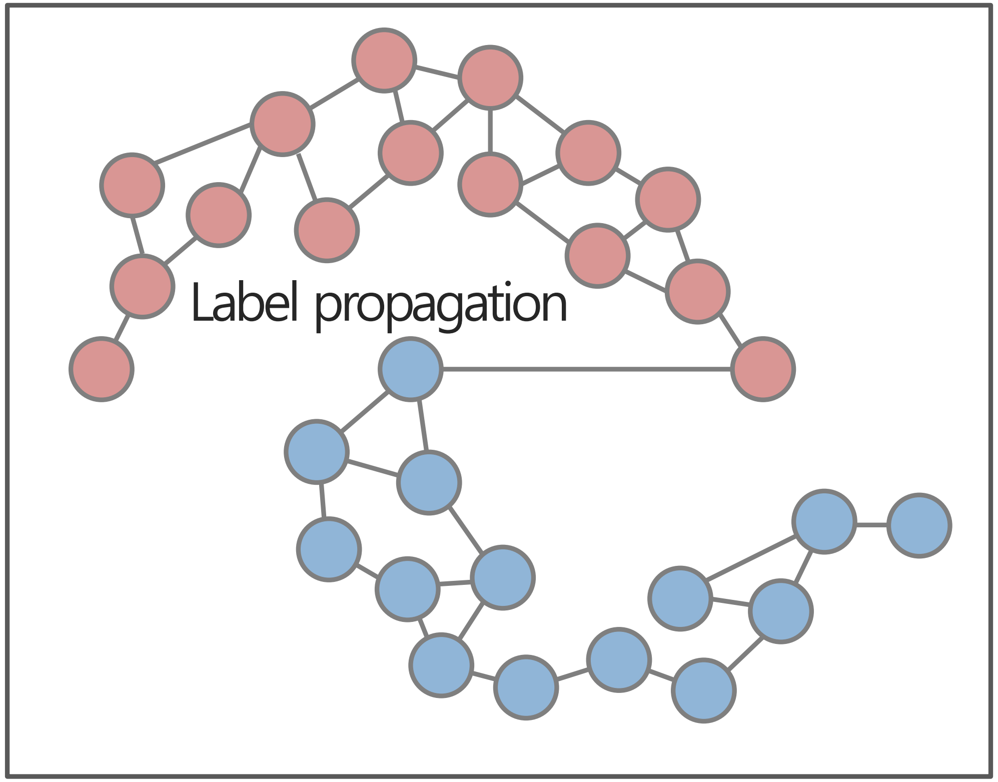


Inductive learning

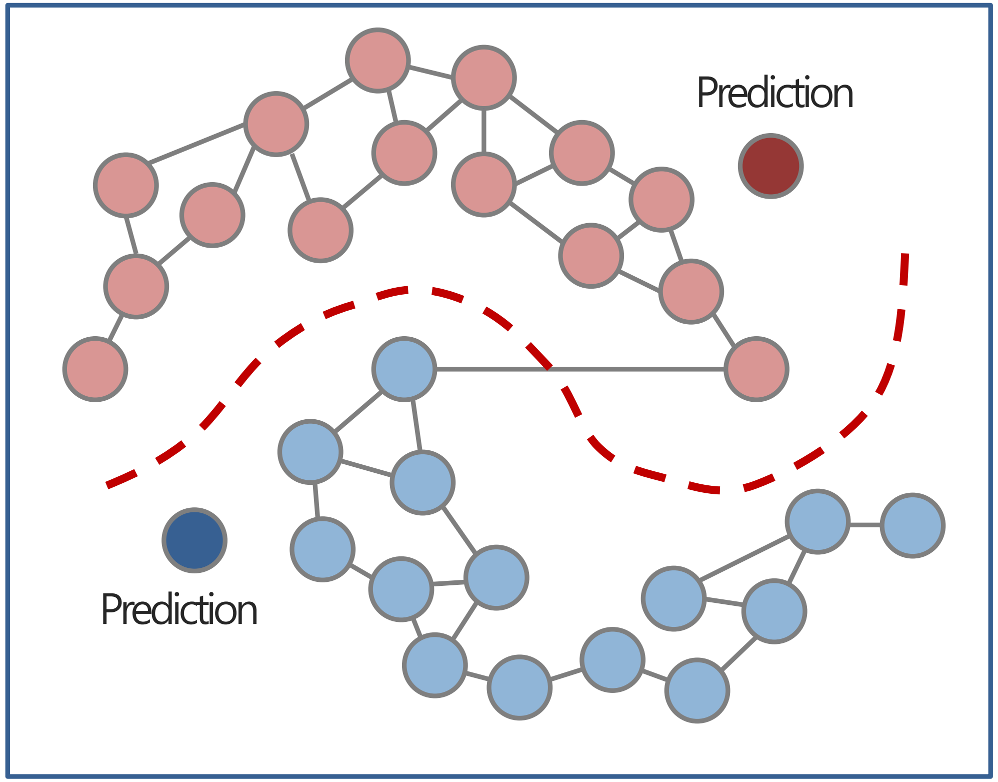
Semi-Supervised Learning

Transductive Learning and Inductive Learning

❖ 크게 두가지 상황을 가정하여 연구가 수행



Transductive learning



Inductive learning

Graph-Based

Semi-Supervised Learning

Label Propagation

Graph Convolutional Neural Network

Graph Convolutional Network

Min-cut

$$\min_{y \in \{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(y_i - y_{L_i}\right)^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}\left(y_i - y_j\right)^2$$

Harmonic Solution

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Local and Global Consistency

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)\left(f(x_i) - y_{L_i}\right)^2 + \mu f^T L f$$

Labeled : Supervised loss

Unlabeled : Semi-Supervised loss

Label Propagation

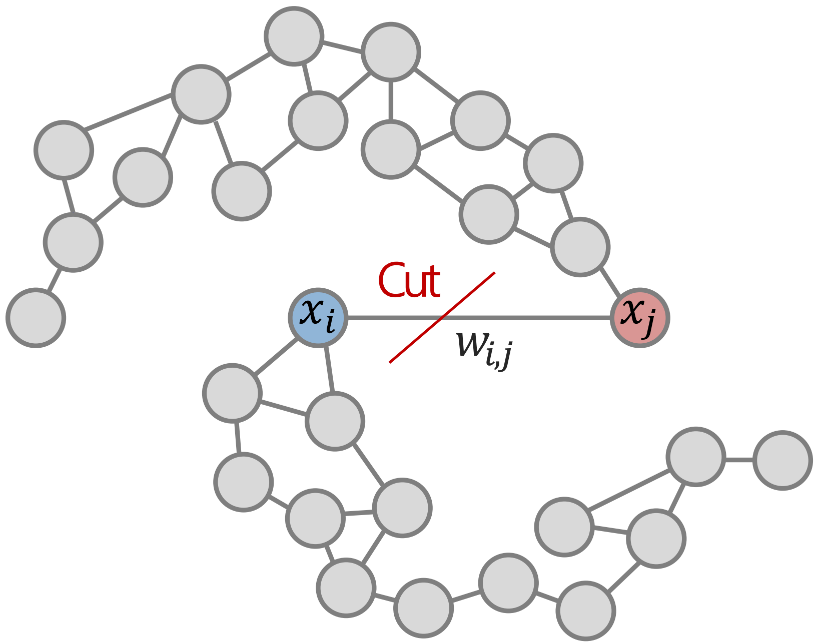
Notation

- Labeled data $\{(x_L, y_L)\} = \{(x_1\ y_1), \dots, (x_{n_L}\ y_{n_L})\}$
- Unlabeled data $x_U = \{(x_{n_L+1}), \dots, (x_{n_L+n_U})\}$
- Input $x = \{x_L \cup x_U\}, n = n_L + n_U, x_i \in R^d$
- Output $y = \{y_L \cup \widehat{y_U}\}, n = n_L + n_U, y_i \in \{0, 1\}$

Label Propagation

Background of Min-cut Algorithm

- Labeled data $\{(x_L, y_L)\} = \{(x_1\ y_1), \ldots, (x_{n_L}\ y_{n_L})\}$
- Unlabeled data $x_U = \{(x_{n_L+1}), \ldots, (x_{n_L+n_U})\}$
- Input $x = \{x_L \cup x_U\}, n = n_L + n_U, x_i \in R^d$
- Output $y = \{y_L \cup \widehat{y_U}\}, n = n_L + n_U, y_i \in \{0, 1\}$



min Cut

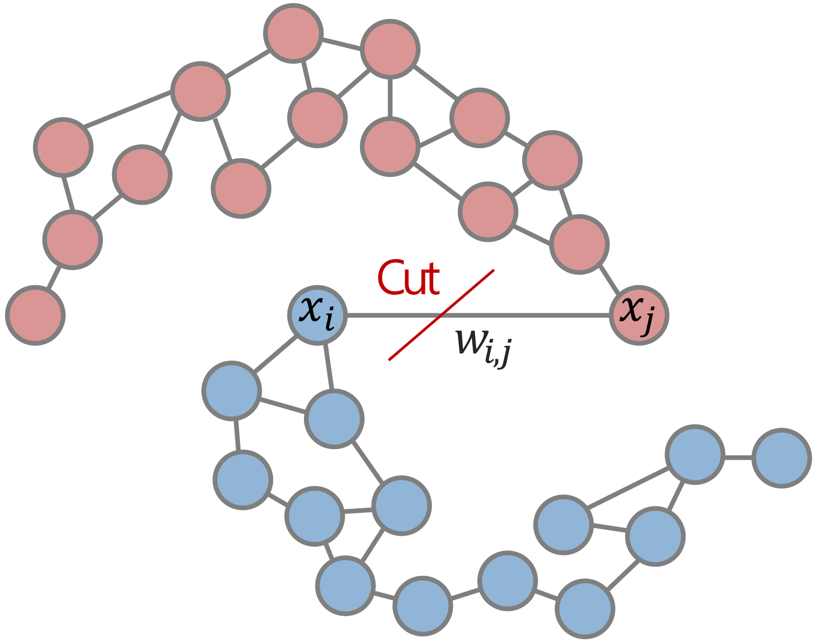
$$cut(set\,1, set2) = \sum_{i \in set\,1,\ j \in set\,2} w_{i,j}$$

서로 다른 set들을 연결하는 edge의 weight합

Label Propagation

Background of Min-cut Algorithm

- Labeled data $\{(x_L, y_L)\} = \{(x_1\ y_1), \ldots, (x_{n_L}\ y_{n_L})\}$
- Unlabeled data $x_U = \{(x_{n_L+1}), \ldots, (x_{n_L+n_U})\}$
- Input $x = \{x_L \cup x_U\}, n = n_L + n_U, x_i \in R^d$
- Output $y = \{y_L \cup \widehat{y_U}\}, n = n_L + n_U, y_i \in \{0, 1\}$



$$min\ \text{Cut}$$

$$cut(set1, set2) = \sum_{i \in set1,\, j \in set2} w_{i,j}$$

서로 다른 set들을 연결하는 edge의 weight합

Label Propagation

Min-cut Algorithm

❖ y_L 이 주어지고, $y_U \in \{0,1\}^{n-n_L}$ 을 도출해보고자 함

discrete

$$\min_{y\in\{0,1\}^{n_L+n_U}} \sum_{i,j=1}^{n_L+n_U} w_{ij}\left|y_i - y_j\right|$$

w_{ij} 크면 $y_i = y_j$ 일 확률이 큼

w_{ij} 작으면 $y_i \neq y_j$ 일 수 있음

$$\min_{y\in\{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(y_i - y_{L_i}\right)^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}\left(y_i - y_j\right)^2$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

레이블링 되지 않은 데이터는
w_{ij} 를 고려하여 레이블링

Label Propagation

Min-cut Algorithm

- y_L 이 주어지고, $y_U \in \{0,1\}^{n-n_L}$ 을 도출해보고자 함

  discrete   an integer program: NP hard

  ➡ **Relaxing to continuous values**

$$\min_{y \in \{0,1\}^{n_L+n_U}} \sum_{i,j=1}^{n_L+n_U} w_{ij} |y_i - y_j|$$

w_{ij} 크면 $y_i = y_j$ 일 확률이 큼

w_{ij} 작으면 $y_i \neq y_j$ 일 수 있음

$$\min_{y \in \{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(y_i - y_{L_i}\right)^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij} \left(y_i - y_j\right)^2$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

레이블링 되지 않은 데이터는
w_{ij} 를 고려하여 레이블링

Label Propagation

Harmonic Solution

- $f(x_i) = y_i \ \ (i = 1, \ldots, n_L)$ 를 만족하는 **harmonic function**을 정의
- Harmonic function은 각 vertex의 label이 이웃한 vertex와 비슷해 지도록 함
  harmonious (similar)

Min-cut

$$\min_{y \in \{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} (y_i - y_{L_i})^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij} (y_i - y_j)^2$$

Relaxing to continuous values

Harmonic solution

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} (f(x_i) - y_{L_i})^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij} (f(x_i) - f(x_j))^2$$

Labeled $i = 1, \ldots, n_L$

Unlabeled $i = n_L + 1, \ldots, n_L + n_U$

$$f(x_i) = y_i$$

$$f(x_i) = \frac{\sum_{j=1}^{n_L+n_U} w_{i,j} f(x_j)}{\sum_{j=1}^{n_L+n_U} w_{i,j}}$$

Label Propagation

Harmonic Solution

- Laplacian matrix를 활용하여 표현가능
- Laplacian matrix = Degree matrix – Adjacency matrix

Harmonic solution

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i) - f(x_j)\right)^2$$

Labeled $i = 1, \ldots, n_L$ **Unlabeled** $i = n_L + 1, \ldots, n_L + n_U$

Graph Laplacian regularization term

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

$$f = (f_L; f_U) = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_L) \\ f(x_{L+1}) \\ \vdots \\ f(x_U) \end{bmatrix}$$

Label Propagation

Learning with Local and Global Consistency

- ❖ Labeled data가 틀린 경우도 있을 수 있음
- ❖ Local consistency, Global consistency를 정의

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

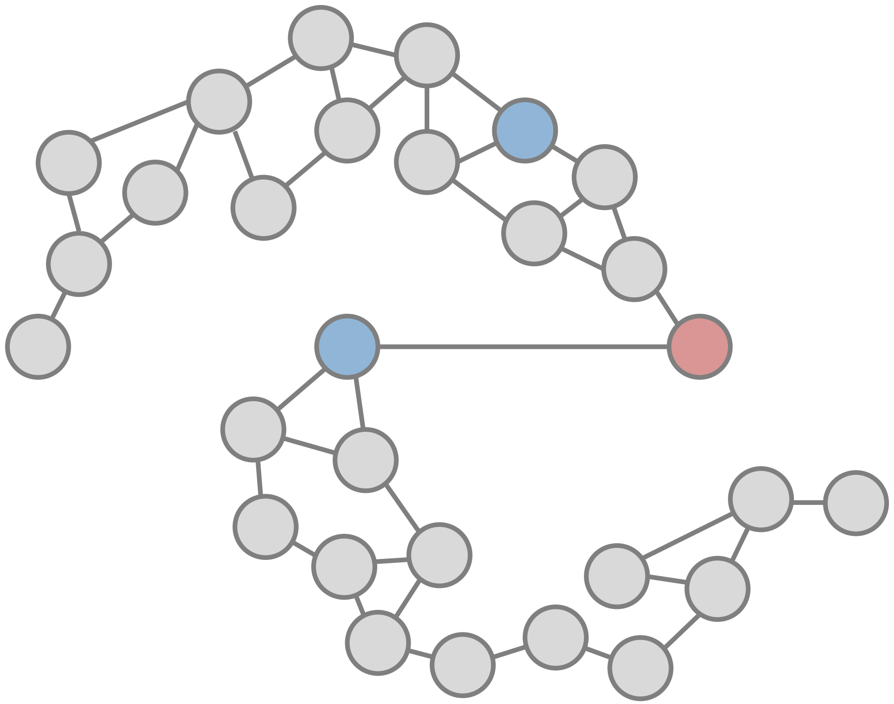
레이블링 되지 않은 데이터는
w_{ij}를 고려하여 레이블링

Local consistency

주변 points들과
동일한 레이블을 갖도록 학습

Global consistency

동일한 구조(cluster, manifold)를
가진다면 동일한 레이블을 갖도록 학습



Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Learning with Local and Global Consistency

❖ Labeled data가 틀린 경우도 있을 수 있음

❖ Local consistency, Global consistency를 정의

$$\min_{f \in \mathbb{R}^{n_L + n_U}} \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

