Data Mining
Quality Analytics

# Introduction to deep metric learning

2021. 11. 12.

Korea Univ, Data Mining & Quality Analytics Lab.

**이민정**

# 발표자 소개

- 이민정 (Minjung Lee)
  - 고려대학교 산업경영공학과
  - 석·박사 통합과정 (2017.03~) Ph.D. Candidate
  - Data Mining & Quality Analytics Lab (김성범 교수님)
- 관심분야
  - Deep learning for multivariate sensor data
  - Incomplete multivariate sensor data
- E-mail
  - leemj2520@korea.ac.kr

# Contents

# Metric 이란?

Deep metric learning ?

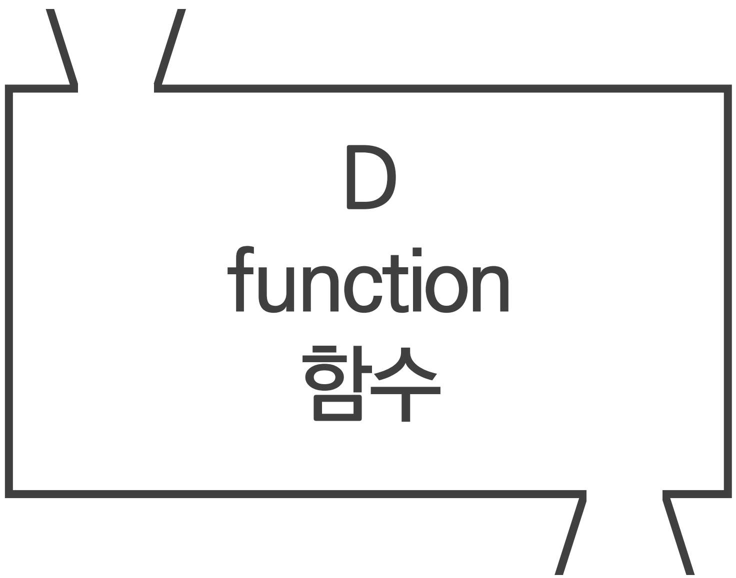
= Distance
거리

# Metric 이란?

## 거리(distance=metric)는 함수

입력 Input

두 객체
$x, y$




음이 아닌 실수값
$[0, \infty)$

출력 Output

# Metric 이란?

## 거리(distance=metric)는 다음 조건을 만족하는 함수를 의미함

$$D(x, y) = 0 \leftrightarrow x = y$$

자기 자신과의 거리는 0이다.

$$D(x, y) = D(y, x)$$

거리는 상호 대칭성을 가진다.

$$D(x, y) \leq D(x, z) + D(z, y)$$

거리는 삼각부등식을 만족한다.

# Metric 이란?

## 자주 쓰이는 거리(distance=metric) 함수의 종류

$x, y$ 가 크기가 $n$인 벡터라면,

유클리디안 거리 = 2차 노름 거리 = $\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$

P차 노름 거리 = $\sqrt[p]{\sum_{i=1}^{n}(x_i - y_i)^p}$

내적 거리 = $x \cdot y = \sum_{i=1}^{n} x_i y_i$

코사인 거리 = $1 - \frac{x \cdot y}{\|x\|\|y\|}$

마할라노비스 거리 = $\sqrt{(x-y)^T \Sigma^{-1} (x-y)}$, $\Sigma$는 데이터의 공분산 행렬

# Deep metric learning이란?

거리 공간(metric space)은
두 개체 사이의 거리가 정의된 공간
= 유사한 개체는 가까이, 유사하지 않은 개체는 멀리 위치한 공간

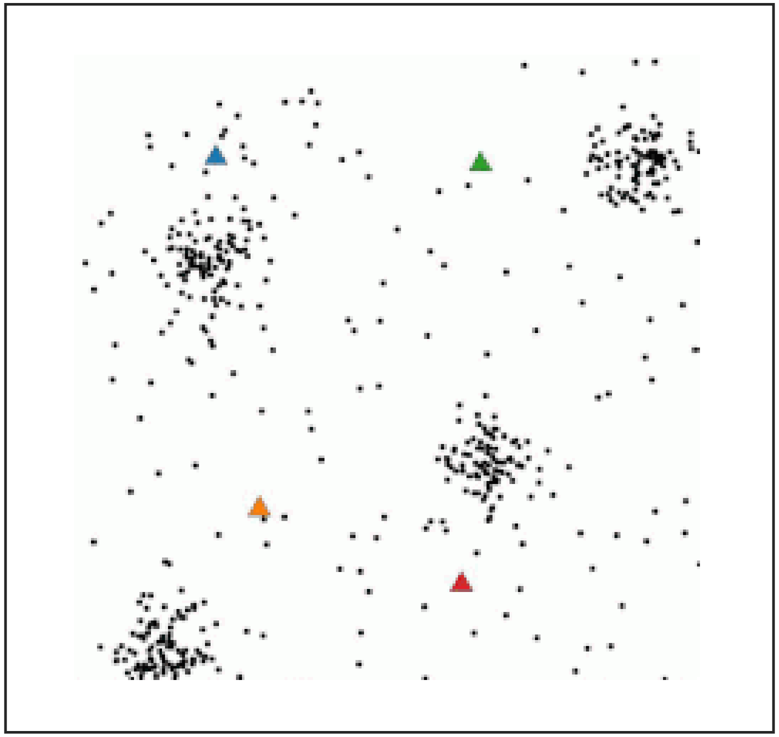
| Deep | metric | learning ? |
|---|---|---|
| = Deep neural network | = Distance | |

딥러닝 모델로 거리 공간을 학습한다

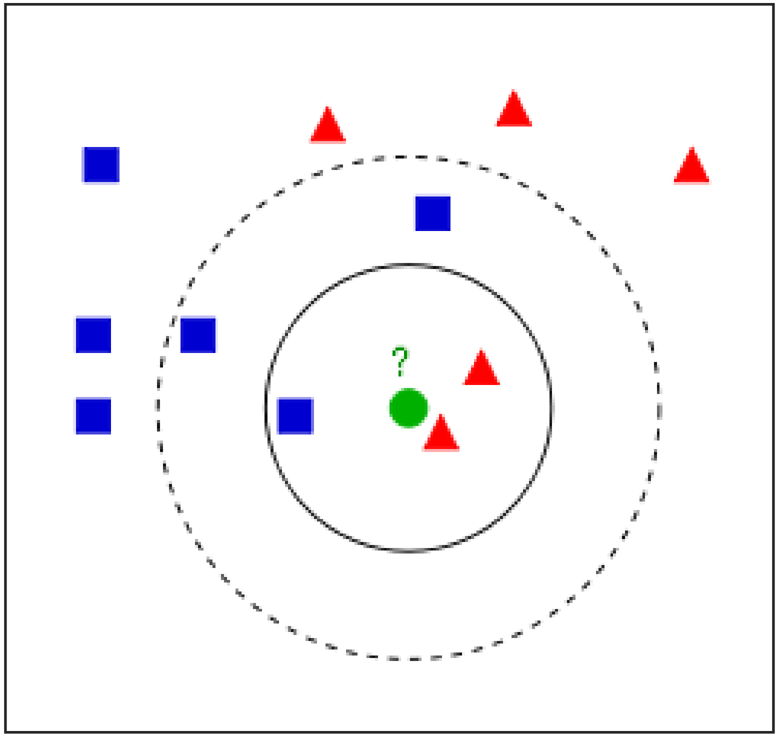
# Deep metric learning이란?

