

딥러닝에서의 준 지도 학습법에 관한 연구

서울대학교
통계학과
김용대, 김동하

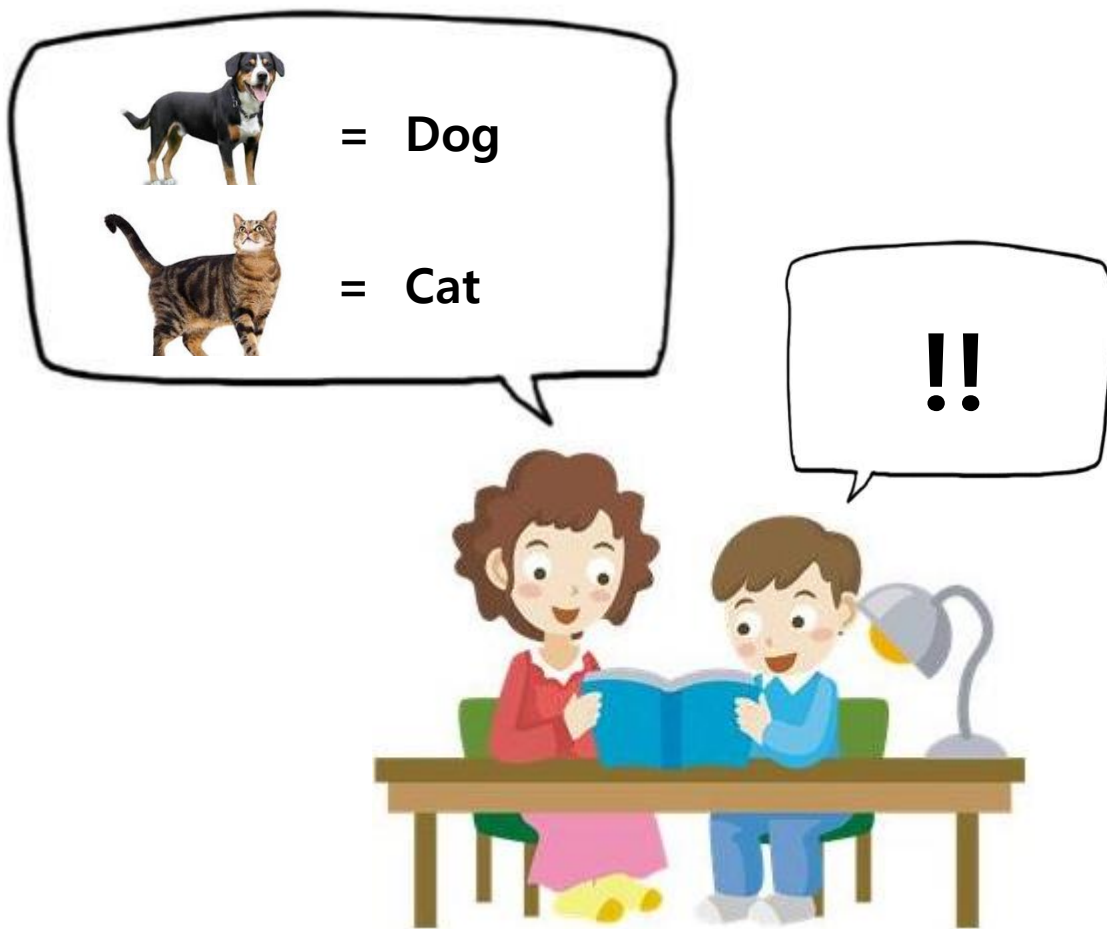
Contents

1. 개요
2. Deep SSL 설명
3. Deep SSL을 이용한 실험

1. 개요-4가지 학습 방법론

1.1 지도 학습(Supervised learning)

- 목표값이 있는 훈련 데이터들을 이용하여 임의의 데이터로부터 예측하고자 하는 값을 올바르게 추측해내는 함수를 학습하는 방법.



1. 개요-4가지 학습 방법론

1.2 비지도 학습(Unsupervised learning)