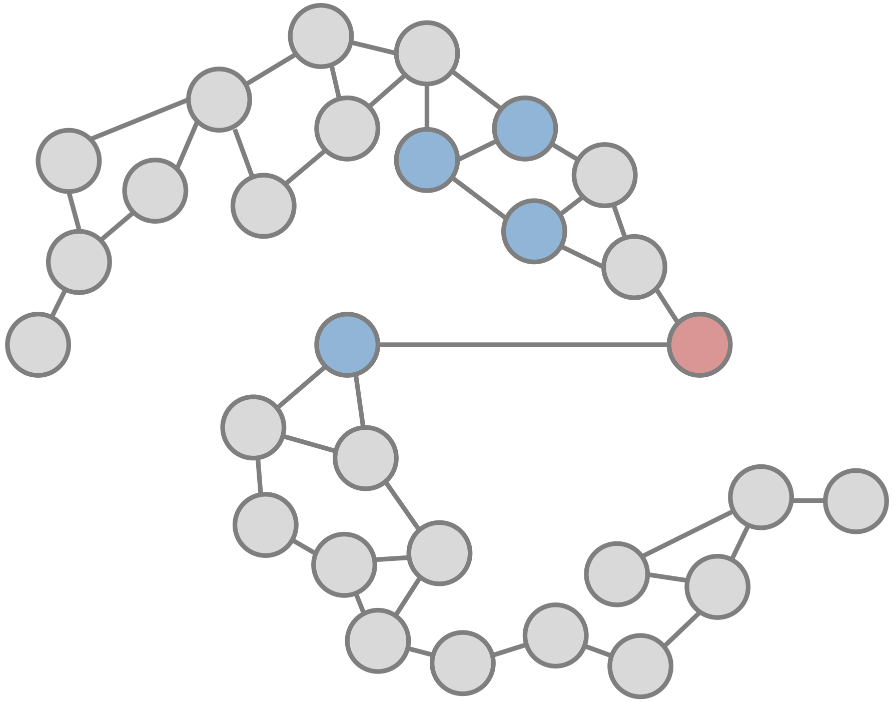
레이블링 되지 않은 데이터는
w_{ij}를 고려하여 레이블링

Local consistency

주변 points들과
동일한 레이블을 갖도록 학습

Global consistency

동일한 구조(cluster, manifold)를
가진다면 동일한 레이블을 갖도록 학습



Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Learning with Local and Global Consistency

- ❖ Labeled data가 틀린 경우도 있을 수 있음
- ❖ Local consistency, Global consistency를 정의

$$\min_{f\in\mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

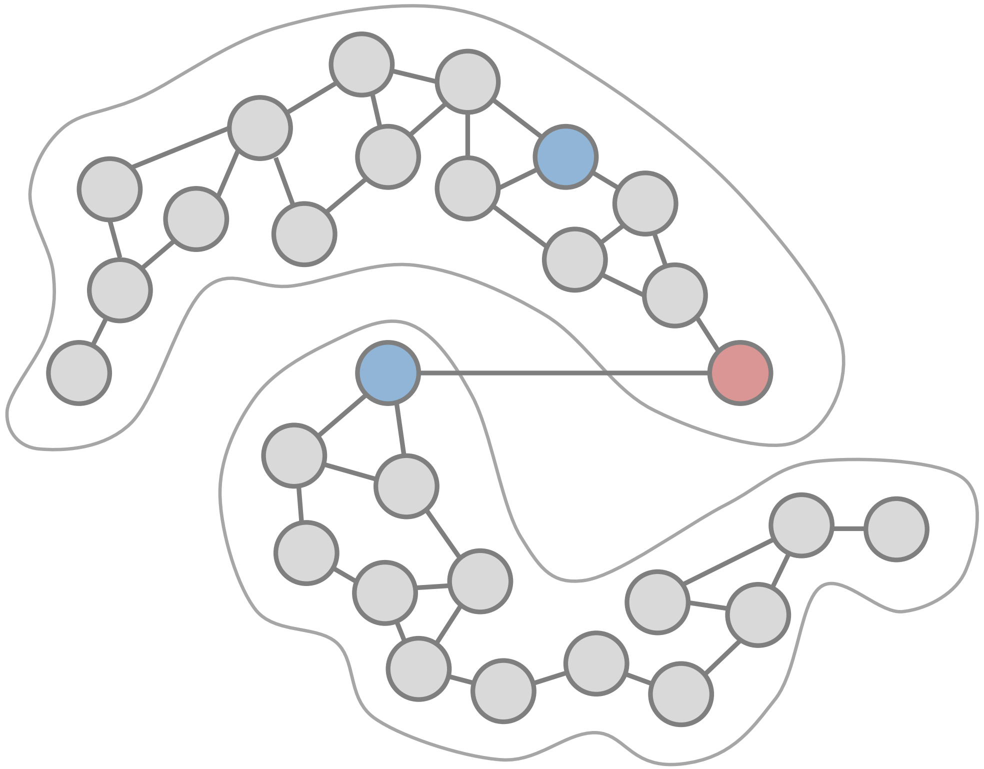
레이블링 되지 않은 데이터는
w_{ij}를 고려하여 레이블링

Local consistency

주변 points들과
동일한 레이블을 갖도록 학습

Global consistency

동일한 구조(cluster, manifold)를
가진다면 동일한 레이블을 갖도록 학습



Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Learning with Local and Global Consistency

- ❖ Labeled data가 틀린 경우도 있을 수 있음
- ❖ Local consistency, Global consistency를 정의

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

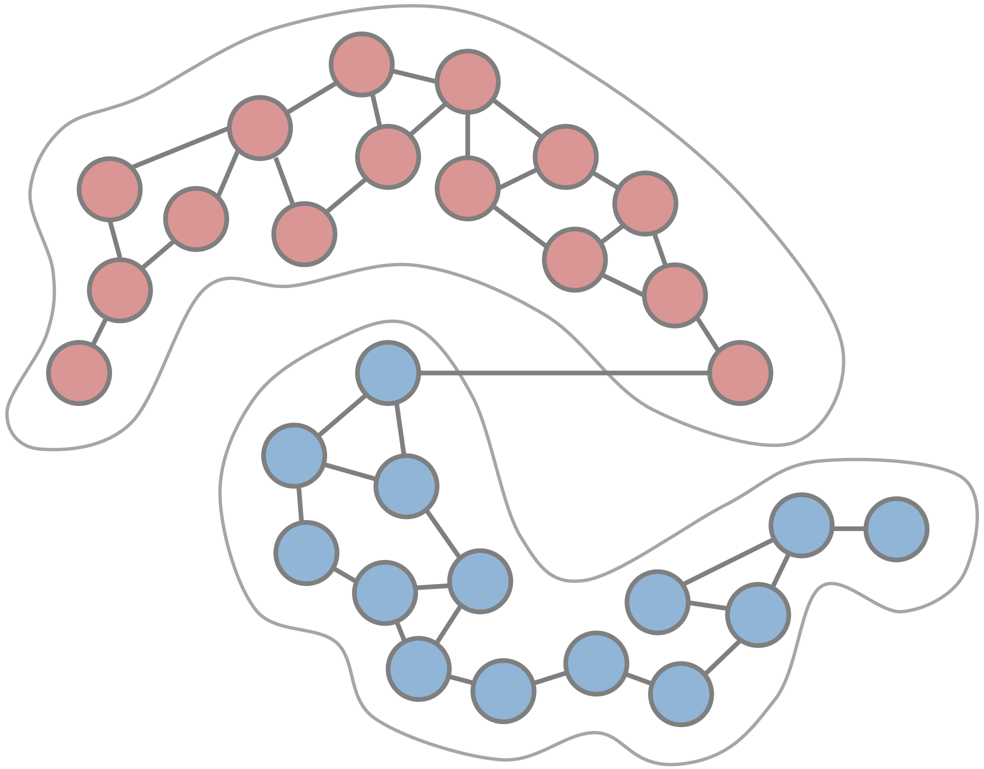
레이블링 되지 않은 데이터는
w_{ij}를 고려하여 레이블링

Local consistency

주변 points들과
동일한 레이블을 갖도록 학습

Global consistency

동일한 구조(cluster, manifold)를
가진다면 동일한 레이블을 갖도록 학습



Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Learning with Local and Global Consistency

- ❖ Labeled data가 틀린 경우도 있을 수 있음
- ❖ Local consistency, Global consistency를 정의

$$\min_{f\in\mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L}\left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Labeled

레이블링 데이터는
실제 레이블과 같도록 학습

Unlabeled

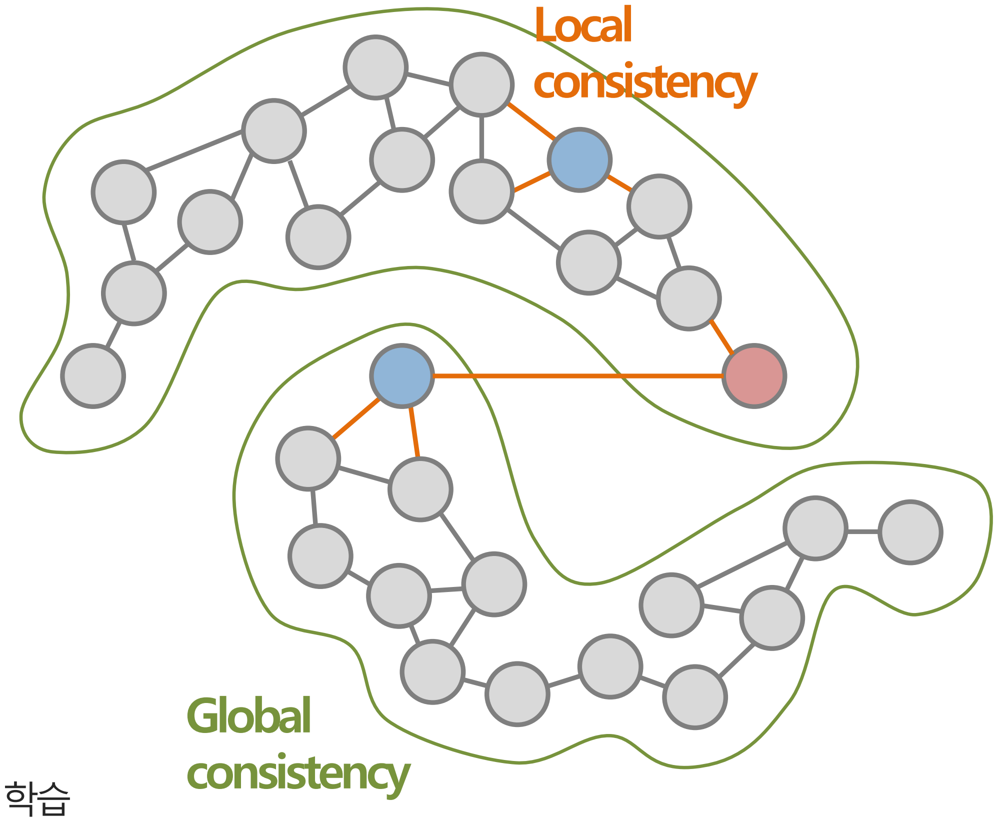
레이블링 되지 않은 데이터는
w_{ij}를 고려하여 레이블링

Local consistency

주변 points들과
동일한 레이블을 갖도록 학습

Global consistency

동일한 구조(cluster, manifold)를
가진다면 동일한 레이블을 갖도록 학습



Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Learning with Local and Global Consistency

- ❖ Labeled data가 틀린 경우도 있을 수 있음
- ❖ Local consistency, Global consistency를 정의 ➡ Add penalty μ (Smoothness)

Local and Global Consistency

$$\min_{f\in\mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)\left(f(x_i) - y_{L_i}\right)^2 + \mu\, f^T L f$$

Local consistency **Global consistency**

μ 크면 Labeled data y_L가 바뀌어 예측될 확률 큼

μ 작으면 Labeled data y_L를 보존할 확률 큼

$$\min_{f\in\mathbb{R}^{n_L+n_U}} (1-\mu)(f-y)^T(f-y) + \mu\, f^T L f$$

$$\frac{\partial L}{\partial f} = (f-y) - \mu(f-y) + \mu\, f^T L f = 0$$

$$f = \beta(I+\alpha L)^{-1} y \qquad \alpha = \frac{1}{1+\mu} \qquad \beta = \frac{\mu}{1+\mu} \qquad f = (f_L; f_U) = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_L) \\ f(x_{L+1}) \\ \vdots \\ f(x_U) \end{bmatrix}$$

Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Graph Convolutional Neural Network

Graph Convolutional Network

Min-cut

$$\min_{y\in\{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} (y_i - y_{L_i})^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}(y_i - y_j)^2$$

Harmonic Solution

$$\min_{f\in\mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} (f(x_i) - y_{L_i})^2 + f^T L f$$

Local and Global Consistency

$$\min_{f\in\mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)(f(x_i) - y_{L_i})^2 + \mu f^T L f$$

Pairwise Constraints

$$\min_{f\in\mathbb{R}^{n_L+n_U}} (1-\mu)\|(f - y)\|^2 + \mu f^T L f$$

Labeled : Supervised loss

Unlabeled : Semi-Supervised loss

Graph-Based

Semi-Supervised Learning

Graph Convolutional Networks

Graph Convolutional Neural Networks

Semi-Supervised Classification with Graph Convolutional networks

❖ GCN : CNN구조가 가진 특징을 반영하여 graph에 적용가능한 GNN구조를 제안

Published as a conference paper at ICLR 2017

SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS

Thomas N. Kipf
University of Amsterdam
T.N.Kipf@uva.nl

Max Welling
University of Amsterdam
Canadian Institute for Advanced Research (CIFAR)
M.Welling@uva.nl

ABSTRACT

We present a scalable approach for semi-supervised learning on graph-structured data that is based on an efficient variant of convolutional neural networks which operate directly on graphs. We motivate the choice of our convolutional architecture via a localized first-order approximation of spectral graph convolutions. Our model scales linearly in the number of graph edges and learns hidden layer representations that encode both local graph structure and features of nodes. In a number of experiments on citation networks and on a knowledge graph dataset we demonstrate that our approach outperforms related methods by a significant margin.

Graph Convolutional Neural Networks

Limitations of Label Propagation

❖ **Smoothness assumption**

- 서로 가까운 데이터(points)는 같은 레이블일 확률이 높음
- 유사도(Similarity)이외에 다른 정보를 담을 수 없음

Min-cut

$$\min_{y \in \{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(y_i - y_{L_i}\right)^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}\left(y_i - y_j\right)^2$$

Harmonic Solution

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$

Local and Global Consistency

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)\left(f(x_i) - y_{L_i}\right)^2 + \mu f^T L f$$

Labeled : Supervised loss

Unlabeled : Semi-Supervised loss

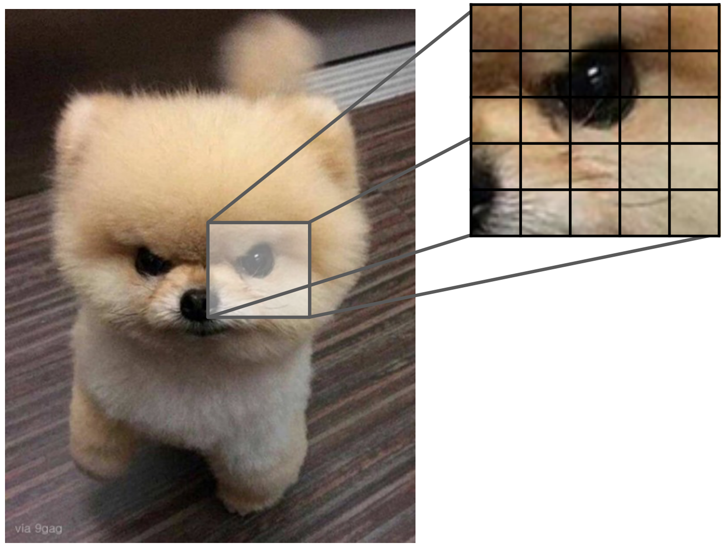
Graph Laplacian regularization term을 없애자

Convolutional Neural Networks

Convolutional Filters

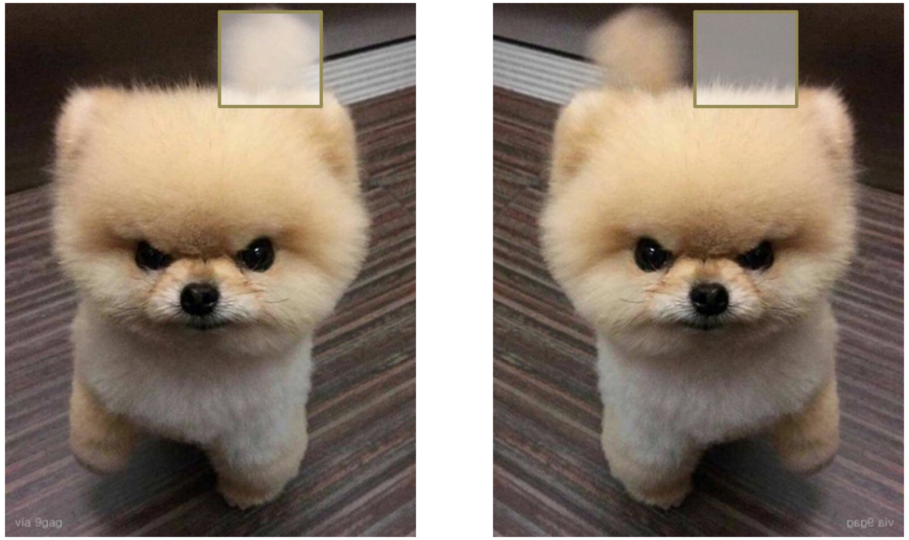
❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님

❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음



Spatially local correlation

→ **Sparse connection**



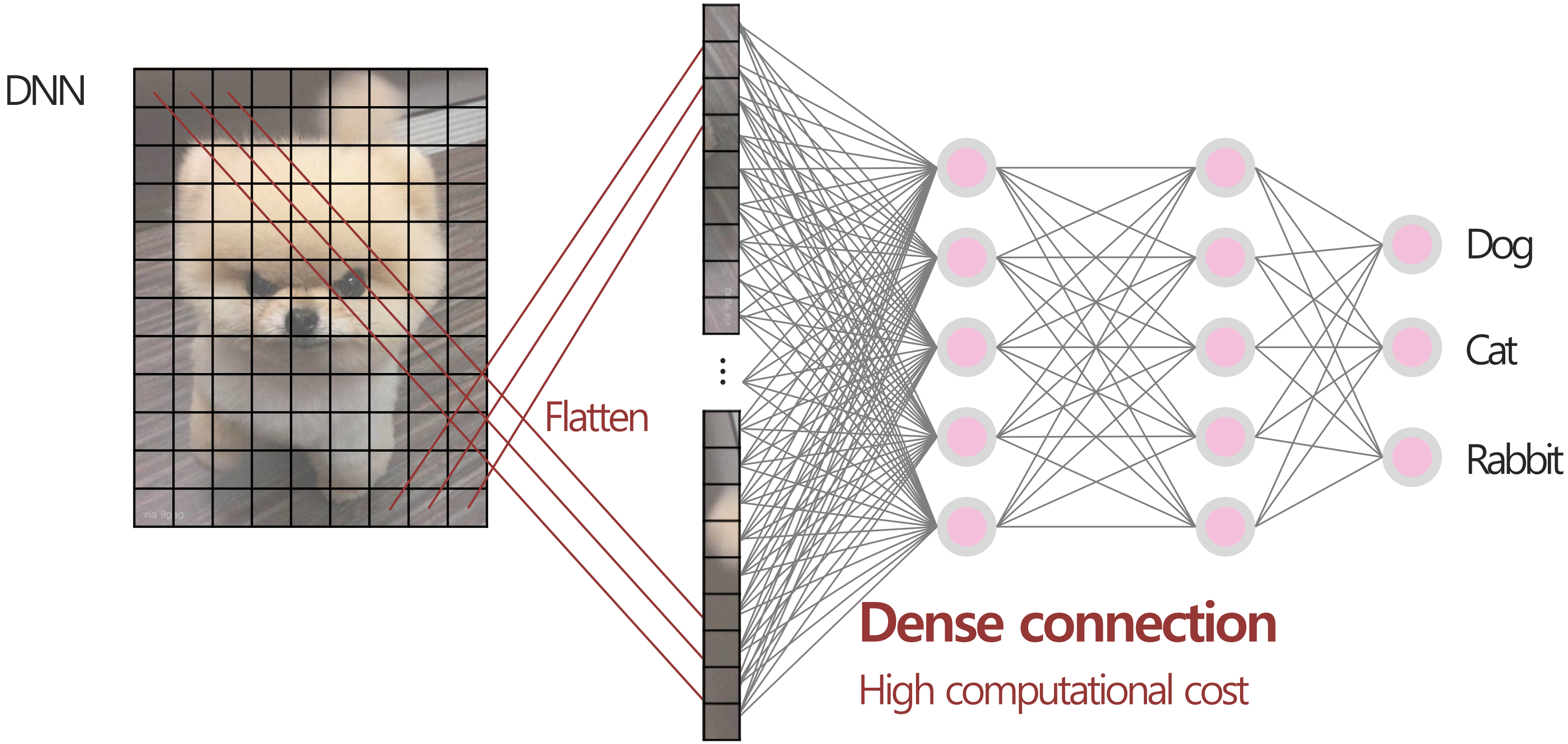
Translation invariance

→ **Weight sharing**

Convolutional Neural Networks

Limitations of Deep Neural Networks

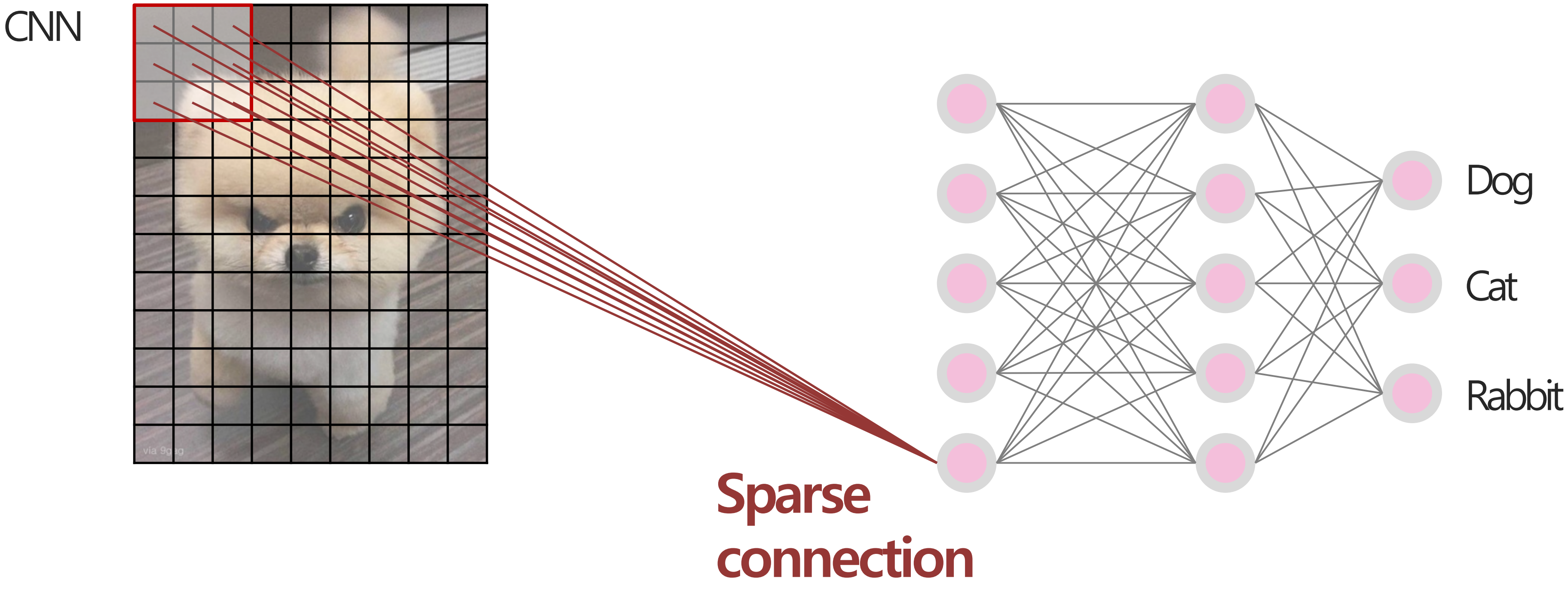
- ❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- ❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음



Convolutional Neural Networks

Convolutional Neural Networks Characteristics : Sparse Connection

- Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음

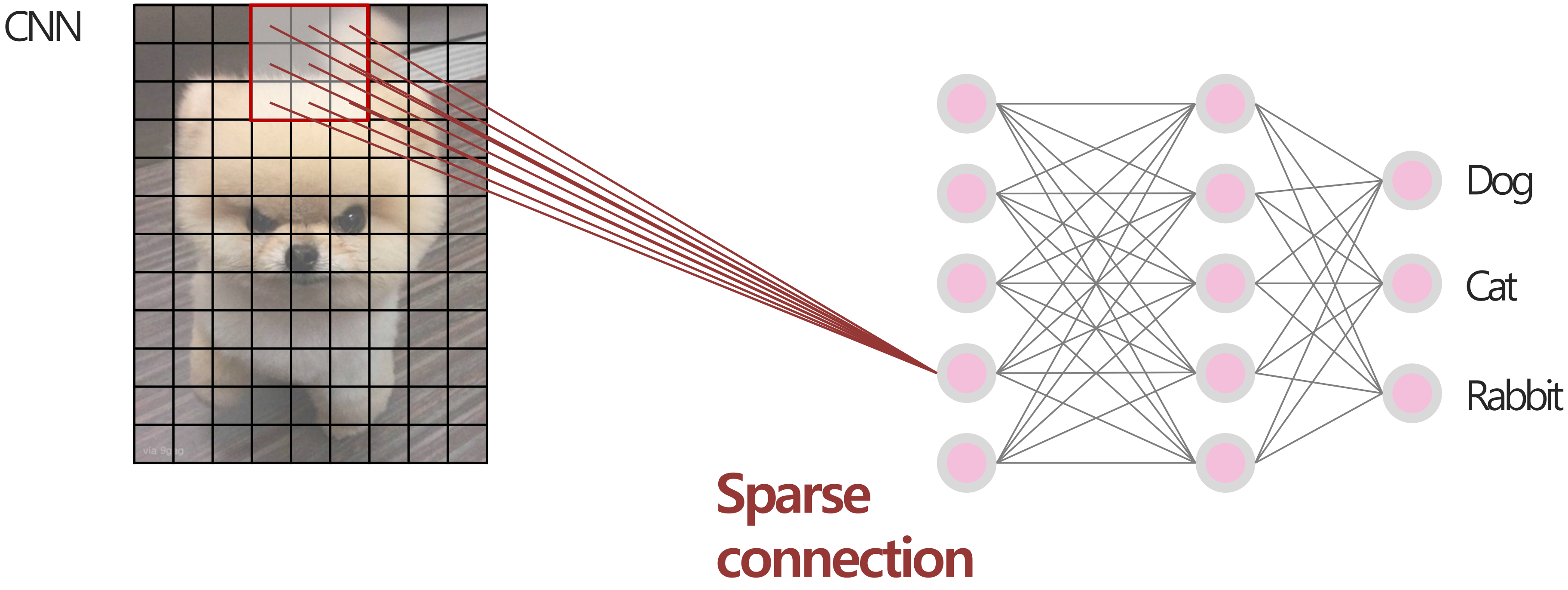


http://dongascience.donga.com/news.php?idx=35340

Convolutional Neural Networks

Convolutional Neural Networks Characteristics : Sparse Connection

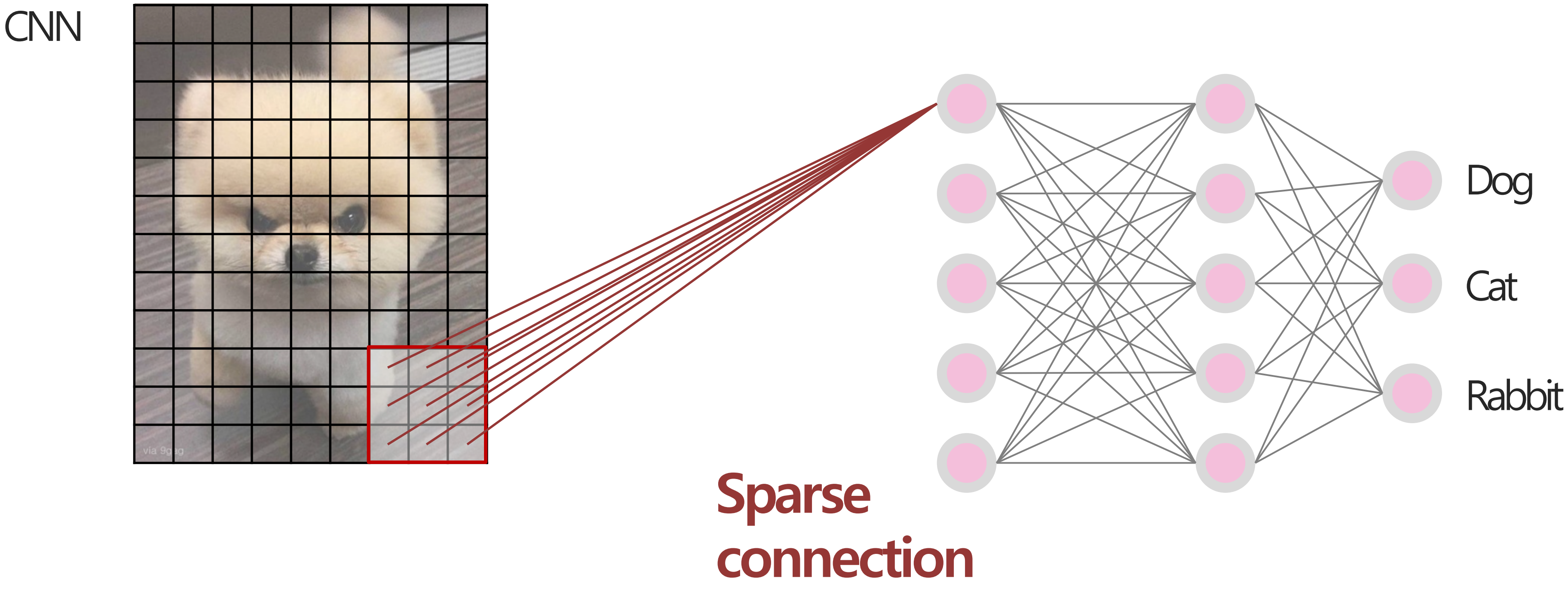
- ❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- ❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음



Convolutional Neural Networks

Convolutional Neural Networks Characteristics : Sparse Connection

- ❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- ❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음



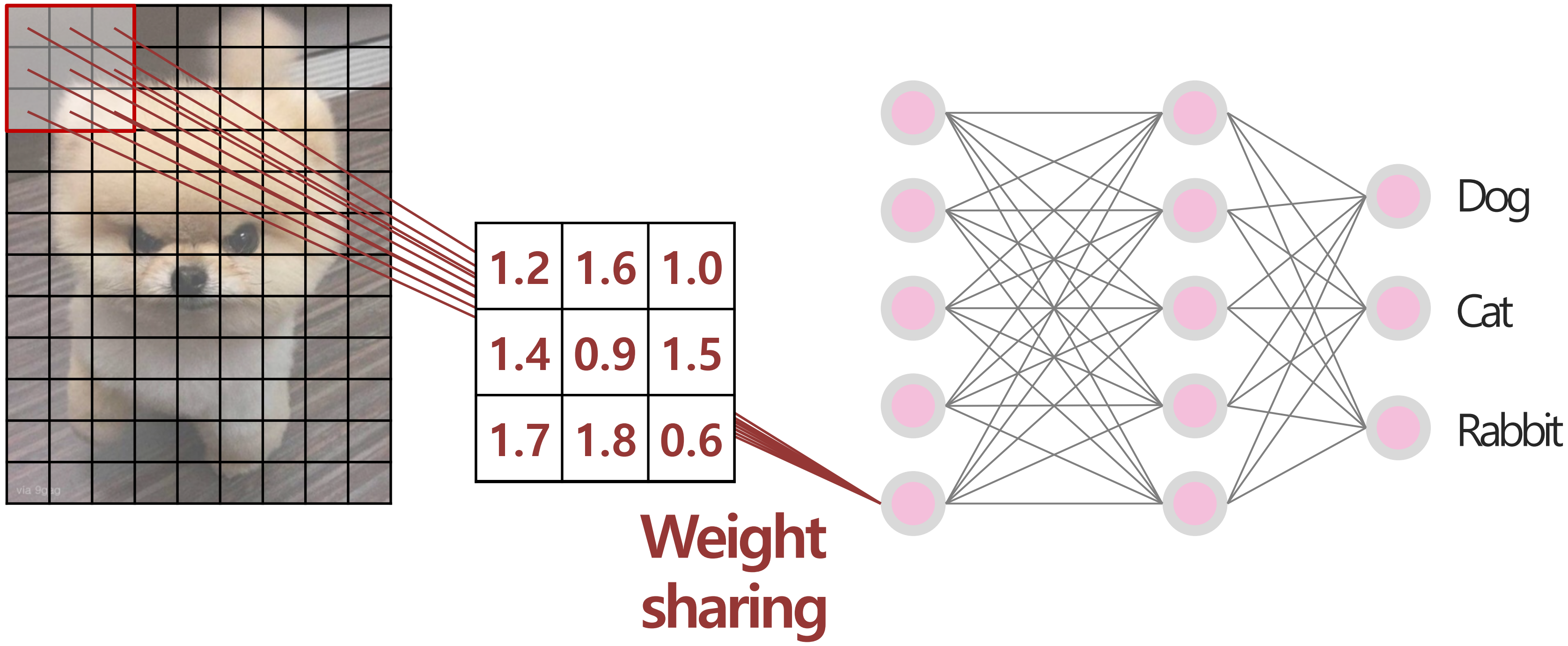
http://dongascience.donga.com/news.php?idx=35340

Convolutional Neural Networks

Convolutional Neural Networks Characteristics : Weight Sharing

- ❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- ❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음