## 거리와 기계학습 알고리즘의 밀접한 관계
## 유사도 측정 → 의사결정

**비지도 학습**
군집 대표 알고리즘
K-means



**지도 학습**
수치/범주 예측 대표 알고리즘
K-nearest neighbor algorithm

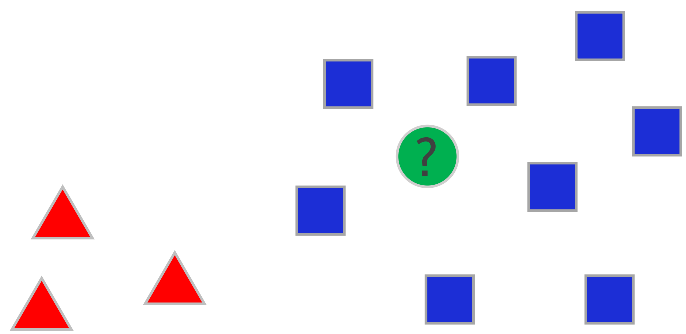


# Deep metric learning이란?

폐암 O

폐암 X

## $X$가 유사하면 $y$도 유사하다



# Deep metric learning이란?

## $X$가 유사하면 $y$도 유사하다 ?

| | 입력변수 | | | | 출력변수 |
|---|---|---|---|---|---|
| | 흡연 유무 | 폐암 가족력 | 결혼 유무 | IQ | 폐암 유무 |
| 관측치 1 | 1 | 1 | X | 100 | 폐암 O |
| 관측치 2 | 1 | 1 | O | 110 | 폐암 X |
| 관측치 3 | 0 | 0 | O | 109 | 폐암 X |
| … | … | … | … | … | … |

# Deep metric learning이란?

## $X$가 유사하면 $y$도 유사하다 ? NO!

| | 흡연 유무 | 폐암 가족력 | 결혼 유무 | IQ | 폐암 유무 |
|---|---|---|---|---|---|
| 관측치 1 | 1 | 1 | X | 100 | 폐암 O |
| 관측치 2 | 1 | 1 | O | 110 | 폐암 X |
| 관측치 3 | 0 | 0 | O | 109 | 폐암 X |
| … | … | … | … | … | … |

폐암과 관련없는 변수가 포함되어 거리가 계산!

# Deep metric learning이란?

## $X$가 유사하면 $y$도 유사하다 ? NO!

| | 흡연 유무 | 폐암 가족력 | 결혼 유무 | IQ | 폐암 유무 |
|---|---|---|---|---|---|
| 관측치 1 | 1 | 1 | X | 100 | 폐암 O |
| 관측치 2 | 1 | 1 | O | 110 | 폐암 X |
| 관측치 3 | 0 | 0 | O | 109 | 폐암 X |
| … | … | … | … | … | … |

동일한 중요도로 거리가 계산

# Deep metric learning이란?

데이터 입력 공간 input space

$X$

| | 흡연 유무 | 폐암 가족력 | 결혼 유무 | IQ | 폐암 유무 |
|---|---|---|---|---|---|
| 관측치1 | 1 | 1 | X | 100 | 폐암O |
| 관측치2 | 1 | 1 | O | 110 | 폐암X |
| 관측치3 | 0 | 0 | O | 109 | 폐암X |
| … | … | … | … | … | … |

$y$

Feature engineering

(selection, extraction…)

$\tilde{X}$

| | 폐암 관련 특징1 | 폐암 관련 특징2 | 폐암 유무 |
|---|---|---|---|
| 관측치1 | 2 | 1 | 폐암O |
| 관측치2 | 1 | 0.5 | 폐암X |
| 관측치3 | 5 | 0 | 폐암X |
| … | … | … | … |

$y$

잠재/임베딩 공간 latent/embedding space

# Deep metric learning이란?

데이터 입력 공간 input space

## 최근 데이터 특징?

Feature engineering

(selection, extraction…)

$\tilde{X}$

| | 폐암 관련 특징1 | 폐암 관련 특징2 | 폐암 유무 |
|---|---|---|---|
| 관측치1 | 2 | 1 | 폐암O |
| 관측치2 | 1 | 0.5 | 폐암X |
| 관측치3 | 5 | 0 | 폐암X |
| … | … | … | … |

$y$

잠재/임베딩 공간 latent/embedding space

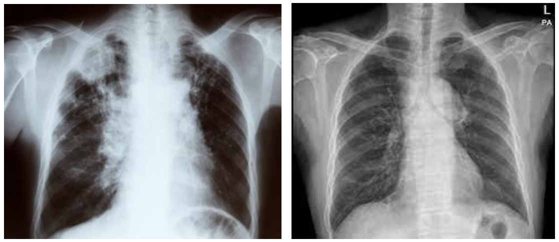
# Deep metric learning이란?

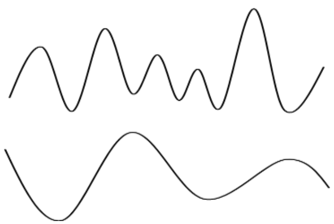
## 데이터 입력 공간 input space

변수의 수 多 정형 데이터

| | 흡연 유무 | 폐암 가족력 | … | IQ | 폐암 유무 |
|---|---|---|---|---|---|
| 관측치1 | 1 | 1 | … | 100 | 폐암O |
| 관측치2 | 1 | 1 | … | 110 | 폐암X |
| 관측치3 | 0 | 0 | … | 109 | 폐암X |
| … | … | … | … | … | … |

비정형 데이터





Feature engineering
(selection, extraction⋯)

$\widetilde{X}$

| | 폐암 관련 특징1 | 폐암 관련 특징2 | 폐암 유무 |
|---|---|---|---|
| 관측치1 | 2 | 1 | 폐암O |
| 관측치2 | 1 | 0.5 | 폐암X |
| 관측치3 | 5 | 0 | 폐암X |
| … | … | … | … |

$y$

## 잠재/임베딩 공간 latent/embedding space

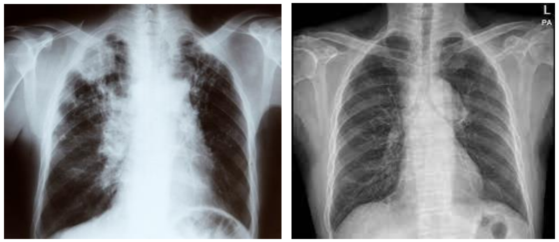
# Deep metric learning이란?

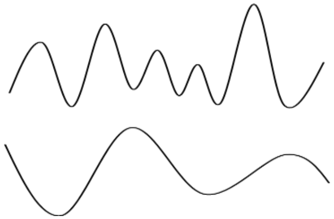
## 데이터 입력 공간 input space

### 변수의 수 多 정형 데이터

| | 흡연 유무 | 폐암 가족력 | … | IQ | 폐암 유무 |
|---|---|---|---|---|---|
| 관측치1 | 1 | 1 | … | 100 | 폐암O |
| 관측치2 | 1 | 1 | … | 110 | 폐암X |
| 관측치3 | 0 | 0 | … | 109 | 폐암X |
| … | … | … | … | … | … |

### 비정형 데이터





Feature engineering

(selection, extraction…)

Deep neural network

$\tilde{X}$

| | 폐암 관련 특징1 | 폐암 관련 특징2 | 폐암 유무 |
|---|---|---|---|
| 관측치1 | 2 | 1 | 폐암O |
| 관측치2 | 1 | 0.5 | 폐암X |
| 관측치3 | 5 | 0 | 폐암X |
| … | … | … | … |