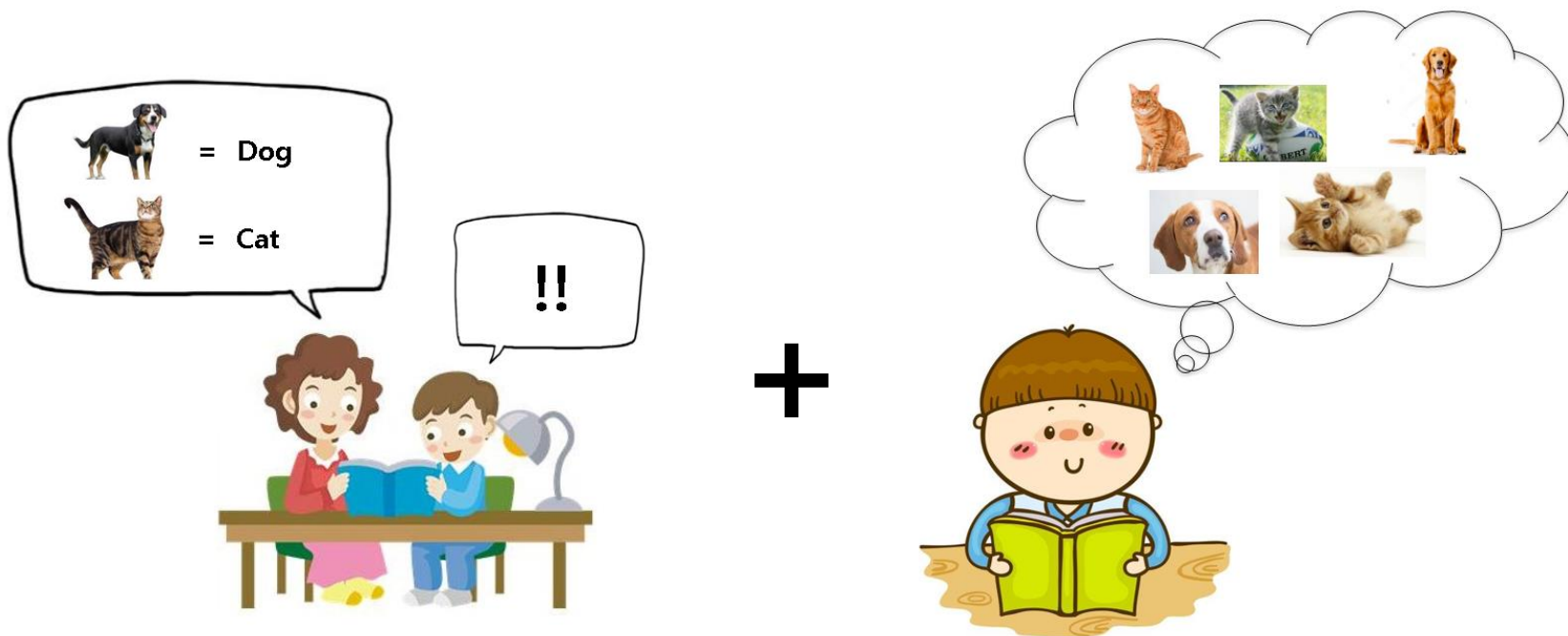
- 목표값이 없는 데이터들을 이용하여 데이터들의 분포 또는 군집에 대한 정보를 학습하는 방법.



1. 개요-4가지 학습 방법론

1.3 준 지도 학습(Semi-supervised learning)

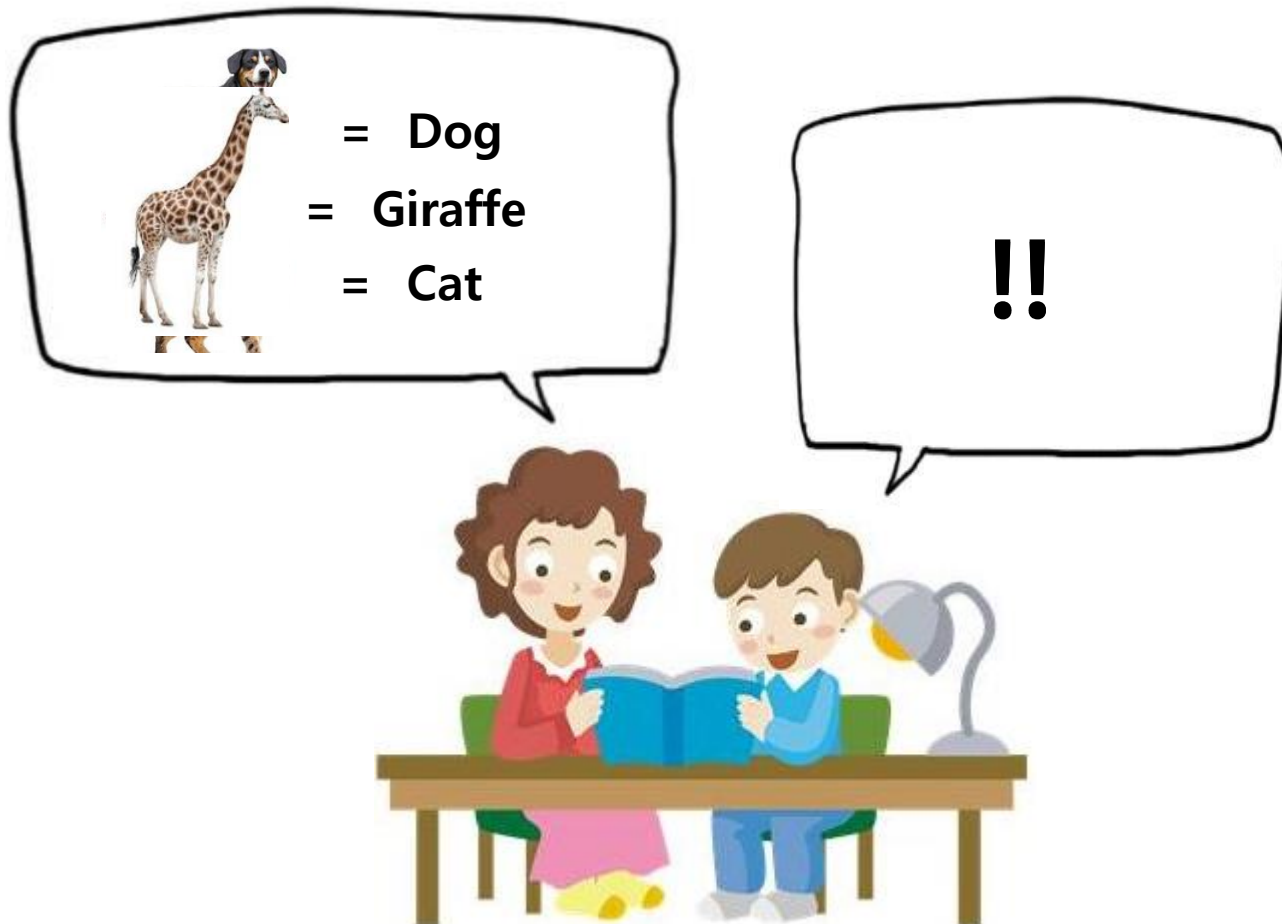
- 목표값이 있는 데이터들과 목표값이 없는 데이터들을 모두 이용하여 임의의 데이터에 대한 목표값을 예측하는 함수를 학습하는 방법.