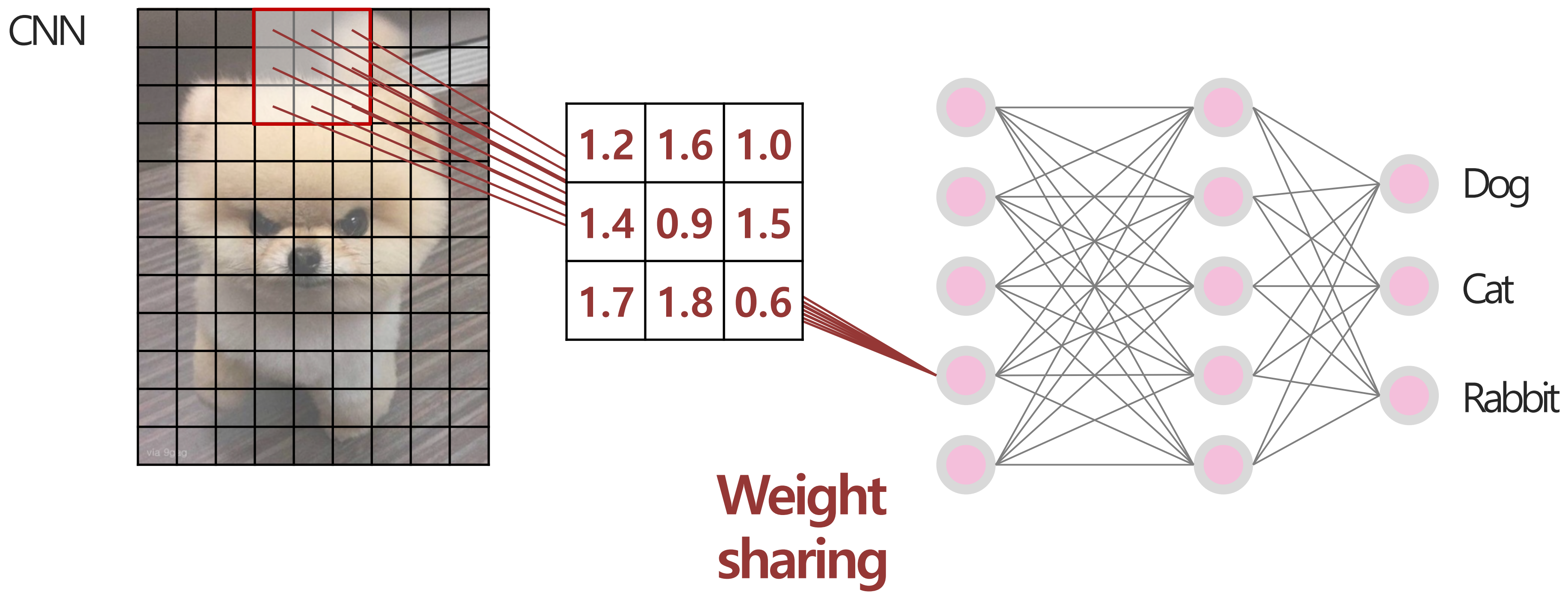
CNN



Convolutional Neural Networks

Convolutional Neural Networks Characteristics : Weight Sharing

- ❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- ❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음

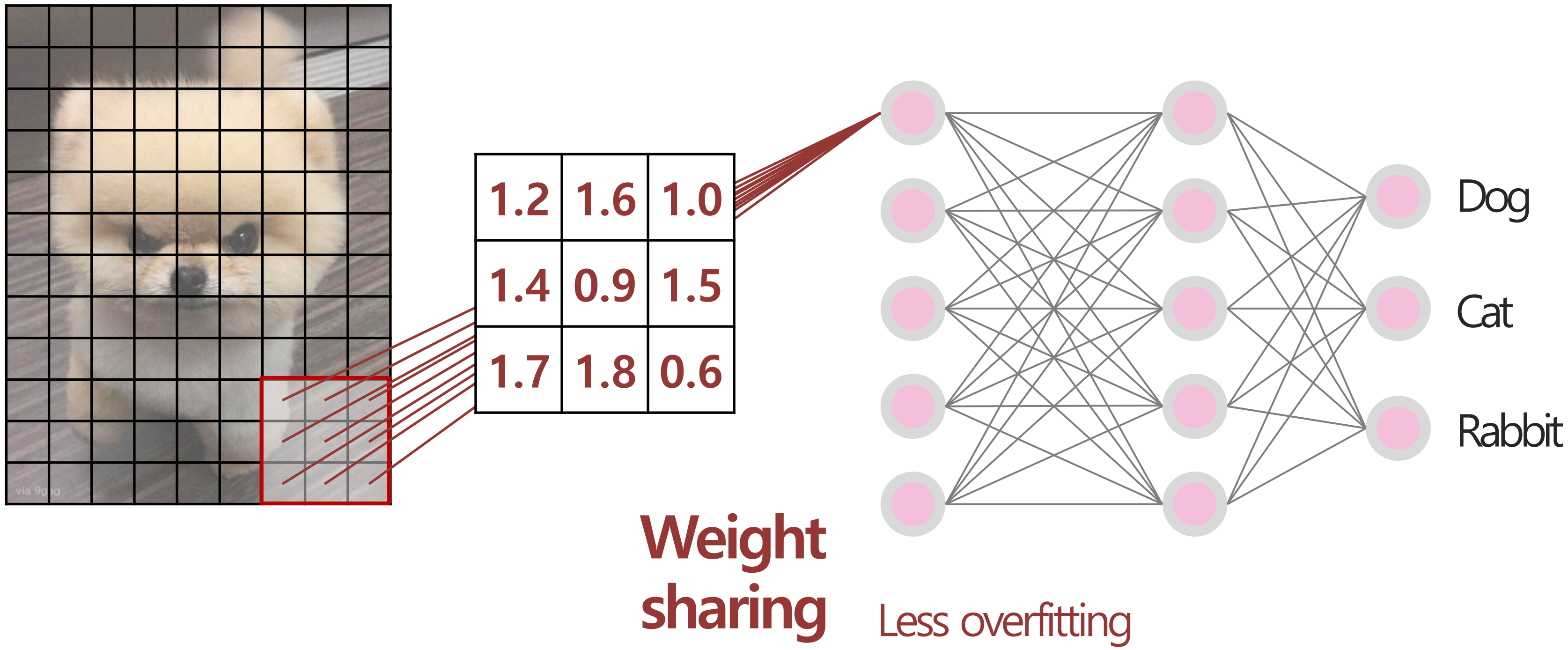


Convolutional Neural Networks

Convolutional Neural Networks Characteristics : Weight Sharing

- ❖ Spatially local correlation : 인접 변수(pixel)간 높은 상관관계 지님
- ❖ Translation invariance : 부분적 특성(e.g., 눈, 귀, 꼬리)은 고정된 위치에 등장하지 않음

CNN



Less overfitting

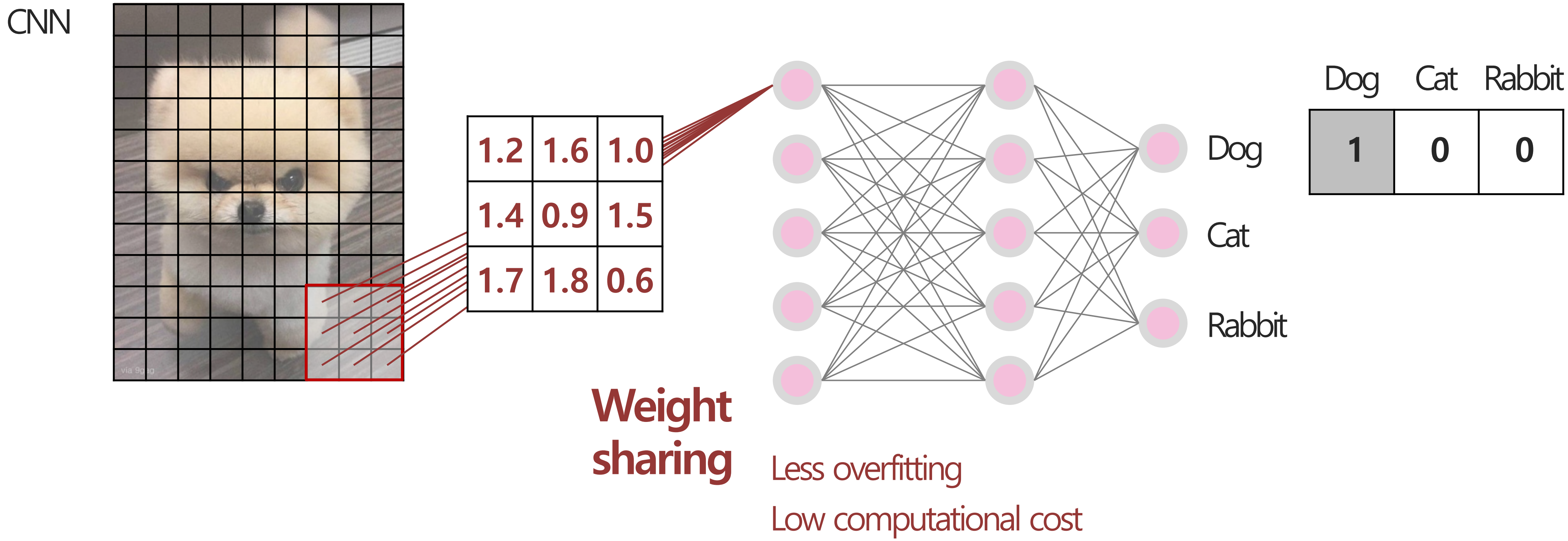
Low computational cost

Convolutional Neural Networks

Feature Extraction of Convolutional Neural Networks

❖ CNN에서의 feature extraction 대상은 입력 이미지

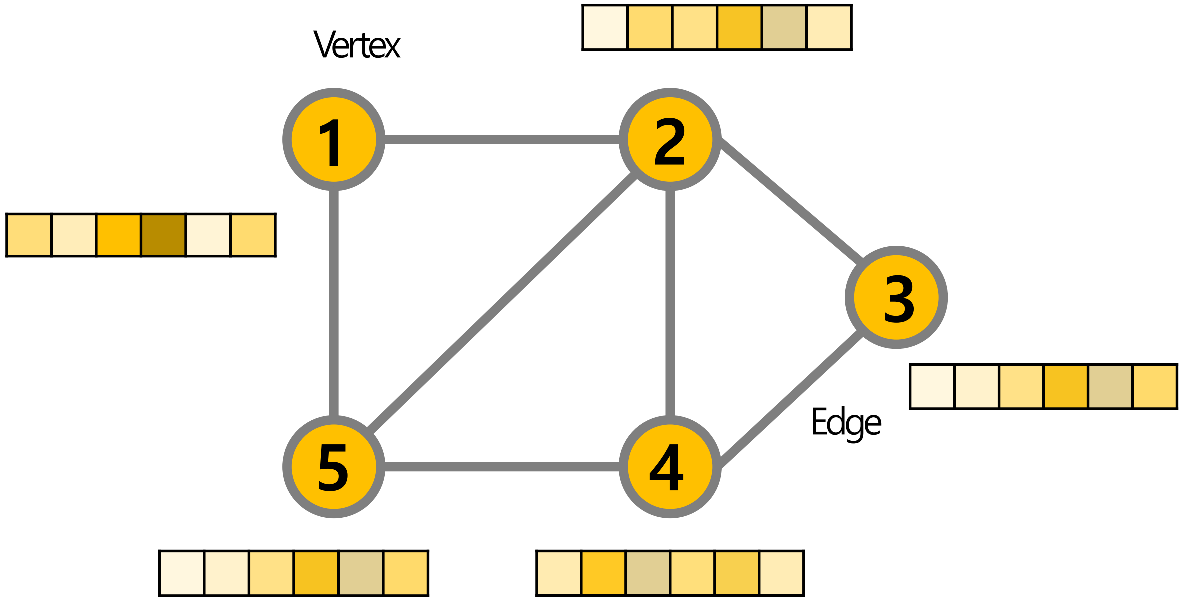
❖ 각 convolution 연산을 통해 적절한 activation map을 형성하고, 최종적으로 예측을 수행



Graph Convolutional Neural Networks

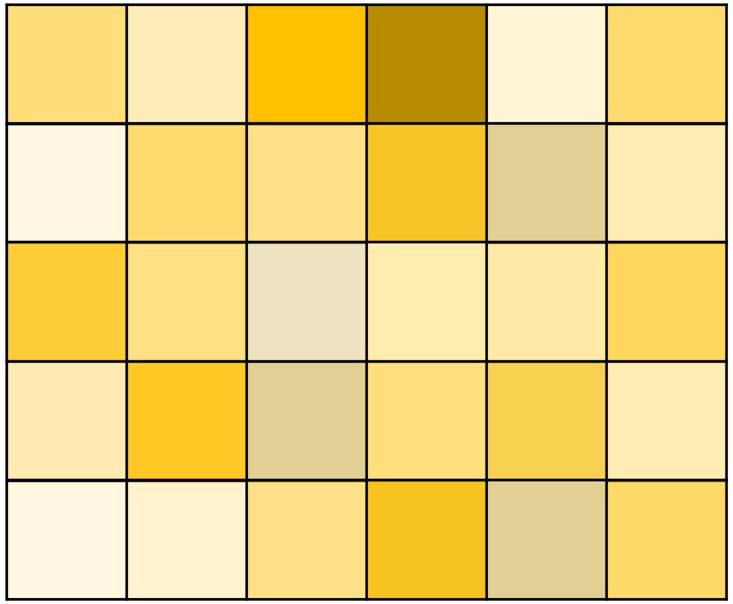
Feature Extraction of Graph Convolutional Networks

❖ GCN에서의 feature extraction 대상은 **Vertex(node)**에 대한 정보 **"Node-feature matrix"**



Node – Feature Matrix

$X \in R^{n \times f}$



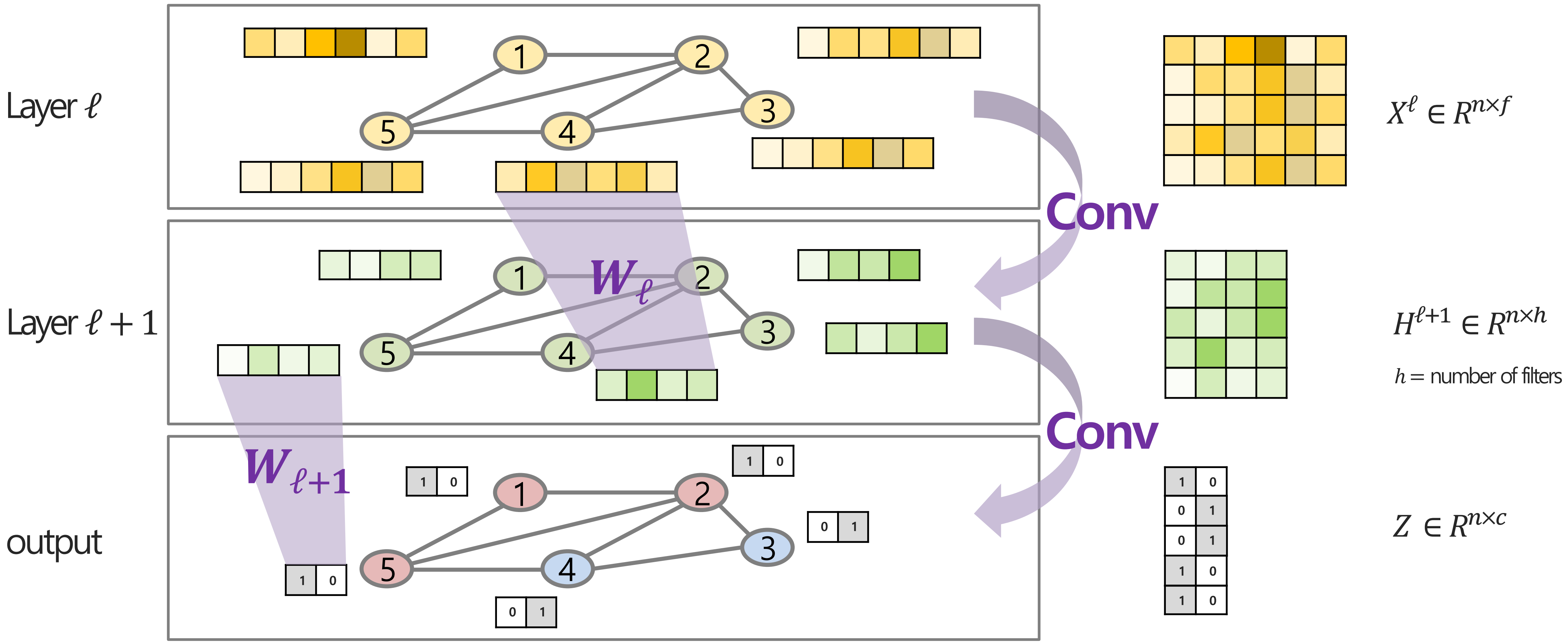
Node – Class Matrix

$Y \in R^{n \times c}$

1	0
0	1
0	1
1	0
1	0

Graph Convolutional Neural Networks

Graphical Overview of GCN



Graph Convolutional Neural Networks

Graphical Overview of GCN

Class 1
Class 2
Unlabeled

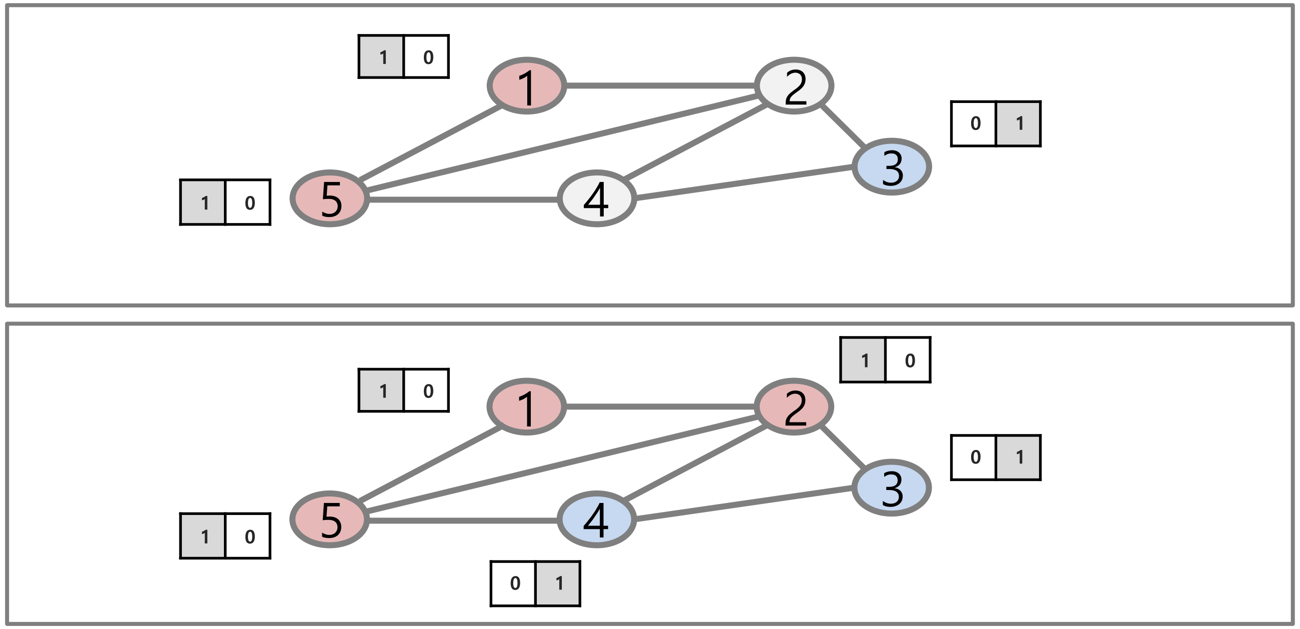
$$Z = \sigma' \left(\hat{A}\sigma\left(\hat{A}X^{\ell}W^{\ell} + b^{\ell}\right)W^{\ell+1} + b^{\ell+1}\right)$$