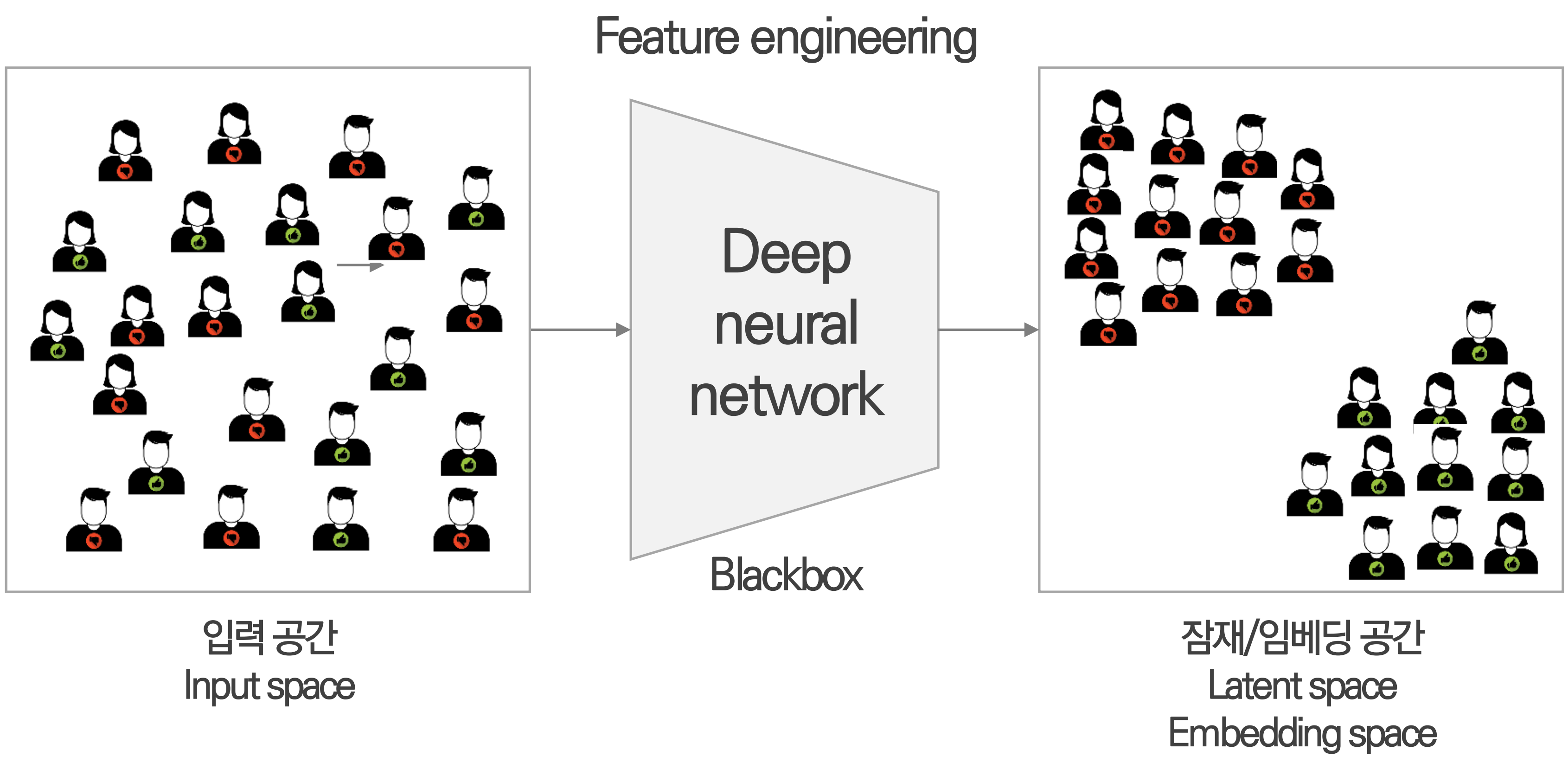
$y$

## 잠재/임베딩 공간 latent/embedding space

# Deep metric learning이란?

## 딥러닝 모델로 거리 공간을 학습한다

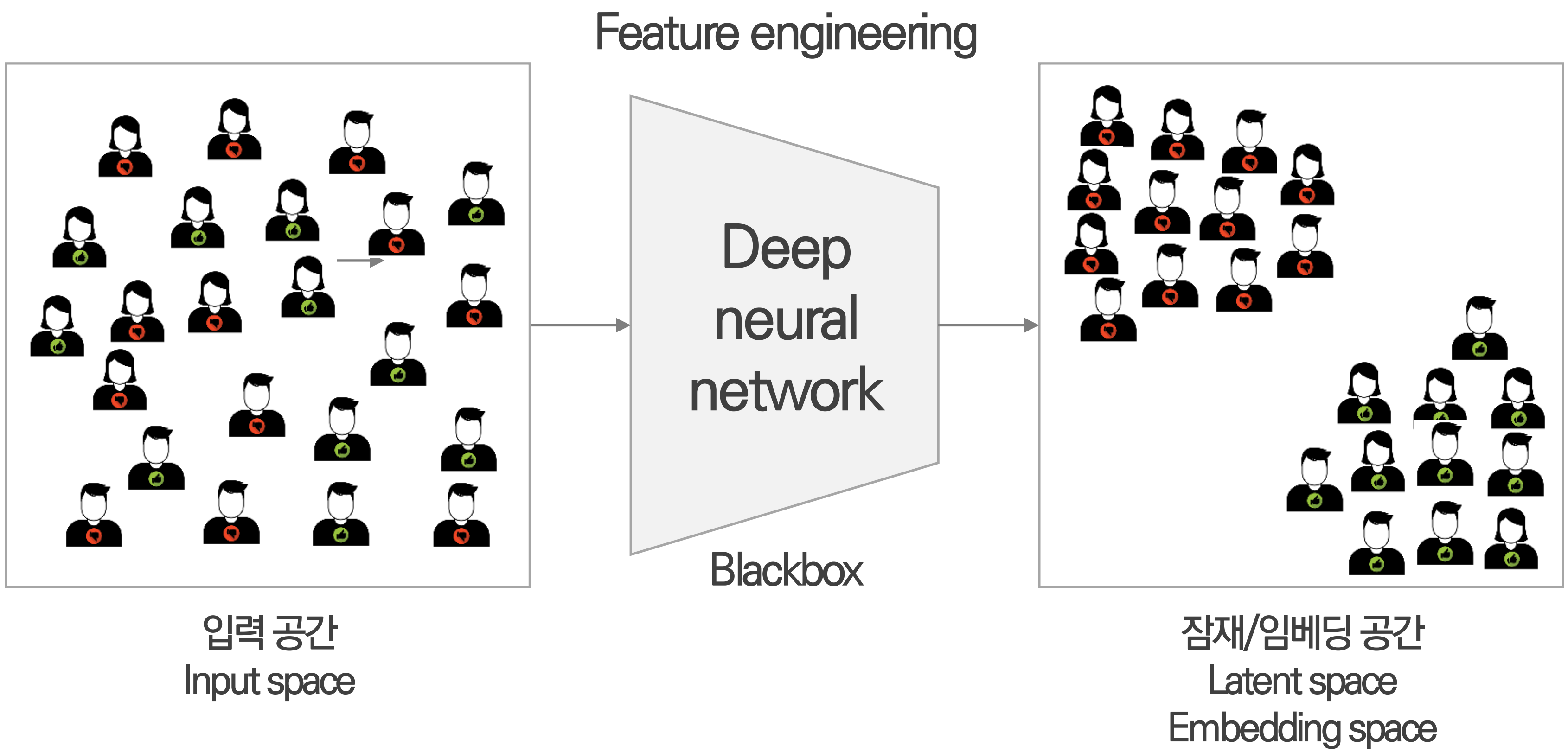
유사한 것(같은 클래스)끼리 가깝게
유사하지 않은 것 (다른 클래스)끼리는 멀게하는 특징을 추출