1. 개요-4가지 학습 방법론

1.4 점진적 학습(Incremental learning)

- 추가되는 데이터로부터 새로운 정보를 학습하여 이미 습득된 기존의 지식을 갱신하는 학습법

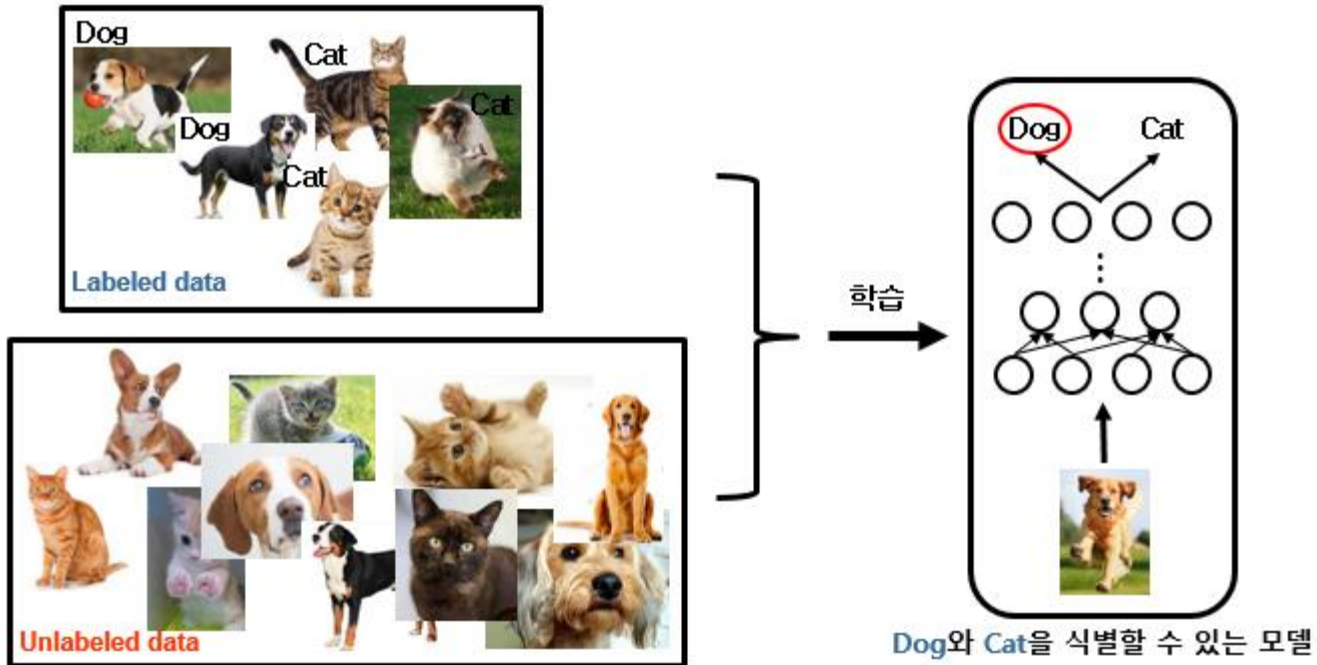


1. 개요-연구 목표

연구 목표 1.

- 수많은 자료들 중에는 목표값이 있는 자료보다 **목표값이 없는 자료가 훨씬 많음.**
(목표값은 사람이 수동으로 입력해줘야 하므로)