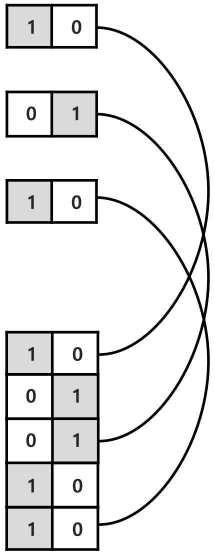
Loss function $$L = -\sum_{l \in Y_L} \sum_{c=1}^{C} Y_{lc} \ln(Z_{lc})$$

Cross-entropy error over all labeled data

ground truth

output





$y_L \in R^{n_L \times c}$

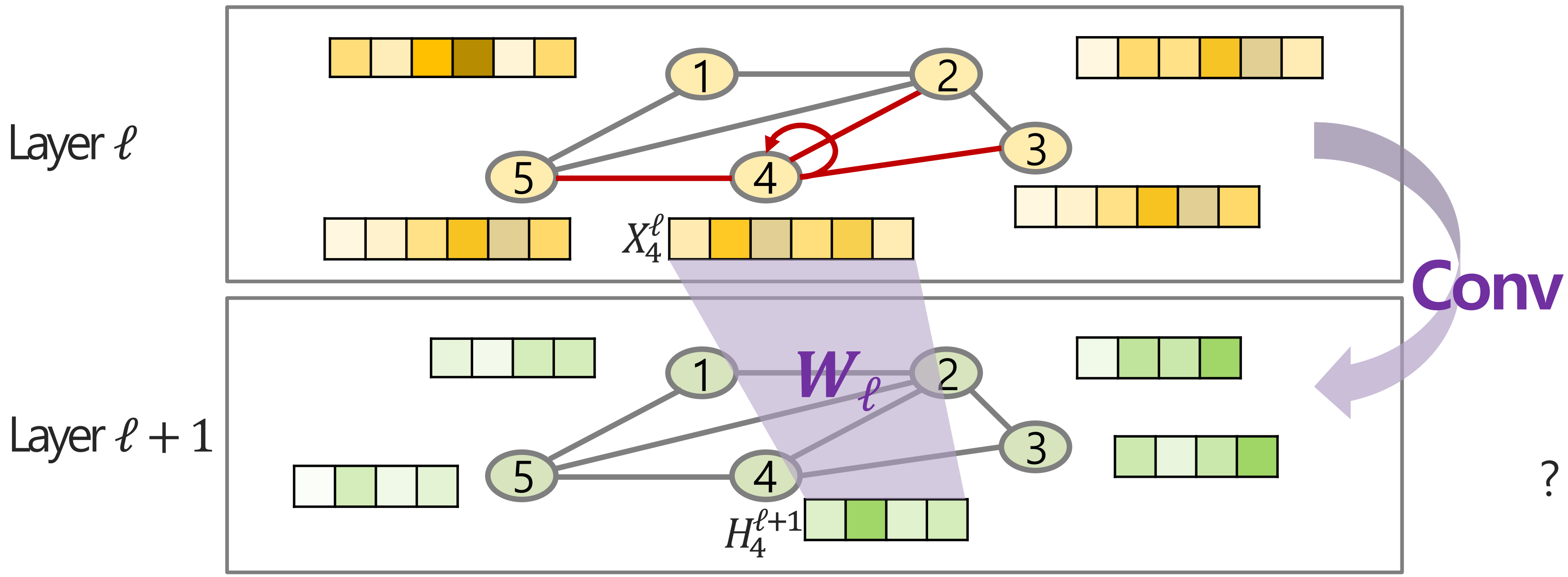
$Z \in R^{n \times c}$

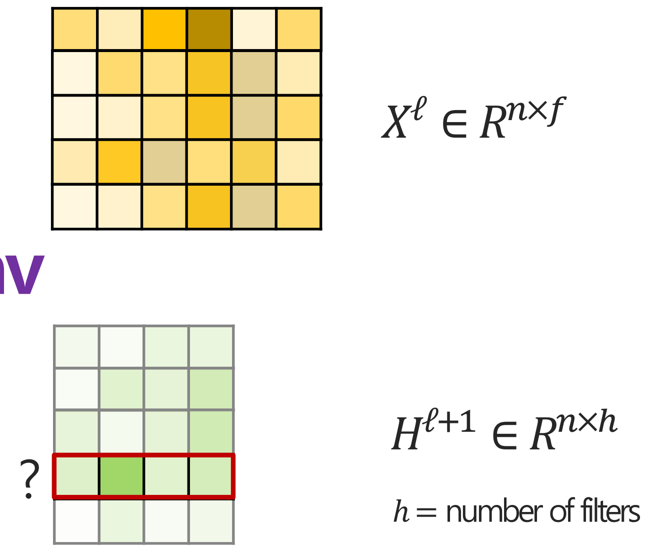
$Z = \sigma'(\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})W^{\ell+1} + b^{\ell+1})$

$H^{\ell+1}$

Graph Convolutional Neural Networks

GCN Mechanism : Layer ℓ to Layer $\ell+1$





자기 자신 이웃들에 대한 정보

$$H_4^{\ell+1} = \sigma(X_4^{\ell}W^{\ell} + X_2^{\ell}W^{\ell} + X_3^{\ell}W^{\ell} + X_5^{\ell}W^{\ell} + b^{\ell})$$

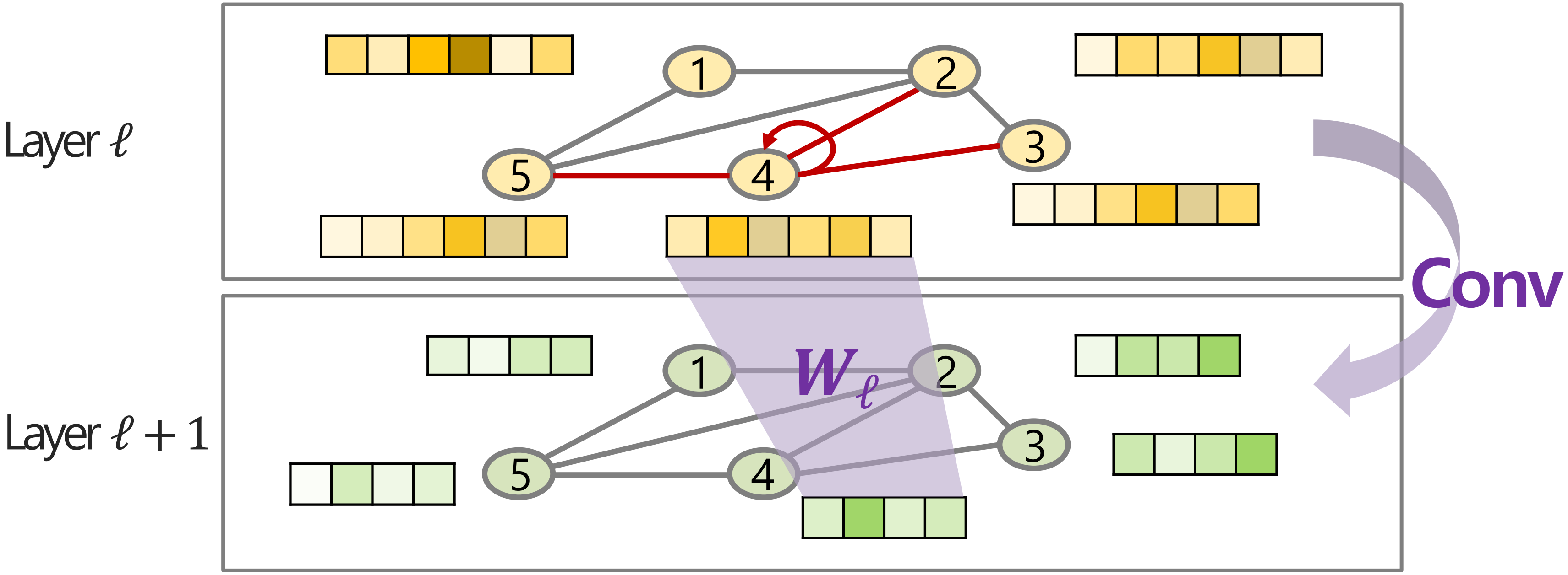
$$= \sigma(\textstyle\sum_{j \in N_{(4)}} X_j^{\ell}W^{\ell} + b^{\ell})$$

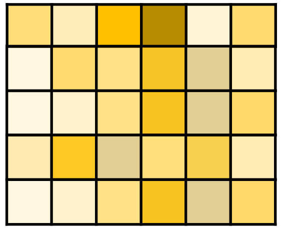
$$Z = \sigma'(\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})W^{\ell+1} + b^{\ell+1})$$

$H^{\ell+1}$

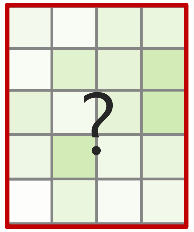
Graph Convolutional Neural Networks

GCN Mechanism : Layer ℓ to Layer $\ell+1$





$X^{\ell} \in R^{n \times f}$



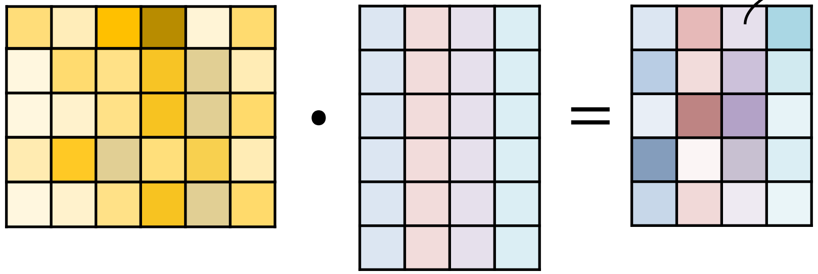
$H^{\ell+1} \in R^{n \times h}$

$h =$ number of filters

$$H^{\ell+1} = \sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})$$

activation map

X^{ℓ} 5×6 W^{ℓ} 6×4 activation map 5×4



첫번째 node-feature에 대해
세번째 filter를 적용시킨
activation map value

$$Z = \sigma'(\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})W^{\ell+1} + b^{\ell+1})$$

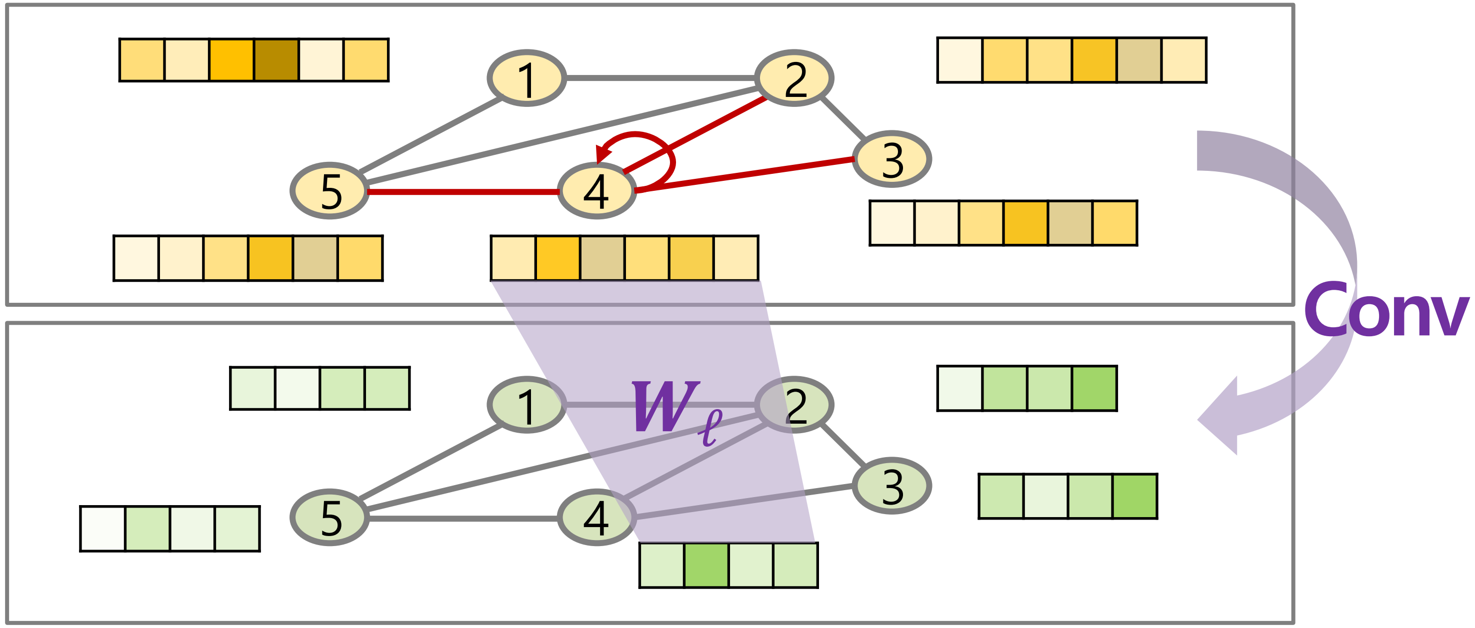
$H^{\ell+1}$

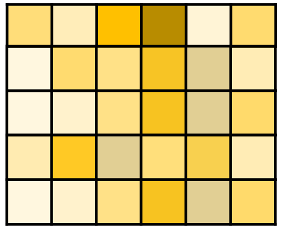
Graph Convolutional Neural Networks

GCN Mechanism : Layer ℓ to Layer $\ell+1$

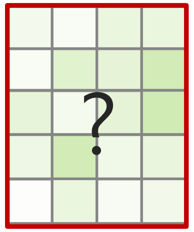
Layer ℓ

Layer $\ell+1$





$X^{\ell} \in R^{n \times f}$



$H^{\ell+1} \in R^{n \times h}$

$h =$ number of filters

$$H^{\ell+1} = \sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})$$

$$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$$

$$\hat{A} = \tilde{D}^{-1/2}(I+A)\tilde{D}^{-1/2}$$

$\tilde{D}^{-1/2}$

5×5

$1/\sqrt{3}$	0	0	0	0
0	$1/\sqrt{5}$	0	0	0
0	0	$1/\sqrt{3}$	0	0
0	0	0	$1/\sqrt{4}$	0
0	0	0	0	$1/\sqrt{4}$

$(I+A)$

5×5

1	1	0	0	1
1	1	1	1	1
0	1	1	1	0
0	1	1	1	1
1	1	0	1	1

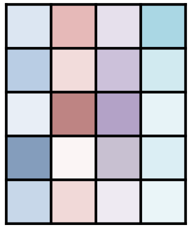
$\tilde{D}^{-1/2}$

5×5

$1/\sqrt{3}$	0	0	0	0
0	$1/\sqrt{5}$	0	0	0
0	0	$1/\sqrt{3}$	0	0
0	0	0	$1/\sqrt{4}$	0
0	0	0	0	$1/\sqrt{4}$

activation map

5×4



$$Z=\sigma'(\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell}+b^{\ell})W^{\ell+1}+b^{\ell+1})$$