# Softmax classifier로 학습된 feature의 한계

## 어떤 손실함수를 써야할까?

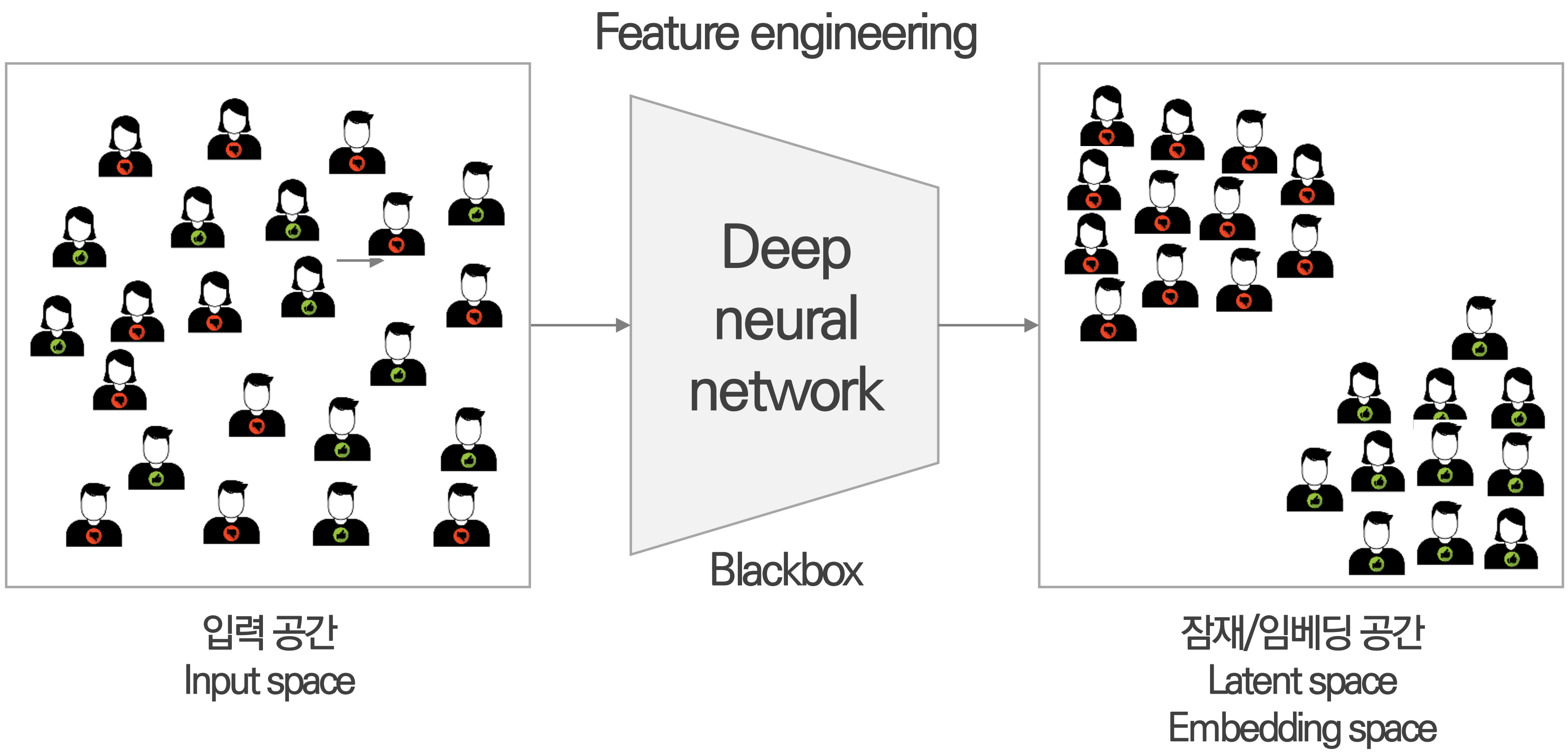
유사한 것(같은 클래스)끼리 가깝게
유사하지 않은 것 (다른 클래스)끼리는 멀게하는 특징을 추출



# Softmax classifier로 학습된 feature의 한계

## Softmax 사용해서 학습한 딥러닝 모델 사용?

유사한 것(같은 클래스)끼리 가깝게
유사하지 않은 것 (다른 클래스)끼리는 멀게하는 특징을 추출



# Softmax classifier로 학습된 feature의 한계

## Deep metric learning만을 위한 손실함수의 필요성

Face recognition 얼굴인식 분야

# Softmax classifier로 학습된 feature의 한계

- 얼굴 인식 기술(face recognition)은 사진이나 동영상의 얼굴에서 정보를 파악하는 기술
- Deep metric learning이 우수한 성능을 보이는 computer vision 분야
- 얼굴 인식은 크게 face identification(식별), verification(검증)으로 나뉨

# Softmax classifier로 학습된 feature의 한계

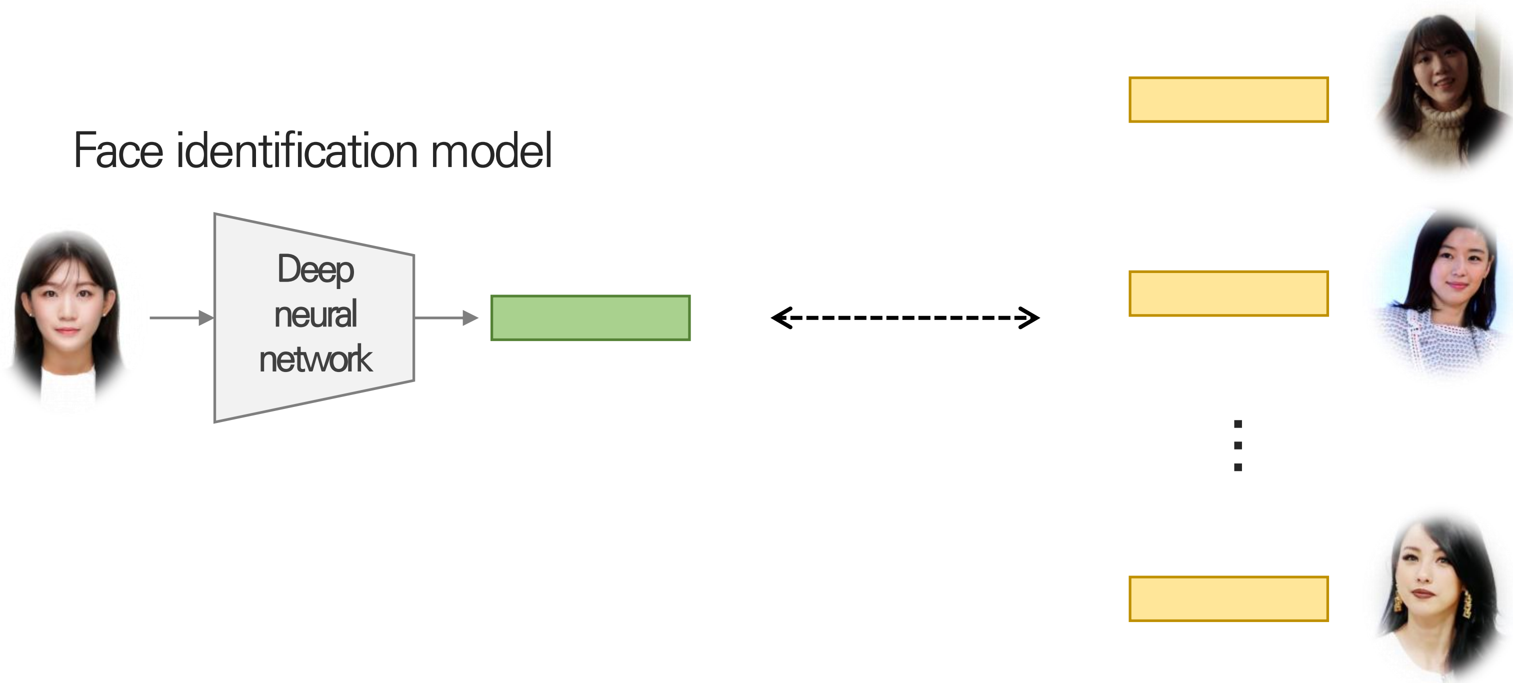
❖ 얼굴 인식 기술(face recognition)은 사진이나 동영상의 얼굴에서 정보를 파악하는 기술

❖ Deep metric learning이 우수한 성능을 보이는 computer vision 분야

❖ 얼굴 인식은 크게 face identification(식별), verification(검증)으로 나뉨

- **Face identification : 데이터베이스 내 인물 중 누구와 가장 유사한지 식별, 일대다 유사성 계산**
- Face verification : 동일인 여부 검증, 일대일 유사성 계산