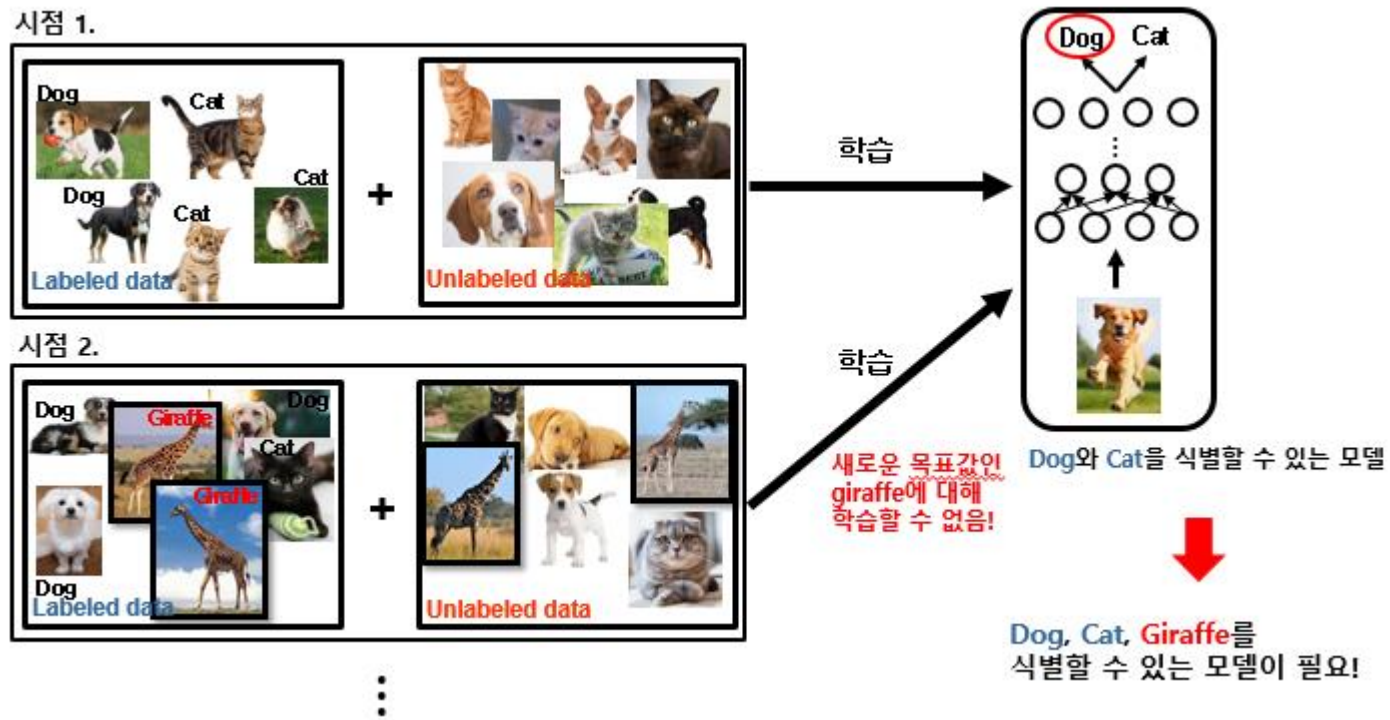
- ✓ 목표값을 잘 예측하는 **딥러닝 모델을 학습**할 때에, 목표값이 있는 데이터 뿐만 아니라 **목표값이 없는 데이터들도 함께 사용**하여 좋은 예측 모델을 학습하는 방법을 개발.



1. 개요-연구 목표

연구 목표 2.

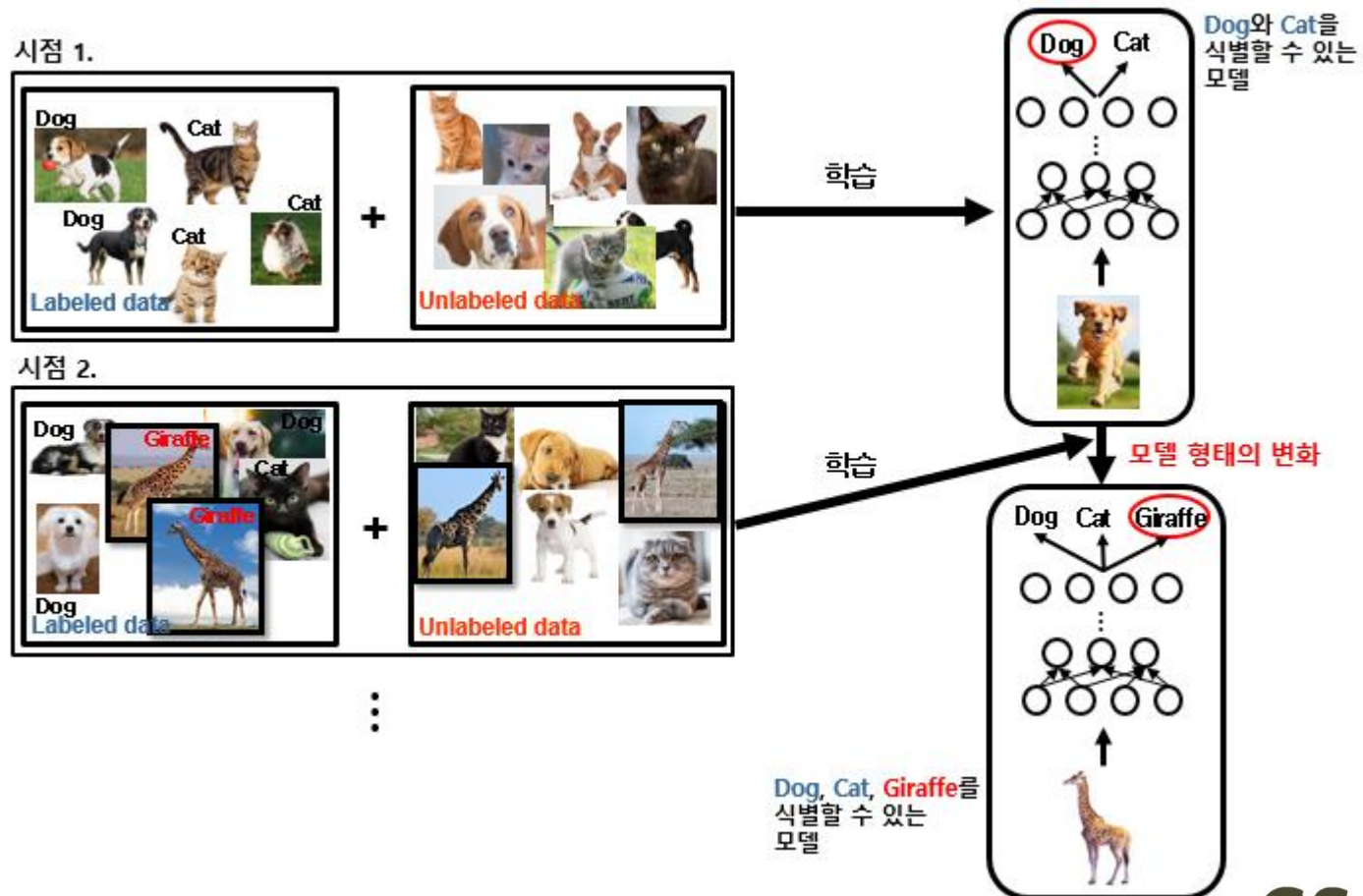
- 여러 시간에 걸쳐서 데이터가 생성되는 경우에는 한 번에 모든 데이터를 사용하여 모델을 학습할 수 없고, 업데이트되는 적은 양의 데이터를 이용하여 **모델을 매번 점진적으로 학습**해야 함.
- 또한, 새로 업데이트되는 자료에는 **기존에 없던 목표값**을 가질 수가 있음.