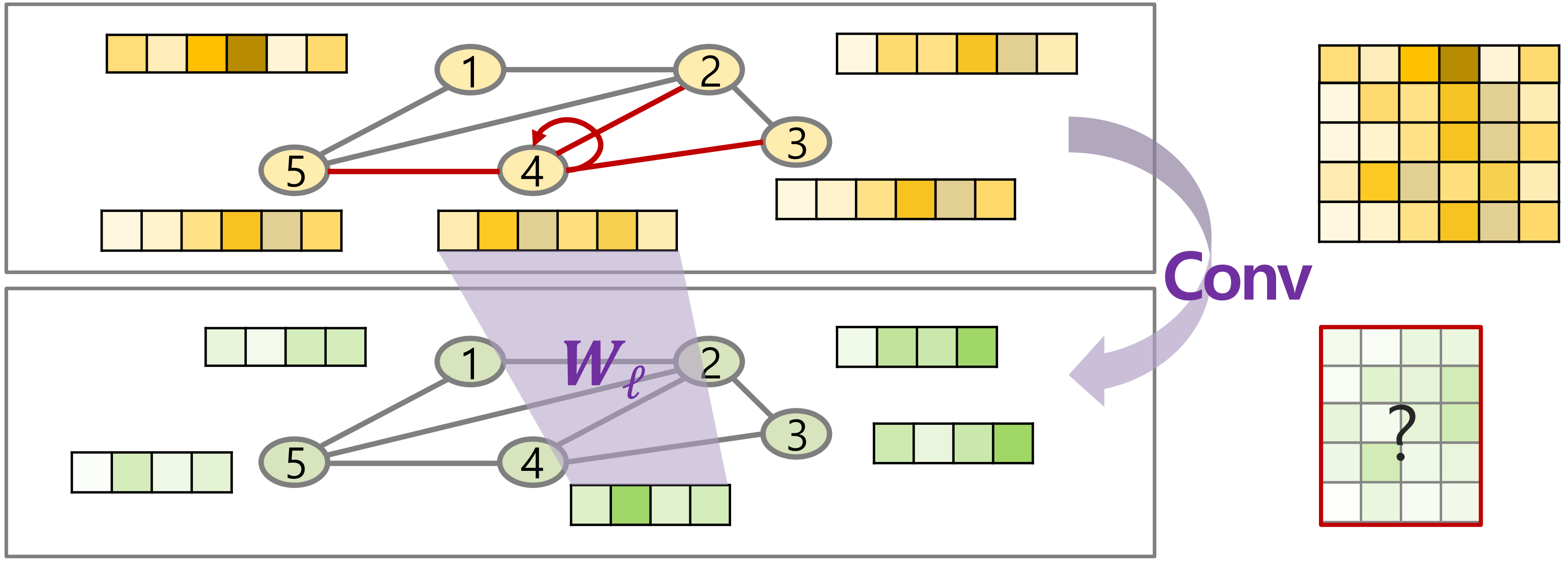
$H^{\ell+1}$

Graph Convolutional Neural Networks

GCN Mechanism : Layer ℓ to Layer $\ell+1$

Layer ℓ

Layer $\ell+1$



$X^{\ell} \in R^{n \times f}$

$H^{\ell+1} \in R^{n \times h}$

$h =$ number of filters

$$H^{\ell+1}=\sigma(\hat{A}X^{\ell}W^{\ell}+b^{\ell})$$

$$\hat{A}=\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$$

$$\hat{A}=\tilde{D}^{-1/2}(I+A)\tilde{D}^{-1/2}$$

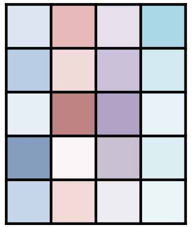
$\hat{A}$

5×5

1/3	1/3	0	0	1/3
1/5	1/5	1/5	1/5	1/5
0	1/3	1/3	1/3	0
0	1/4	1/4	1/4	1/4
1/4	1/4	0	1/4	1/4

activation map

5×4

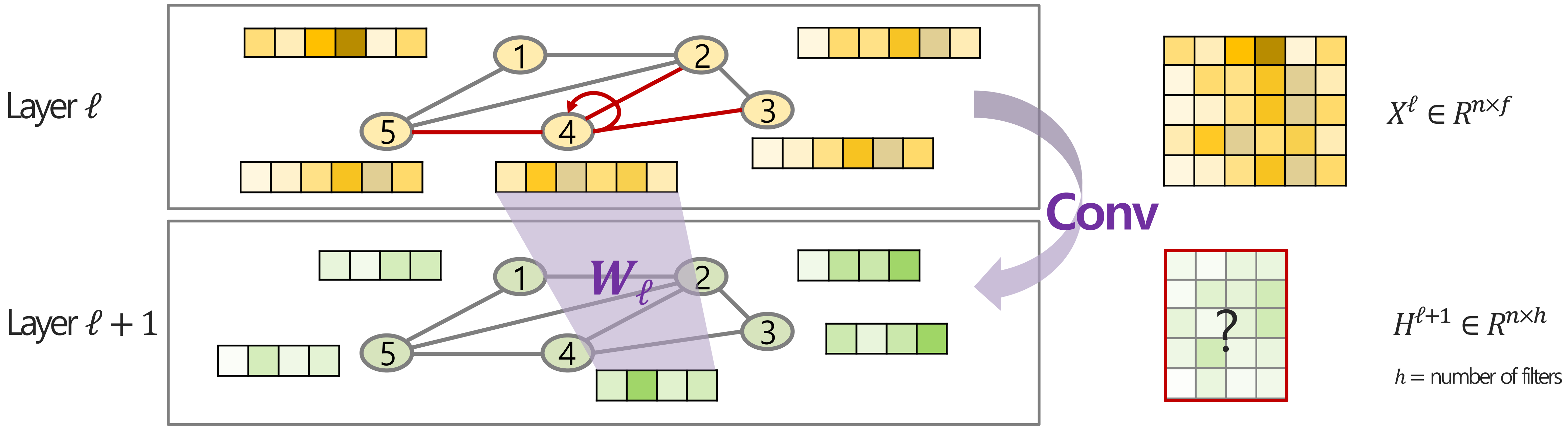


Graph Convolutional Neural Networks

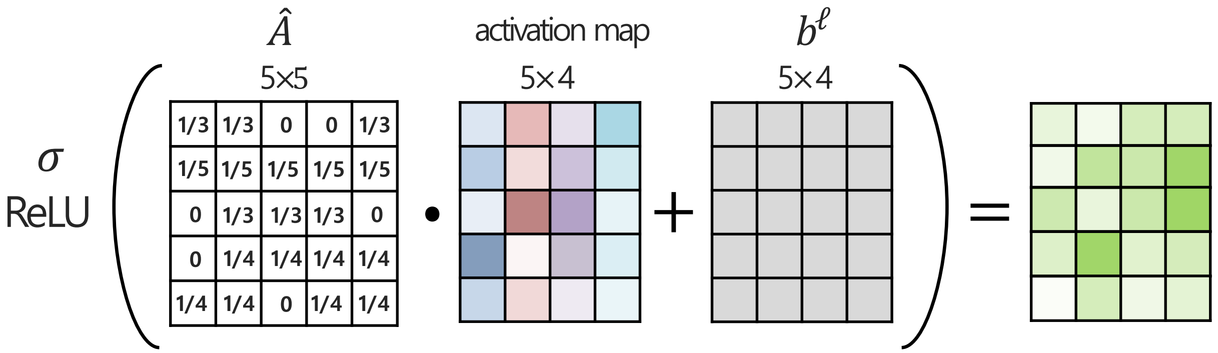
$$Z = \sigma'(\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})W^{\ell+1} + b^{\ell+1})$$

$H^{\ell+1}$

GCN Mechanism : Layer ℓ to Layer $\ell+1$



$$H^{\ell+1} = \sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})$$



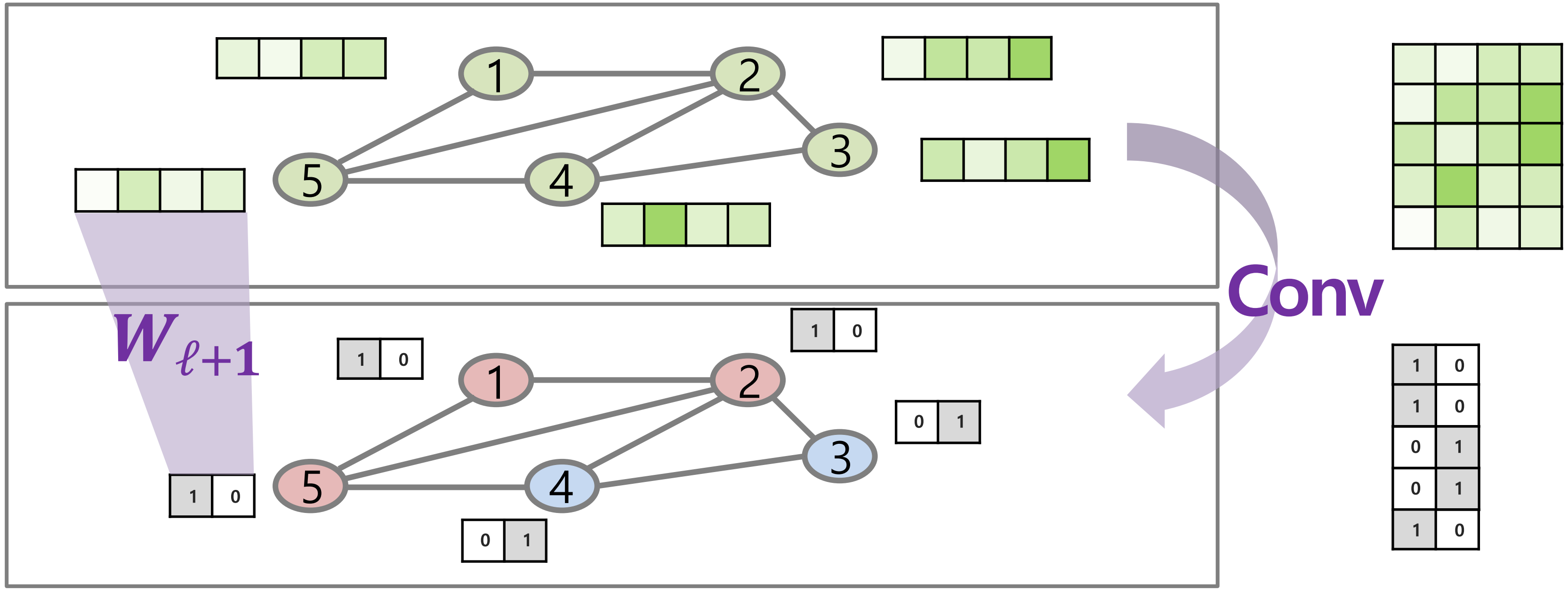
$Z = \sigma'(\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})W^{\ell+1} + b^{\ell+1})$

Graph Convolutional Neural Networks

GCN Mechanism : Layer $\ell + 1$ to Output Layer

Layer $\ell + 1$

output



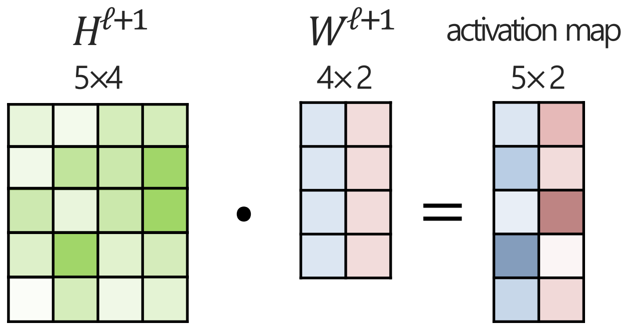
$H^{\ell+1} \in R^{n \times h}$

$h =$ number of filters

$Z \in R^{n \times c}$

$$Z = \sigma'(\underline{\hat{A}H^{\ell+1}W^{\ell+1}} + b^{\ell+1})$$

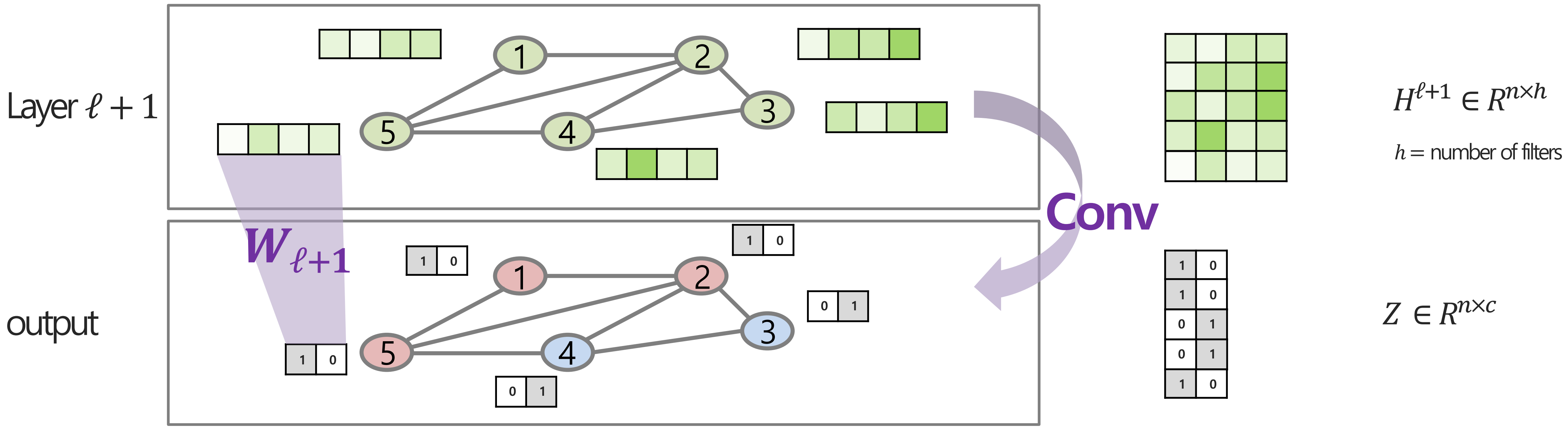
activation map



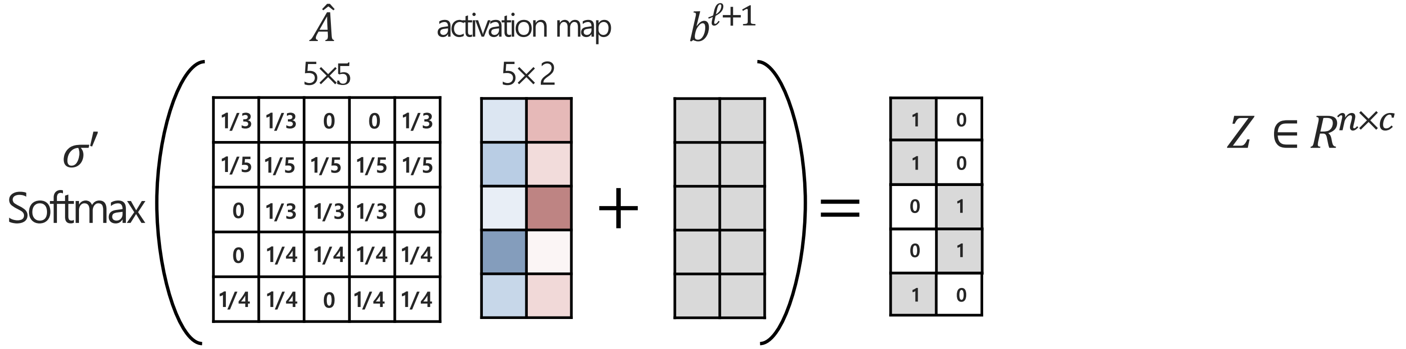
$Z = \sigma' (\hat{A}\sigma(\hat{A}X^{\ell}W^{\ell} + b^{\ell})W^{\ell+1} + b^{\ell+1})$