Face identification model



# Softmax classifier로 학습된 feature의 한계

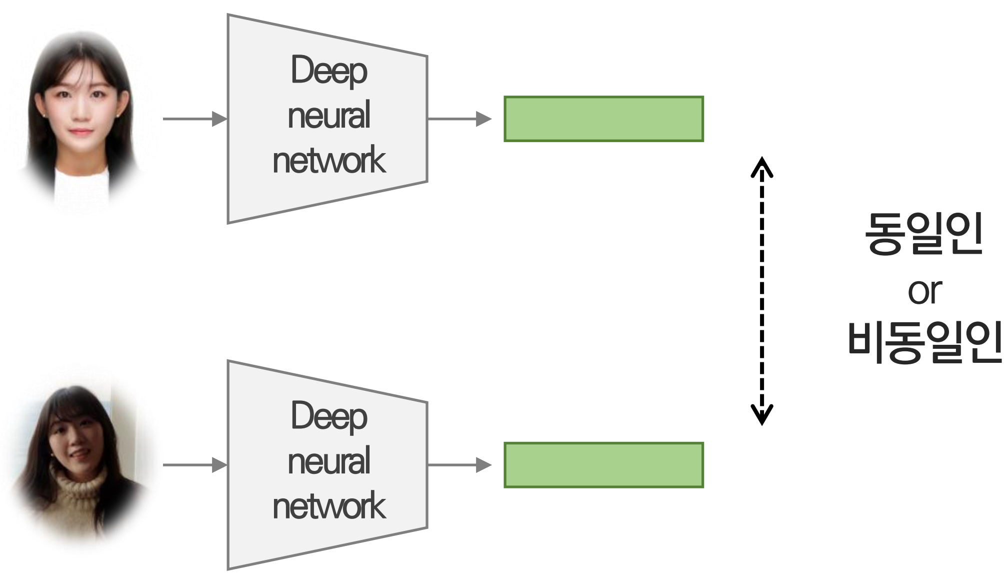
❖ 얼굴 인식 기술(face recognition)은 사진이나 동영상의 얼굴에서 정보를 파악하는 기술

❖ Deep metric learning이 우수한 성능을 보이는 computer vision 분야

❖ 얼굴 인식은 크게 face identification(식별), verification(검증)으로 나뉨

- Face identification : 데이터베이스 내 인물 중 누구와 가장 유사한지 식별, 일대다 유사성 계산
- **Face verification : 동일인 여부 검증, 일대일 유사성 계산**

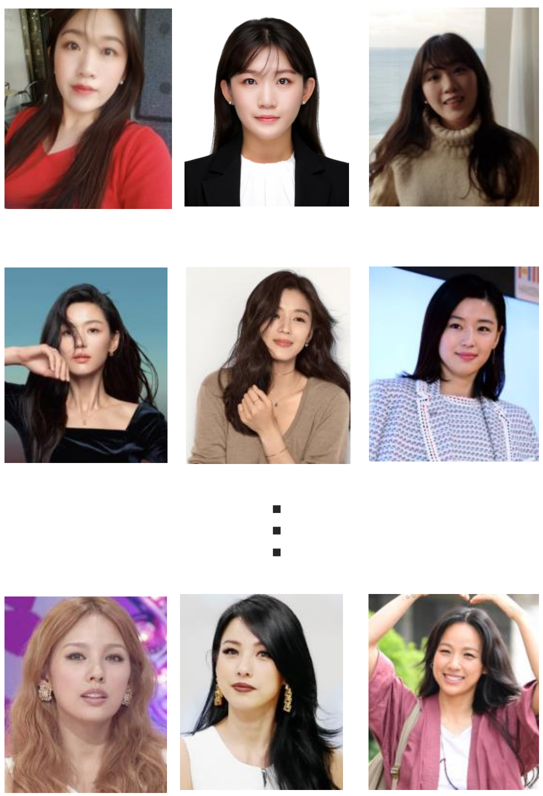
Face verification model



# Softmax classifier로 학습된 feature의 한계

❖ 기본적인 딥러닝 분류모델(Deep neural network + softmax classifier)의 한계

이미지 데이터



레이블

| | | | | |
|---|---|---|---|---|
| 1 | 0 | 0 | ⋯ | 0 |

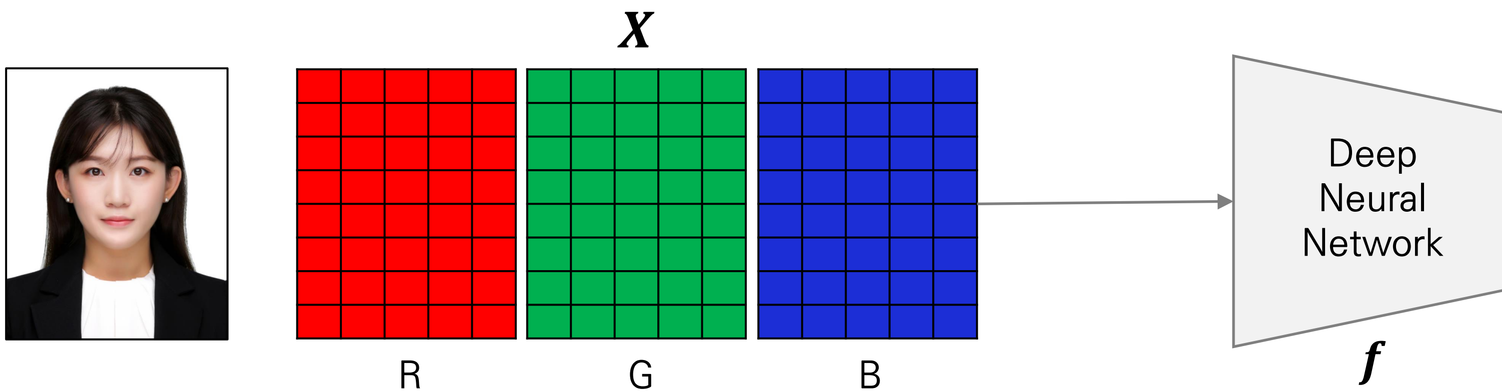
| | | | | |
|---|---|---|---|---|
| 0 | 1 | 0 | ⋯ | 0 |

K개 categories

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | ⋯ | 1 |

# Softmax classifier로 학습된 feature의 한계

❖ 기본적인 딥러닝 분류모델(Deep neural network + softmax classifier)의 한계




**요약벡터**
embedding vector

$$f(X) = Z$$

| $z_1$ | $z_2$ | $z_3$ | $\cdots$ | $z_K$ |
|---|---|---|---|---|
| 2 | 1 | 0.2 | $\cdots$ | -0.1 |

K개 categories

**소프트맥스 함수**
softmax
$\sigma(z)$

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

**확률값 $P$**
probability

| $p_1$ | $p_2$ | $p_3$ | $\cdots$ | $p_K$ |
|---|---|---|---|---|
| 0.3 | 0.12 | 0.01 | $\cdots$ | 0.00 |

**레이블 $L$**
label

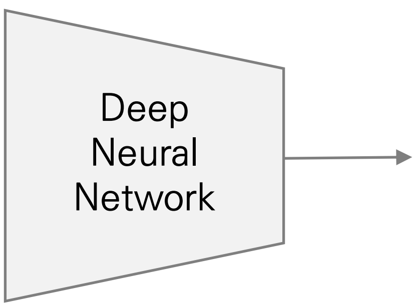
| $l_1$ | $l_2$ | $l_3$ | $\cdots$ | $l_K$ |
|---|---|---|---|---|
| 1 | 0 | 0 | $\cdots$ | 0 |