1. 개요-연구 목표

연구 목표 2. (Cont.)

- ✓ 점진적 학습을 하면서 새롭게 업데이트되는 목표값들을 이용하여 **모델 자체의 형태가 변화**하도록 하는 학습법을 개발.



1. 개요-예제

목적 : 새로운 종류의 이미지 자동 탐지 및 tagging

예 : 오픈마켓의 신상품 분류



갤럭시 기어 S2

휴대폰 ▾ > 웨어러블 디바이스 ▾ > **스마트 워치** ▾



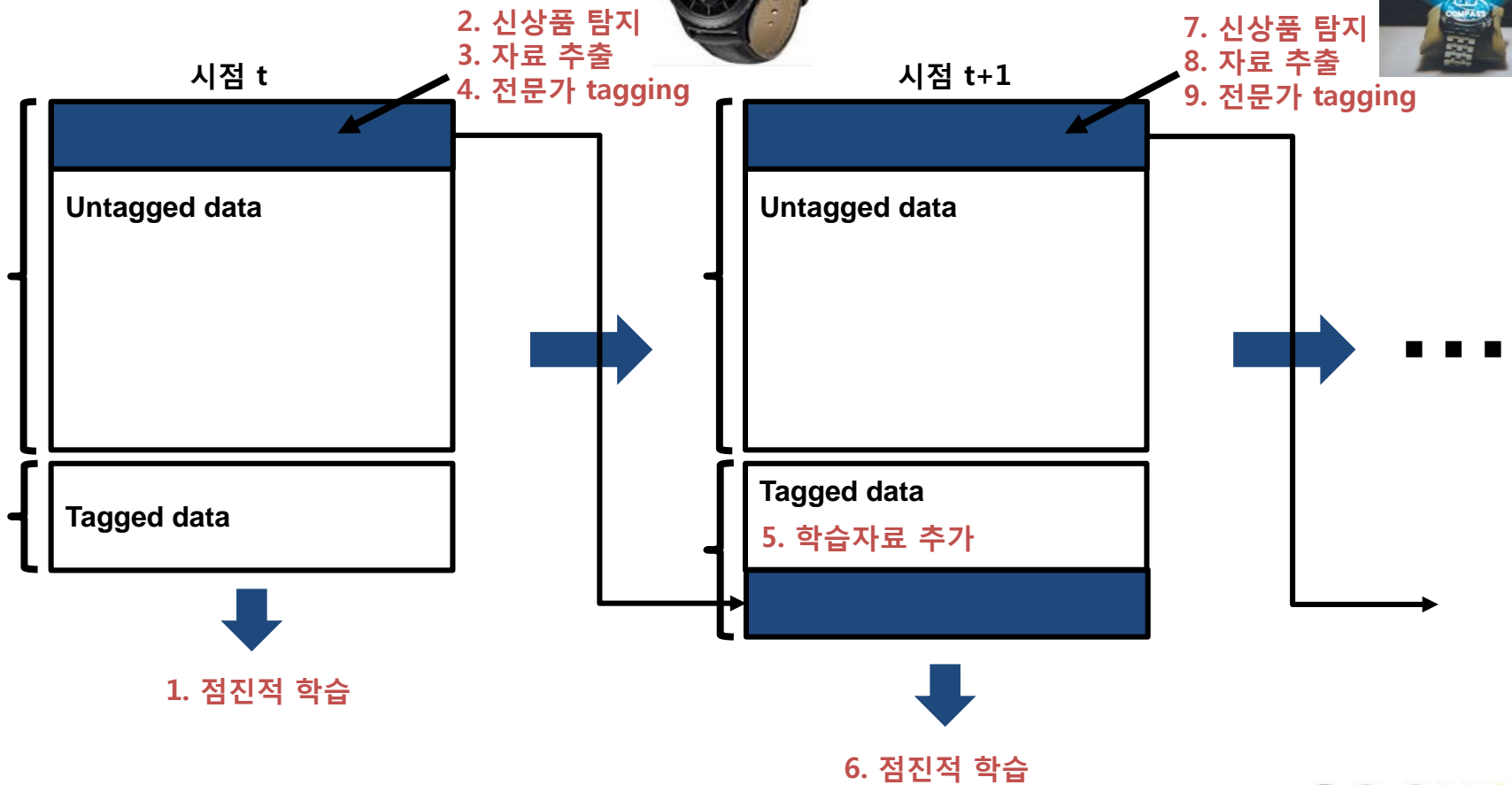
갤럭시 S7 엣지

휴대폰 ▾ > 공기계 ▾ > **삼성** ▾

1. 개요-예제

목적 : 새로운 종류의 이미지 자동 탐지 및 tagging

방법론



1. 딥러닝에서의 준지도 학습법

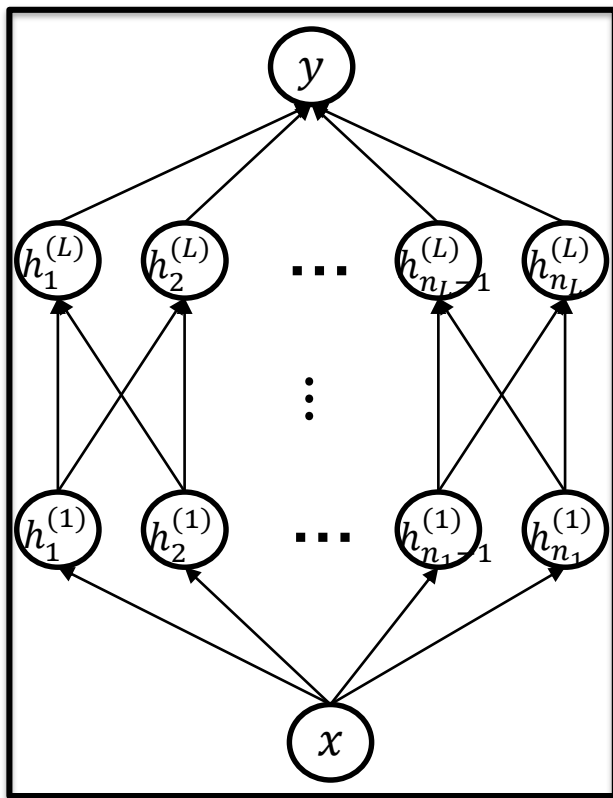
- Deep learning에 사용하는 자료들은 목표값이 없는 자료들의 비율이 높음.
 - ✓ 준 지도 학습의 필요성
- 손실 함수는 주로 지도 학습과 비지도 학습의 손실 함수의 합으로 나타내어짐.

$$l^{semi}(\theta; D^l, D^u) = l^l(\theta; D^l) + l^u(\theta; D^u)$$

- θ : 추정하고자 하는 모수의 집합.
- D^l : 목표값이 있는 데이터.
- D^u : 목표값이 없는 데이터.

2. Deep SSL 설명

Supervised learning part : Deep Neural Networks