Graph Convolutional Neural Networks

GCN Mechanism : Layer $\ell + 1$ to Output Layer

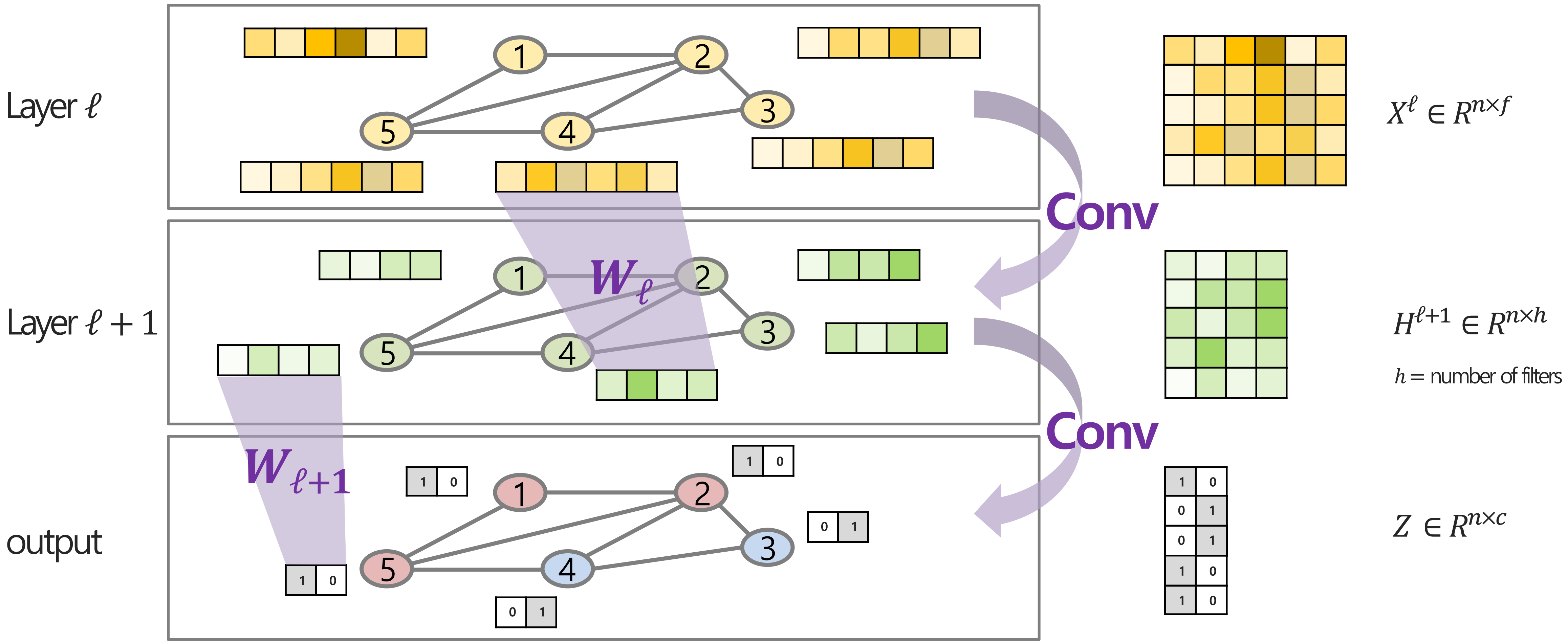


$Z = \sigma' (\hat{A}H^{\ell+1}W^{\ell+1} + b^{\ell+1})$



Graph Convolutional Neural Networks

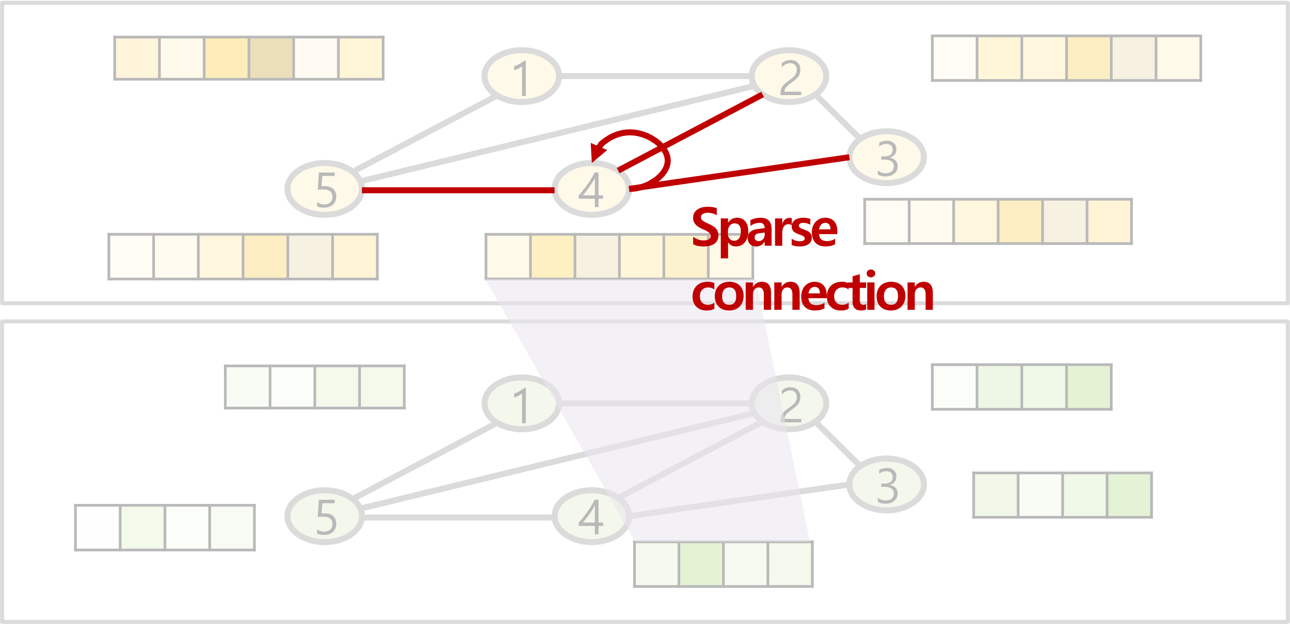
Graphical Overview of GCN

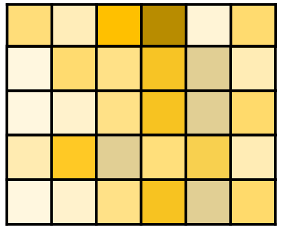


Graph Convolutional Neural Networks

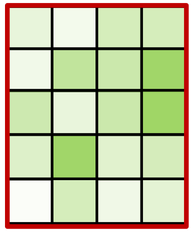
GCN ≈ CNN

GCN





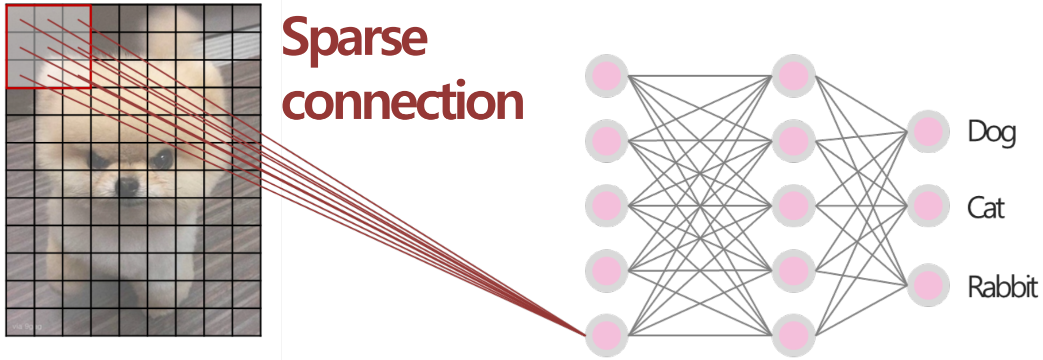
$X^{\ell} \in R^{n \times f}$



$H^{\ell+1} \in R^{n \times h}$

$h =$ number of filters

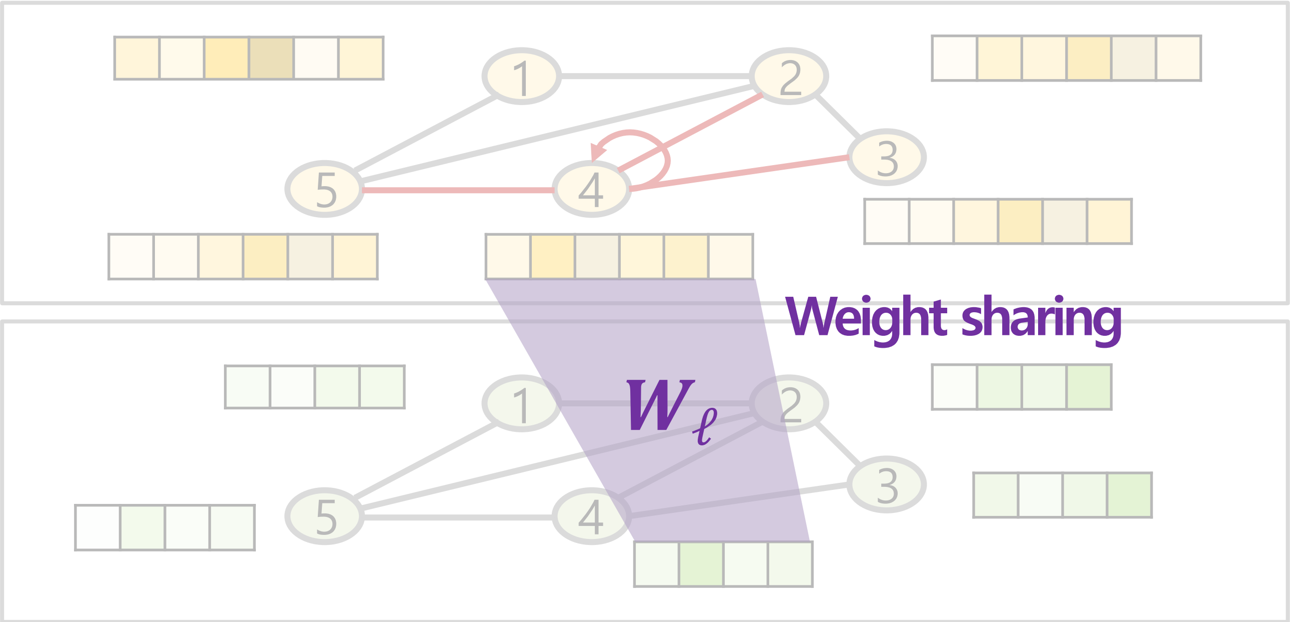
CNN

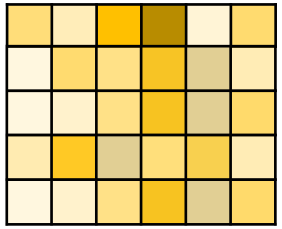


Graph Convolutional Neural Networks

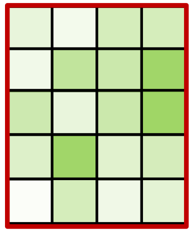
Advanced Techniques of GCN

GCN





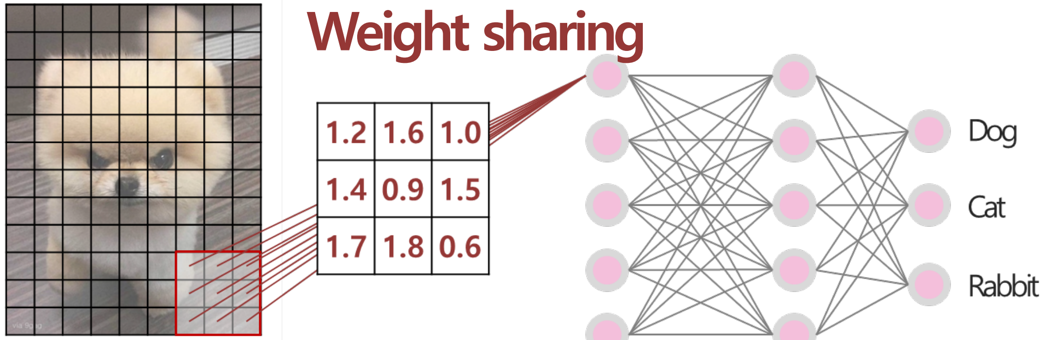
$X^\ell \in R^{n \times f}$



$H^{\ell+1} \in R^{n \times h}$

$h =$ number of filters

CNN



Conclusions

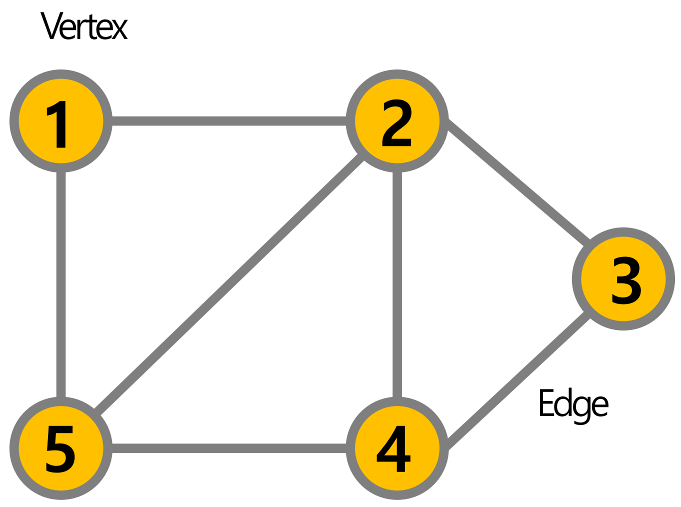
Key point

- ❖ Graph data에 대한 이해
- ❖ Semi-supervised learning에 대한 이해
- ❖ Label propagation 연구들의 수식 전개 흐름
- ❖ Graph data를 CNN에 적용하고자 개발된 GCN의 구조

Conclusions

Key point

- ❖ Graph data에 대한 이해
- ❖ Semi-supervised learning에 대한 이해
- ❖ Label propagation 연구들의 수식 전개 흐름
- ❖ Graph data를 CNN에 적용하고자 개발된 GCN의 구조



Degree Matrix

$D \in R^{n \times n},\ D_{i,i} = \sum_j A_{i,j}$

2	0	0	0	0
0	4	0	0	0
0	0	2	0	0
0	0	0	3	0
0	0	0	0	3

Adjacency Matrix

$A \in R^{n \times n}$

0	1	0	0	1
1	0	1	1	1
0	1	0	1	0
0	1	1	0	1
1	1	0	1	0

Laplacian Matrix

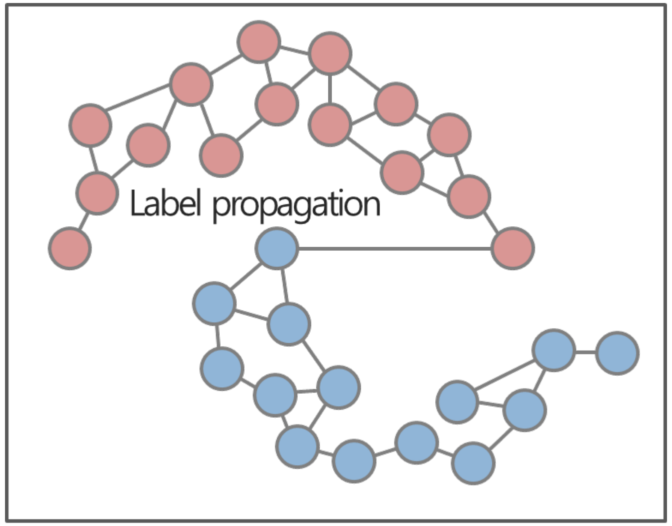
$L \in R^{n \times n},\ L = D - A$

2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3

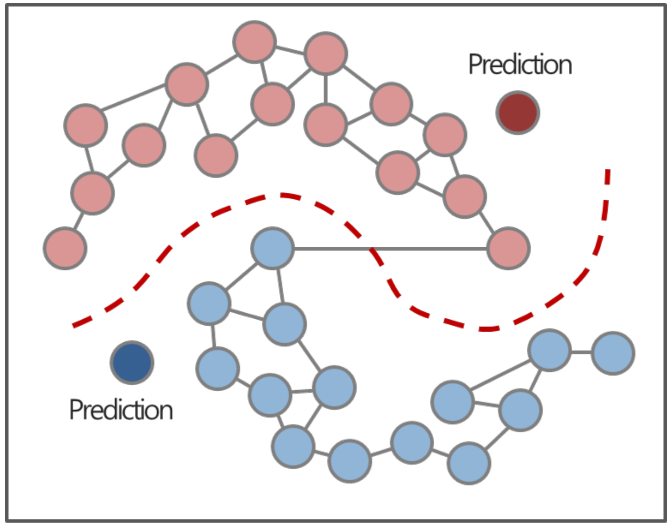
Conclusions

Key point

- ❖ Graph data에 대한 이해
- ❖ Semi-supervised learning에 대한 이해
- ❖ Label propagation 연구들의 수식 전개 흐름
- ❖ Graph data를 CNN에 적용하고자 개발된 GCN의 구조



Transductive learning



Inductive learning

Conclusions

Key point

❖ Graph data에 대한 이해

❖ Semi-supervised learning에 대한 이해

❖ Label propagation 연구들의 수식 전개 흐름

Min-cut	$\min_{y \in \{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} (y_i - y_{L_i})^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}(y_i - y_j)^2$
Harmonic Solution	$\min_{f \in \mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} (f(x_i) - y_{L_i})^2 + f^T L f$
Local and Global Consistency	$\min_{f \in \mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)(f(x_i) - y_{L_i})^2 + \mu f^T L f$
Pairwise Constraints	$\min_{f \in \mathbb{R}^{n_L+n_U}} (1-\mu)\lVert(f-y)\rVert^2 + \mu f^T L f$