**손실함수 교차 엔트로피**
cross entropy

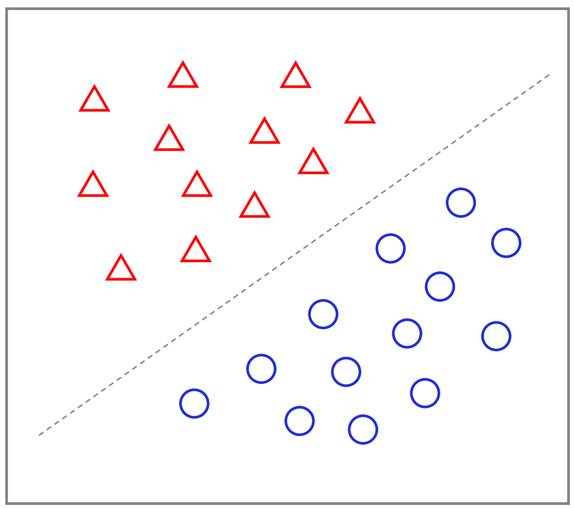
$$L(P, L) = \sum_{j=1}^{K} -l_j log(p_i)$$

# Softmax classifier로 학습된 feature의 한계

❖ 기본적인 딥러닝 분류모델(Deep neural network + softmax classifier)의 한계



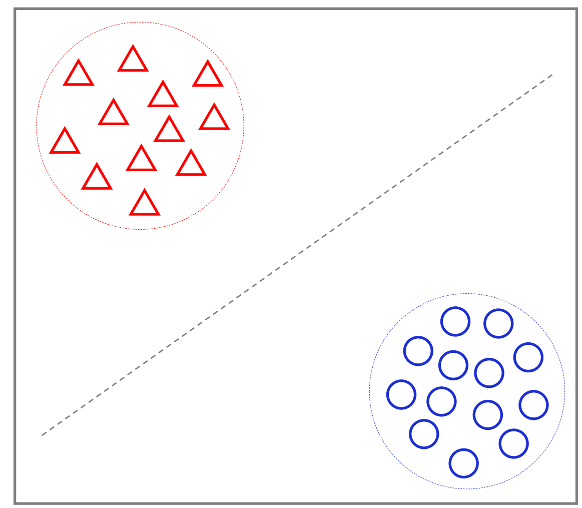

Softmax classifier로 형성되는
embedding space 특징



**Separable** features

적당히 구분 가능한 특징이 학습됨
– 학습에 사용되는 category의
수가 너무 많기 때문

정확한 얼굴 인식을 위해 필요한
embedding space 특징



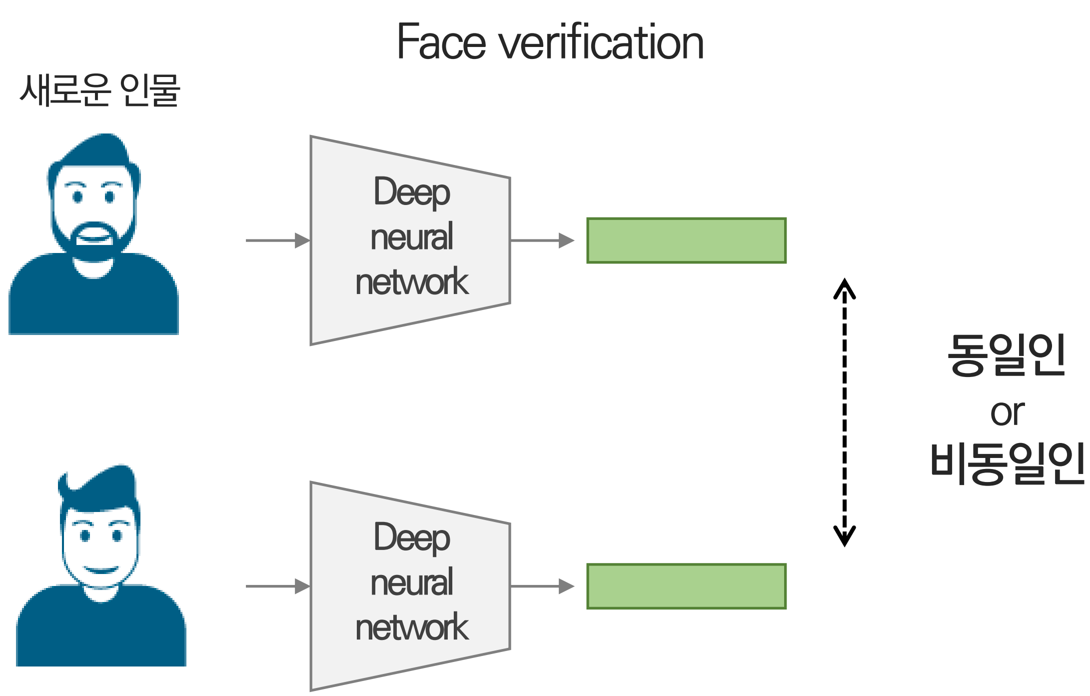
**Discriminative** features

inter–class variation 大
intra–class variation 小

# Softmax classifier로 학습된 feature의 한계

❖ 기본적인 딥러닝 분류모델(Deep neural network + softmax classifier)의 한계

## Openset 특징 추출 효과적 X



# Deep metric learning 손실함수 종류

- Contrastive loss – 2005 CVPR
- Triplet loss – 2015 CVPR
- Center loss – 2016 ECCV
- Additive angular margin loss – 2019 CVPR

# Deep metric learning 손실함수 종류

❖ Contrastive loss – 2005 CVPR

❖ 2021년 11월 8일 기준 3241회 인용

**Learning a Similarity Metric Discriminatively, with Application to Face Verification**

Sumit Chopra Raia Hadsell Yann LeCun

Courant Institute of Mathematical Sciences
New York University
New York, NY, USA
{sumit, raia, yann}@cs.nyu.edu

**Abstract**

*We present a method for training a similarity metric from data. The method can be used for recognition or verification applications where the number of categories is very large and not known during training, and where the number of training samples for a single category is very small. The idea is to learn a function that maps input patterns into a target space such that the $L_1$ norm in the target space approximates the "semantic" distance in the input space. The method is applied to a face verification task. The learning process minimizes a discriminative loss function that drives the similarity metric to be small for pairs of faces from the same person, and large for pairs from different persons. The mapping from raw to the target space is a convolutional network whose architecture is designed for robustness to geometric distortions. The system is tested on the Purdue/AR face database which has a very high degree of variability in the pose, lighting, expression, position, and artificial occlusions such as dark glasses and obscuring scarves.*

per category. A common approach to this kind of problem is distance-based methods, which consist in computing a similarity metric between the pattern to be classified or verified and a library of stored prototypes. Another common approach is to use non-discriminative (generative) probabilistic methods in a reduced-dimension space, where the model for one category can be trained without using examples from other categories. To apply discriminative learning techniques to this kind of application, we must devise a method that can extract information about the problem from the available data, without requiring specific information about the categories.

The solution presented in this paper is to *learn a similarity metric from data*. This similarity metric can later be used to compare or match new samples from previously-unseen categories (e.g. faces from people not seen during training). We present a new type of discriminative training method that is used to train the similarity metric. The method can be applied to classification problems where the number of categories is very large and/or where examples from all categories are not available at the time of training

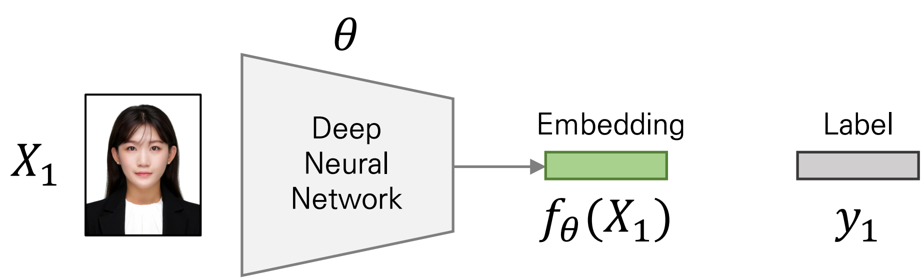
Chopra, S., Hadsell, R., & LeCun, Y. (2005, June). Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 539–546). IEEE.

# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$

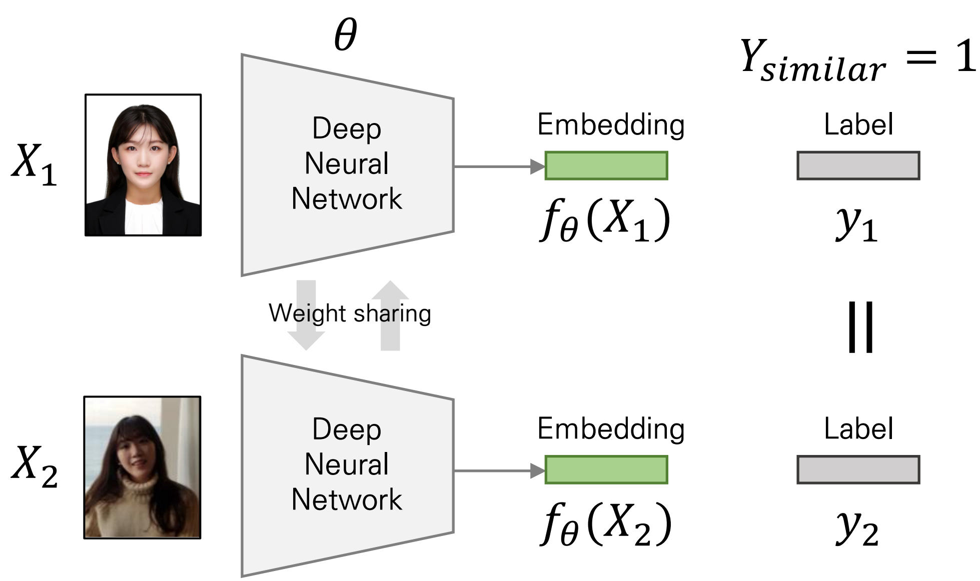


# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$

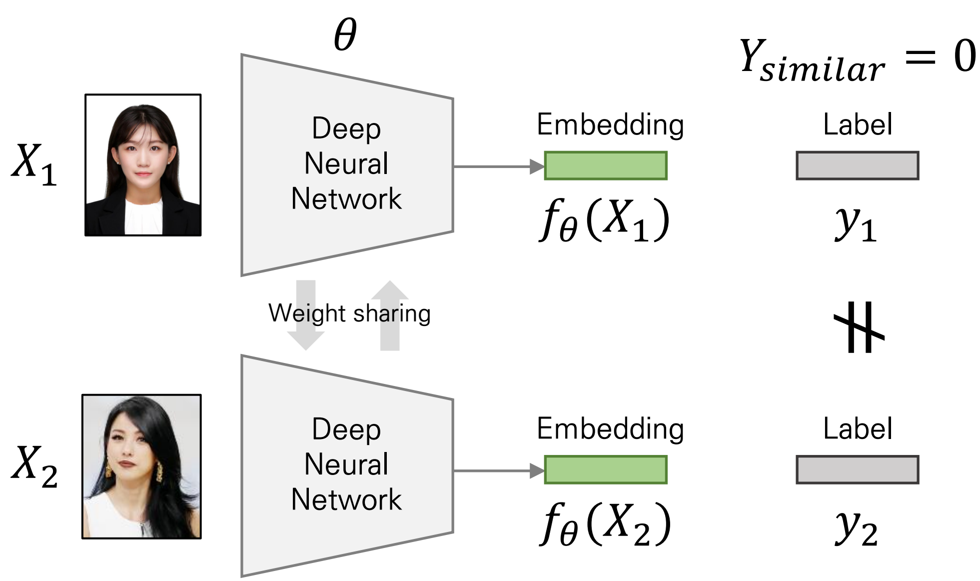


# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$



# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$
$$= Y_{similar}\left(\boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right) + (1 - Y_{similar})\max\left(0, m - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right)$$

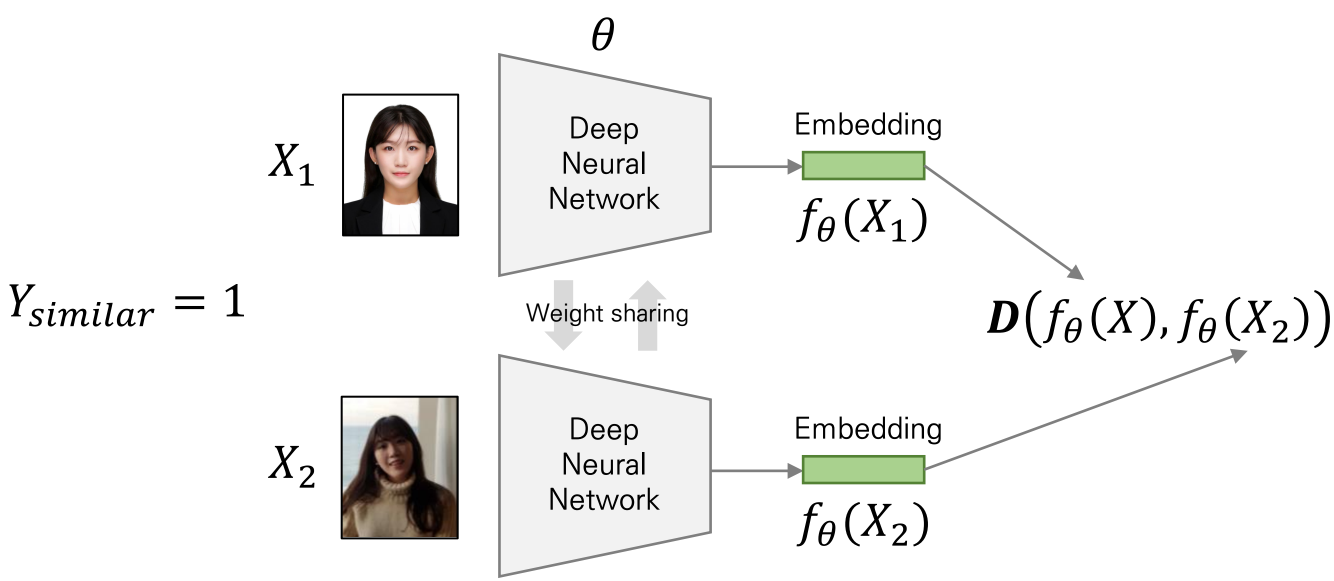
# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$

$$= Y_{similar}\left(\boldsymbol{D}^2\left(f_\theta(X_1), f_\theta(X_2)\right)\right) + \underbrace{(1 - Y_{similar})\max\left(0, m - \boldsymbol{D}^2\left(f_\theta(X_1), f_\theta(X_2)\right)\right)}_{=0}$$