$$z_i^{(1)} = b_i^{(1)} + \mathbf{w}_i^{(1)T} \mathbf{x}, \quad i = 1, \dots, n_1$$

$$h_i^{(1)} = \sigma \left(z_i^{(1)} \right), \quad i = 1, \dots, n_1$$

$$z_i^{(2)} = b_i^{(2)} + \mathbf{w}_i^{(2)T} \mathbf{z}^{(1)}, \quad i = 1, \dots, n_2$$

$$h_i^{(2)} = \sigma \left(z_i^{(2)} \right), \quad i = 1, \dots, n_2$$

...

$$z_i^{(L)} = b_i^{(L-1)} + \mathbf{w}_i^{(L-1)T} \mathbf{z}^{(L-1)}, \quad i = 1, \dots, n_L$$

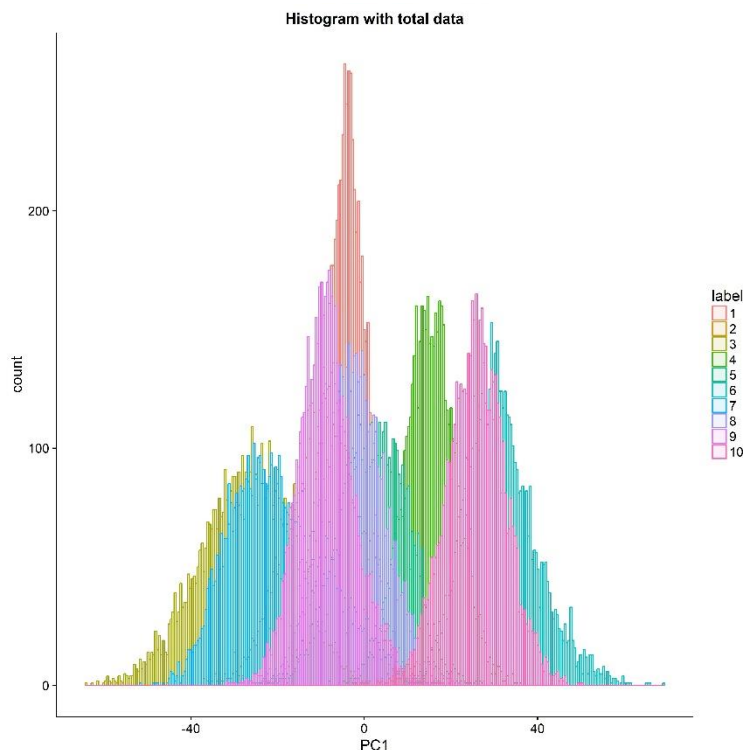
$$h_i^{(L)} = \sigma \left(z_i^{(L-1)} \right), \quad i = 1, \dots, n_L$$

$$P(y = k | \mathbf{x}) = \text{softmax} \left(\mathbf{h}^{(L)'} \mathbf{w}_k^{(L+1)} + b_k^{(L+1)} \right), \quad k = 1, \dots, K$$

2. Deep SSL 설명

Unsupervised learning part :