❖ Graph data를 CNN에 적용하고자 개발된 GCN의 구조

Conclusions

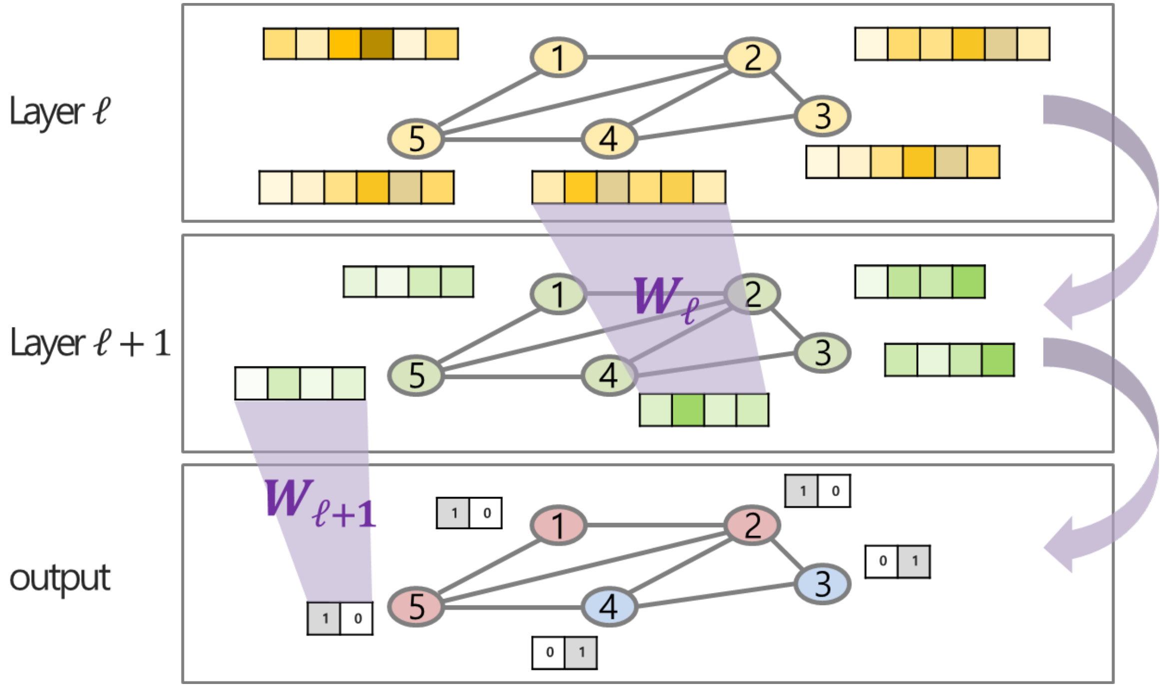
Key point

❖ Graph data에 대한 이해

❖ Semi-supervised learning에 대한 이해

❖ Label propagation 연구들의 수식 전개 흐름

❖ Graph data를 CNN에 적용하고자 개발된 GCN의 구조



Thank you

References

1. Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.
2. Bai, L., Wang, J., Liang, J., & Du, H. (2020). New label propagation algorithm with pairwise constraints. Pattern Recognition, 106, 107411.
3. Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
4. Chong, Y., Ding, Y., Yan, Q., & Pan, S. (2020). Graph-based semi-supervised learning: A review. Neurocomputing, 408, 216-230.
5. http://dmqa.korea.ac.kr/activity/seminar/267
6. https://github.com/pilsung-kang/Business-Analytics-IME654-/tree/master/05%20Semi-supervised%20Learning

Appendix

Semi-Supervised Learning

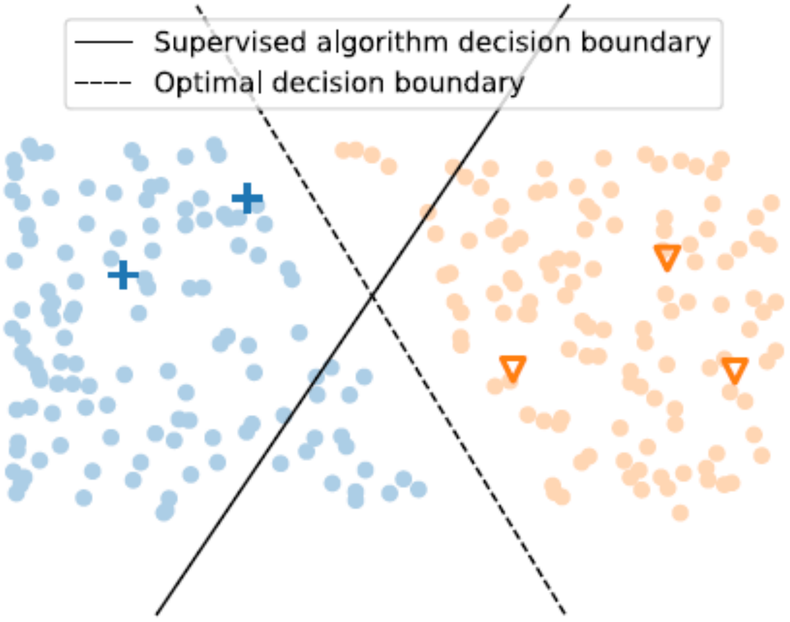
Assumptions

❖ **Smoothness assumption**

- 서로 가까운 데이터(points)는 같은 레이블일 확률이 높음

❖ **Cluster assumption**

- 같은 군집을 갖는 데이터(points)는 같은 레이블일 확률이 높음



Smoothness assumption example

❖ **Manifold assumption**

- 고차원상(input space)의 데이터(points)가 저차원상(feature space)에서 특정 구조(manifold)를 따라 놓여있음
- 저차원상에서 같은 특정 구조(manifold)를 갖는 데이터(points)는 같은 레이블일 확률이 높음

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373-440.

Semi-Supervised Learning

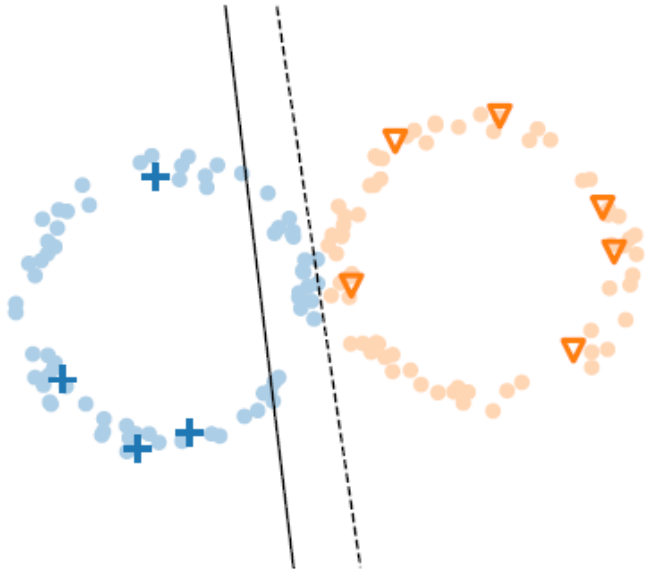
Assumptions

❖ **Smoothness assumption**

- 서로 가까운 데이터(points)는 같은 레이블일 확률이 높음

❖ **Cluster assumption**

- 같은 군집을 갖는 데이터(points)는 같은 레이블일 확률이 높음



Manifold assumption example

❖ **Manifold assumption**

- 고차원상(input space)의 데이터(points)가 저차원상(feature space)에서 특정 구조(manifold)를 따라 놓여있음
- 저차원상에서 같은 특정 구조(manifold)를 갖는 데이터(points)는 같은 레이블일 확률이 높음

Label Propagation

Harmonic Solution

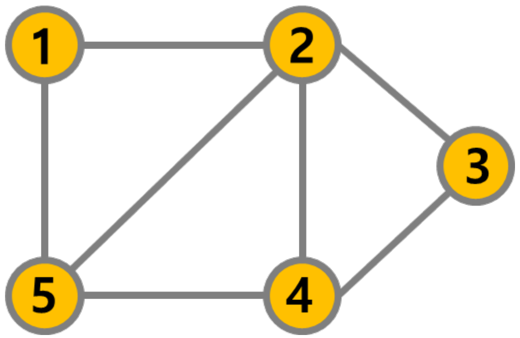
- ❖ Laplacian matrix를 활용하여 표현가능
- ❖ Laplacian matrix = Degree matrix – Adjacency matrix

$$\sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i) - f(x_j)\right)^2 = f^T L f$$

Laplacian Matrix

$L \in R^{n \times n},\ L = D - A$

2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3



$$\sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i) - f(x_j)\right)^2$$

$$= 1\times\{f(x_1) - f(x_2)\}^2 + 0\times\{f(x_1) - f(x_3)\}^2$$
$$+0\times\{f(x_1) - f(x_4)\}^2 + 1\times\{f(x_1) - f(x_5)\}^2$$
$$+1\times\{f(x_2) - f(x_3)\}^2 + 1\times\{f(x_2) - f(x_4)^2\}$$
$$+1\times\{f(x_2) - f(x_5)\}^2 + \cdots + 1\times\{f(x_4) - f(5)\}^2$$

Label Propagation

Harmonic Solution

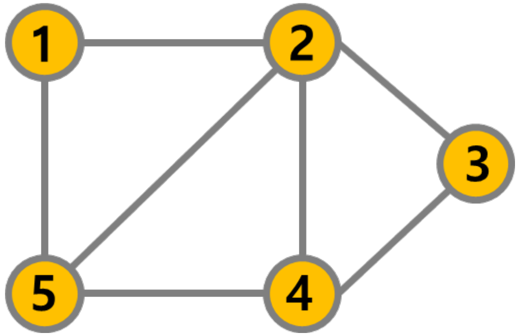
- ❖ Laplacian matrix를 활용하여 표현가능
- ❖ Laplacian matrix = Degree matrix – Adjacency matrix

$$\sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i)-f(x_j)\right)^2 = f^T L f$$

Laplacian Matrix

$L \in R^{n\times n},\ L = D - A$

2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3



$$\sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i)-f(x_j)\right)^2$$

$$= f(x_1)\{2f(x_1) - f(x_2) - f(x_5)\}$$
$$+f(x_2)\{-f(x_1) + 4f(x_2) - f(x_3) - f(x_4) - f(x_5)\}$$
$$+f(x_3)\{-f(x_2) + 2f(x_3) - f(x_4)\}$$
$$+f(x_4)\{-f(x_2) - f(x_3) + 3f(x_4) - f(x_5)\}$$
$$+f(x_5)\{-f(x_1) - f(x_2) - f(x_4) + 3f(x_5)\}$$

Label Propagation

Harmonic Solution

❖ Laplacian matrix를 활용하여 표현가능

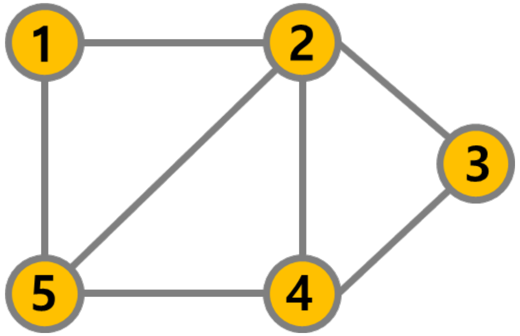
❖ Laplacian matrix = Degree matrix – Adjacency matrix

$$\sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i) - f(x_j)\right)^2 = f^T L f$$

Laplacian Matrix

$L \in R^{n \times n},\ L = D - A$

2	-1	0	0	-1
-1	4	-1	-1	-1
0	-1	2	-1	0
0	-1	-1	3	-1
-1	-1	0	-1	3



$$\sum_{i,j=1}^{n_L+n_U} w_{ij}\left(f(x_i) - f(x_j)\right)^2$$

$$= [f(x_1)\ f(x_2)\ f(x_3)\ f(x_4)\ f(x_5)] \times \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 \\ -1 & -1 & 0 & -1 & 3 \end{bmatrix} \times \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ f(x_4) \\ f(x_5) \end{bmatrix}$$

Label Propagation

Label Propagation with Pairwise Constraints

- 최근에는 negative label, multi-label 등 label의 형태에 따라 연구가 수행
- Pairwise constraints는 예측된 clustering label ($f(x_i)$)의 순서가 다른 경우를 반영

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

ELSEVIER

New label propagation algorithm with pairwise constraints

Liang Bai[a,b,*], Junbin Wang[b], Jiye Liang[a], Hangyuan Du[b]

[a] *Institute of Intelligent Information Processing, Shanxi University, Taiyuan, 030006, Shanxi, China*
[b] *School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, Shanxi, China*

ARTICLE INFO

Article history:
Received 10 May 2019
Revised 17 March 2020
Accepted 29 April 2020
Available online 11 May 2020

Keywords:
Cluster analysis
Semi-supervised clustering
Label propagation algorithm
Pairwise constraint
Spectral clustering

ABSTRACT

The label propagation algorithm is a well-known semi-supervised clustering method, which uses pre-given partial labels as constraints to predict the labels of unlabeled data. However, the algorithm has the following limitations: (1) it does not fully consider the misalignment between the pre-given labels and clustering labels, and (2) it only uses label information as clustering constraints. Real applications not only contain partial label information but pairwise constraints on a dataset. To overcome these deficiencies, a new version of the label propagation algorithm is proposed, which makes use of pairwise relations of labels as constraints to construct an optimization model for spreading labels. Experimental analysis was used to compare the proposed algorithm with 8 other semi-supervised clustering algorithms on 11 benchmark datasets. The experimental results demonstrated that the proposed algorithm is more effective than other algorithms.

Bai, L., Wang, J., Liang, J., & Du, H. (2020). New label propagation algorithm with pairwise constraints. Pattern Recognition, 106, 107411.

Label Propagation

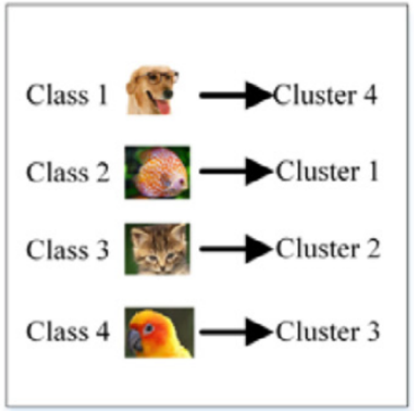
Label Propagation with Pairwise Constraints

❖ Pairwise constraints는 예측된 clustering label ($f(x_i)$)의 순서가 다른 경우를 반영

❖ 다양한 형태의 label (y)에 적용 가능하도록 변경

Objects \ Classes	Class 1	Class 2	Class 3	Class 4
x_1	1	0	0	0
x_2	0	1	0	0
x_3	0	1	0	0
x_4	0	0	1	0
x_5	0	0	0	1
x_6	0	0	1	0

Membership matrix Y of objects to classes



Relation between classes and clusters

Objects \ Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4
x_1	0	0	0	1
x_2	1	0	0	0
x_3	1	0	0	0
x_4	0	1	0	0
x_5	0	0	1	0
x_6	0	1	0	0

Membership matrix F of objects to clusters

Fig. 1. Misalignment between class labels and cluster labels.

Bai, L., Wang, J., Liang, J., & Du, H. (2020). New label propagation algorithm with pairwise constraints. Pattern Recognition, 106, 107411.

Label Propagation

Label Propagation with Pairwise Constraints

- ❖ Pairwise constraints는 예측된 clustering label ($f(x_i)$)의 순서가 다른 경우를 반영
- ❖ 다양한 형태의 label (y)에 적용 가능하도록 변경

Harmonic solution

$$\min_{f \in \mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)\left(f(x_i) - y_{L_i}\right)^2 + \mu f^T L f$$

Local consistency **Global consistency**

$$\min_{f \in \mathbb{R}^{n_L+n_U}} (1-\mu)\left(f(x_i) - y_{L_i}\right)^2 + \mu f^T L f$$

$$\min_{f \in \mathbb{R}^{n_L+n_U}} (1-\mu)\|(f - y)\|^2 + \mu f^T L f$$

Pairwise Constraints

$$\min_{f \in \mathbb{R}^{n_L+n_U}} (1-\mu)\|(ff^T - P)\|^2 + \mu f^T L f$$

$$P = \begin{cases} yy^T, & given\ positive\ labels \\ \frac{1}{k-1}(y^- y^{-T}), & given\ negative\ labels \end{cases}$$

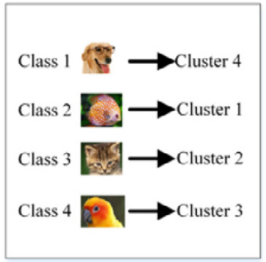
Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Label Propagation with Pairwise Constraints

Classes / Objects	Class 1	Class 2	Class 3	Class 4
x_1	1	0	0	0
x_2	0	1	0	0
x_3	0	1	0	0
x_4	0	0	1	0
x_5	0	0	0	1
x_6	0	0	1	0

Membership matrix Y of objects to classes



Relation between classes and clusters

Clusters / Objects	Cluster 1	Cluster 2	Cluster 3	Cluster 4
x_1	0	0	0	1
x_2	1	0	0	0
x_3	1	0	0	0
x_4	0	1	0	0
x_5	0	0	1	0
x_6	0	1	0	0

Membership matrix F of objects to clusters

Fig. 1. Misalignment between class labels and cluster labels.

$$\min_{f \in \mathbb{R}^{n_L + n_U}} (1-\mu)\|(ff^T - P)\|^2 + \mu f^T L f$$

$$P = \begin{cases} yy^T, & given\ positive\ labels \\ \frac{1}{k-1}(y^- y^{-T}), & given\ negative\ labels \end{cases}$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad f = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$yy^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \qquad ff^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

=

Class label 과 Clustering label 사이 매칭되지 않는 문제를 해결

Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. Advances in neural information processing systems, 16(16), 321-328.

Label Propagation

Summary

Min-cut	$$\min_{y\in\{0,1\}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(y_i - y_{L_i}\right)^2 + \sum_{i,j=1}^{n_L+n_U} w_{ij}\left(y_i - y_j\right)^2$$	
Harmonic Solution	$$\min_{f\in\mathbb{R}^{n_L+n_U}} \infty \sum_{i=1}^{n_L} \left(f(x_i) - y_{L_i}\right)^2 + f^T L f$$	
Local and Global Consistency	$$\min_{f\in\mathbb{R}^{n_L+n_U}} \sum_{i=1}^{n_L} (1-\mu)\left(f(x_i) - y_{L_i}\right)^2 + \mu f^T L f$$	
Pairwise Constraints	$$\min_{f\in\mathbb{R}^{n_L+n_U}} (1-\mu)\|(f-y)\|^2 + \mu f^T L f$$	

Labeled : Supervised loss

Unlabeled : Semi-Supervised loss