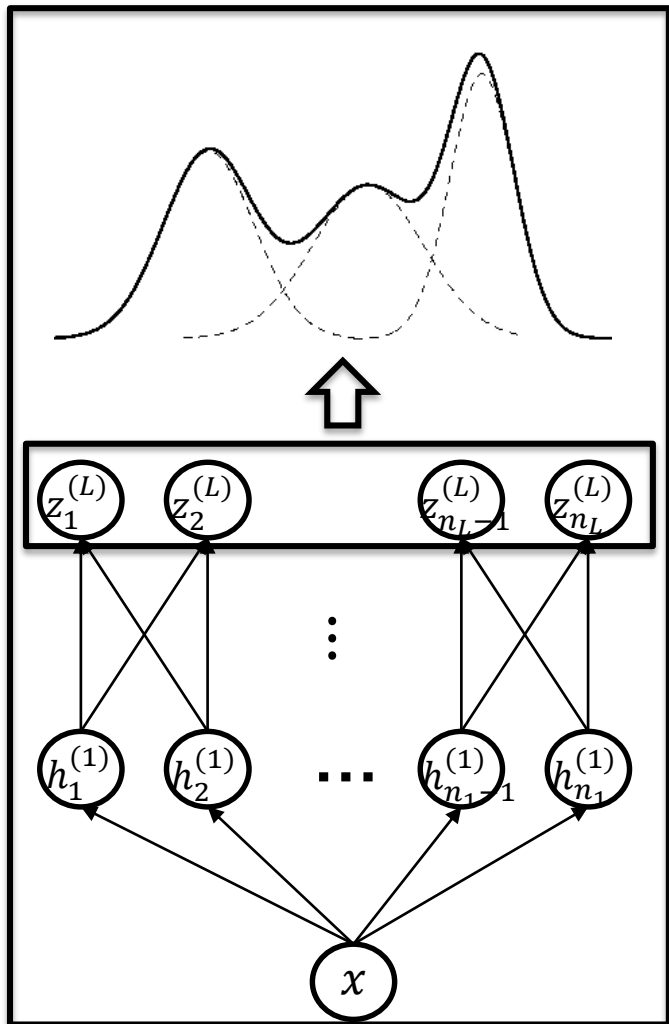
- Input data에 어떤 분포를 가정?
- DNN의 결과로부터 얻은 최상층의 hidden node들의 분포를 살펴보자.



혼합 정규분포 모형을
가정하자!

2. Deep SSL 설명

Unsupervised learning part : DNN + GMM



$$z_i^{(1)} = b_i^{(1)} + \mathbf{w}_i^{(1)T} \mathbf{x}, \quad i = 1, \dots, n_1$$

$$h_i^{(1)} = \sigma \left(\mathbf{z}_i^{(1)} \right), \quad i = 1, \dots, n_1$$

$$z_i^{(2)} = b_i^{(2)} + \mathbf{w}_i^{(2)T} \mathbf{z}^{(1)}, \quad i = 1, \dots, n_2$$

$$h_i^{(2)} = \sigma \left(\mathbf{z}_i^{(2)} \right), \quad i = 1, \dots, n_2$$

...

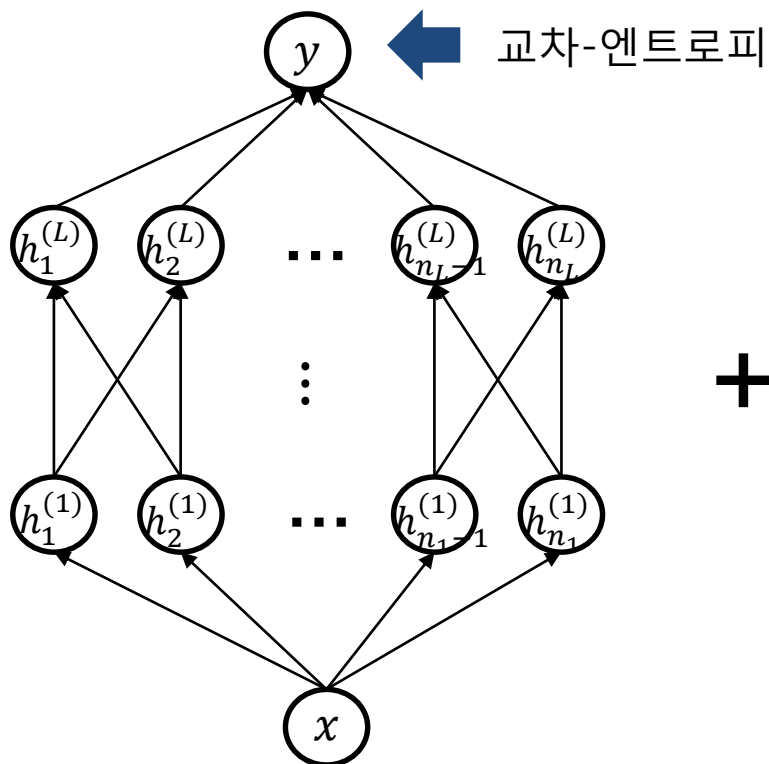
$$z_i^{(L)} = b_i^{(L-1)} + \mathbf{w}_i^{(L-1)T} \mathbf{z}^{(L-1)}, \quad i = 1, \dots, n_L$$

$$\nu \sim \text{Multi}(\pi_1, \dots, \pi_K)$$

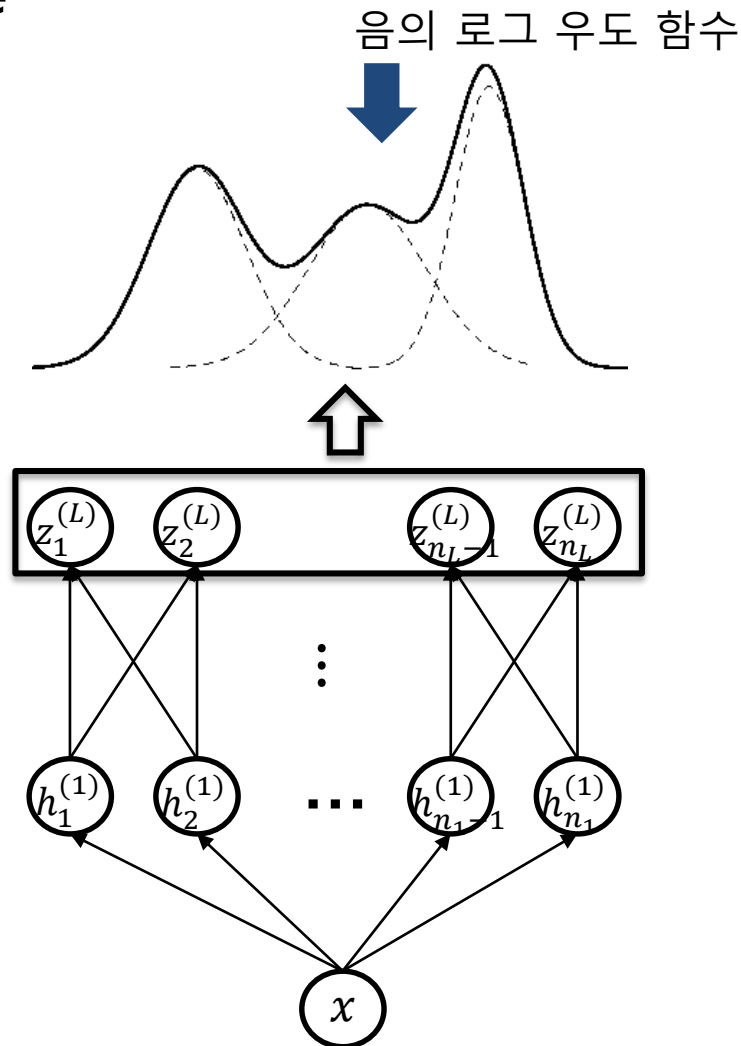
$$\mathbf{z}^{(L)} | (\nu = k) \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$$

2. Deep SSL 설명

Semi-supervised learning procedure



+



2. Deep SSL 설명

- 계산 알고리즘
 - 다음의 목적 함수를 최소화

$$l(\mathbf{W}, \mathbf{b}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}; \mathbf{D}^l, \mathbf{D}^u) =$$

지도 학습 손실 함수
(=교차 엔트로피)

$$-\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K I(y_i^l = k) \log \left(\text{softmax} \left((\mathbf{h}^{(L)}(\mathbf{x}_i^l; \mathbf{W}, \mathbf{b}))' \mathbf{w}_k^{(L+1)} + b_k^{(L+1)} \right) \right)$$

$$-\frac{\lambda_1}{n} \sum_{i=1}^n \log \left(\sum_{k=1}^K \pi_k \phi \left(\mathbf{z}^{(L)}(\mathbf{x}_i^l; \mathbf{W}, \mathbf{b}); \boldsymbol{\mu}_k, \boldsymbol{\Sigma} \right) \right)$$

$$-\frac{\lambda_2}{m} \sum_{j=1}^m \log \left(\sum_{k=1}^K \pi_k \phi \left(\mathbf{z}^{(L)}(\mathbf{x}_j^u; \mathbf{W}, \mathbf{b}); \boldsymbol{\mu}_k, \boldsymbol{\Sigma} \right) \right)$$

비지도 학습 손실 함수
(=음의 로그 우도 함수)

$$\text{subject to } \sum_{k=1}^K \pi_k = 1 \text{ and } \lambda_1, \lambda_2 > 0.$$

2. Deep SSL 설명

- 계산 알고리즘
 - DNN에 관련된 모수 추정
 - SGD(Stochastic Gradient Descent) 기반의 알고리즘
 - Ex) RMSprop, ADAM, ...
 - GMM에 관련된 모수 추정
 - EM 알고리즘

3. Deep GMM을 이용한 실험

Semi-supervised learning 성능 확인

- 제안한 방법을 이용하여 semi-supervised learning을 할 때 기존의 supervised learning과 비교하여 성능이 얼마나 향상되는지 확인.
- ✓ MNIST 데이터 사용
 - Train set : 30000 samples with labeled data and 20000 samples without labeled data
 - Test set : 20000 samples with labeled data
- ✓ Activation function : ReLU
- ✓ Drop-out 사용 X
- ✓ Optimizer : ADAM algorithm

3. Deep GMM을 이용한 실험

실험 결과

Model Structure	DNN		DNN_SSL	
	Train	Test	Train	Test
100-100-100	0.9947	0.9671	0.99973	0.9714
100-100-50	0.992333	0.9652	0.9991	0.9675
100-100	0.9928	0.9683	0.9994	0.9694
100-50	0.986433	0.9653	0.995933	0.9695
100	0.9816	0.9639	0.991533	0.9675
50	0.9675	0.9547	0.9771	0.9579

3. Deep GMM을 이용한 실험

실험 결과 (추가..)

- Convolutional neural networks를 사용.

Model	Conv_DNN		Conv_DNN_SSL	
	Train	Test	Train	Test
[5,5,1,32] [5,5,32,64] 100-100	1.	0.9908	0.996167	0.9914