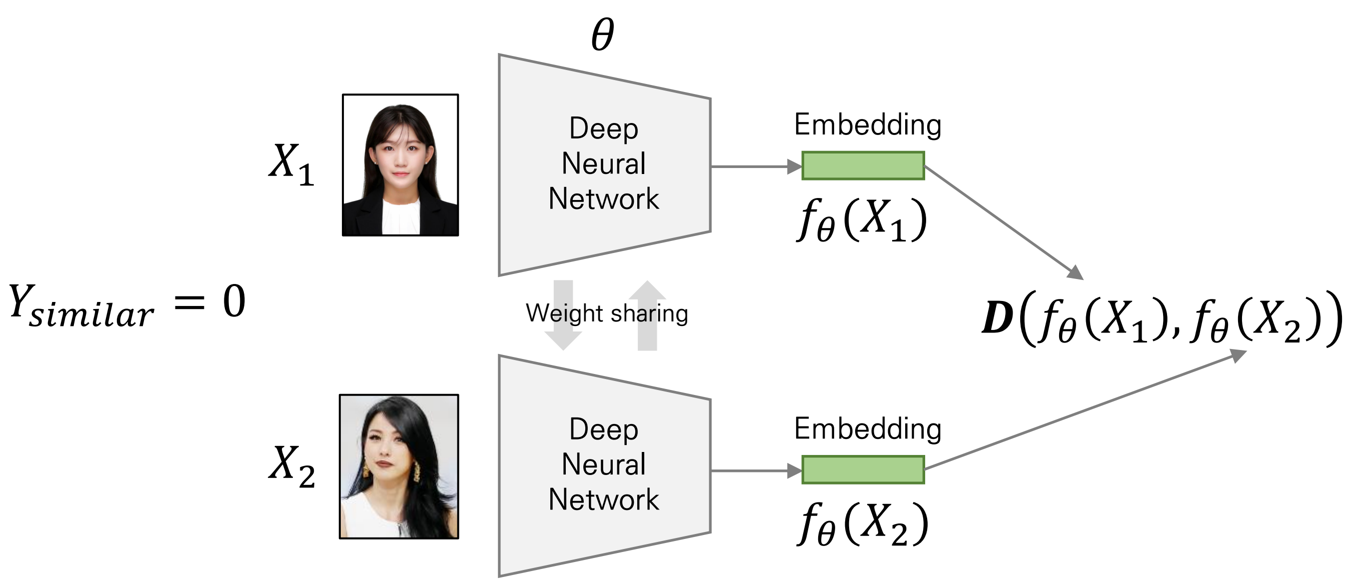
# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$

$$= \underbrace{Y_{similar}\left(\boldsymbol{D}^2(f_\theta(X_1), f_\theta(X_2))\right)}_{=0} + (1 - Y_{similar})\max\left(0, m - \boldsymbol{D}^2(f_\theta(X_1), f_\theta(X_2))\right)$$



# Deep metric learning 손실함수 종류

❖ Contrastive loss – 가장 간단하며 직관적인 손실함수

❖ 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 **유사하지 않은 데이터라면 거리가 멀어지도록** 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$

$$= \underbrace{Y_{similar}\left(\boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right)}_{=0} + (1 - Y_{similar}) \max\left(0, \boxed{m} - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right)$$

if $m < \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)$ then $\mathrm{m} - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big) \langle 0$

$m$ = margin
하이퍼파라미터

$$\max\left(0, m - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right) = 0$$

margin 보다 큰 거리 값을 가진다면 거리를 멀게 조정하지 않아도 됨

# Deep metric learning 손실함수 종류

- Contrastive loss – 가장 간단하며 직관적인 손실함수
- 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 **유사하지 않은 데이터라면 거리가 멀어지도록** 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$
$$= \underbrace{Y_{similar}\left(\boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right)}_{=0} + (1 - Y_{similar})\max\left(0, \boxed{m} - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right)$$

$m$ = margin
하이퍼파라미터

$$\text{if } m > \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big) \text{then m} - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big) \rangle\ 0$$

$$\max\left(0, m - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right) = m - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)$$

margin 보다 작은 거리 값을 가진다면 거리를 더 멀게 조정

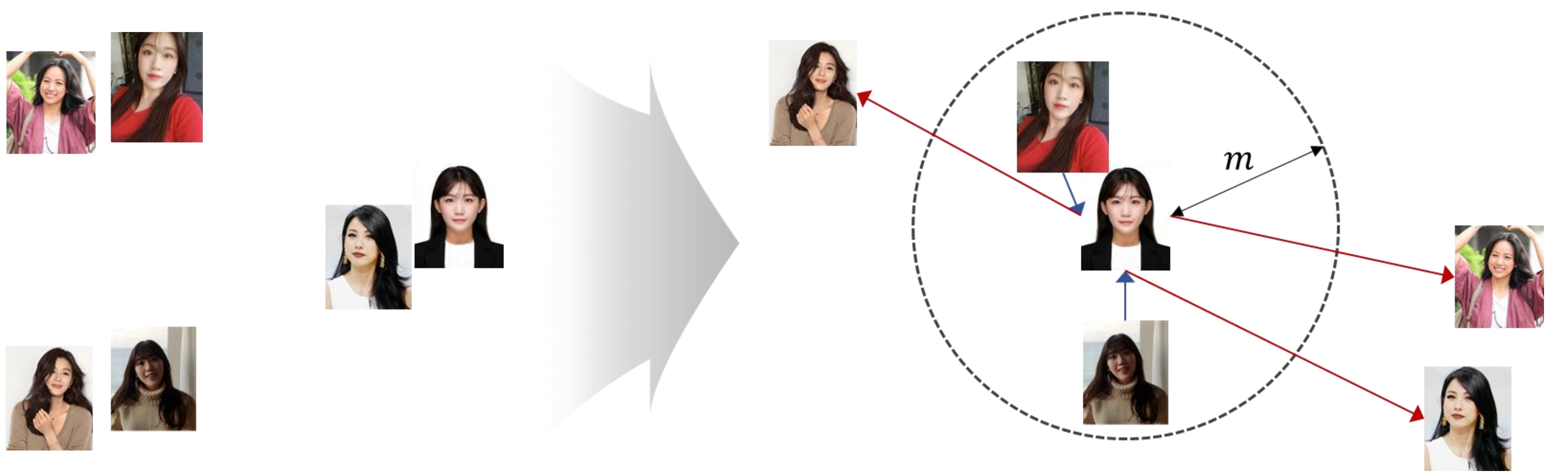
# Deep metric learning 손실함수 종류

- Contrastive loss – 가장 간단하며 직관적인 손실함수
- 두 개의 데이터가 유사하다면 거리가 작아지도록 두 개의 데이터가 유사하지 않은 데이터라면 거리가 멀어지도록 손실함수가 고안됨

$$L_{contrastive}(\theta, X_1, X_2, Y_{similar})$$
$$= Y_{similar}\left(\boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right) + (1 - Y_{similar})\max\left(0, m - \boldsymbol{D}^2\big(f_\theta(X_1), f_\theta(X_2)\big)\right)$$

입력 공간
Input space

잠재/임베딩 공간
Latent space
Embedding space



# Deep metric learning 손실함수 종류

- ❖ Triplet loss – 2015 CVPR
- ❖ 2021년 11월 8일 기준 9067회 인용



**FaceNet: A Unified Embedding for Face Recognition and Clustering**

Florian Schroff
fschroff@google.com
Google Inc.

Dmitry Kalenichenko
dkalenichenko@google.com
Google Inc.

James Philbin
jphilbin@google.com
Google Inc.

**Abstract**

*Despite significant recent advances in the field of face recognition [10, 14, 15, 17], implementing face verification and recognition efficiently at scale presents serious challenges to current approaches. In this paper we present a system, called FaceNet, that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. Once this space has been produced, tasks such as face recognition, verification and clustering can be easily implemented using standard techniques with FaceNet embeddings as feature vectors.*

*Our method uses a deep convolutional network trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer as in previous deep learning approaches. To train, we use triplets of roughly aligned matching / non-matching face patches generated using a novel online triplet mining method. The benefit of our approach is much greater representational efficiency: we achieve state-of-the-art face recognition performance using only 128-bytes per face.*

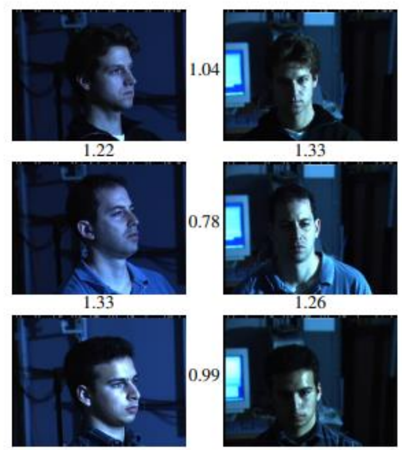


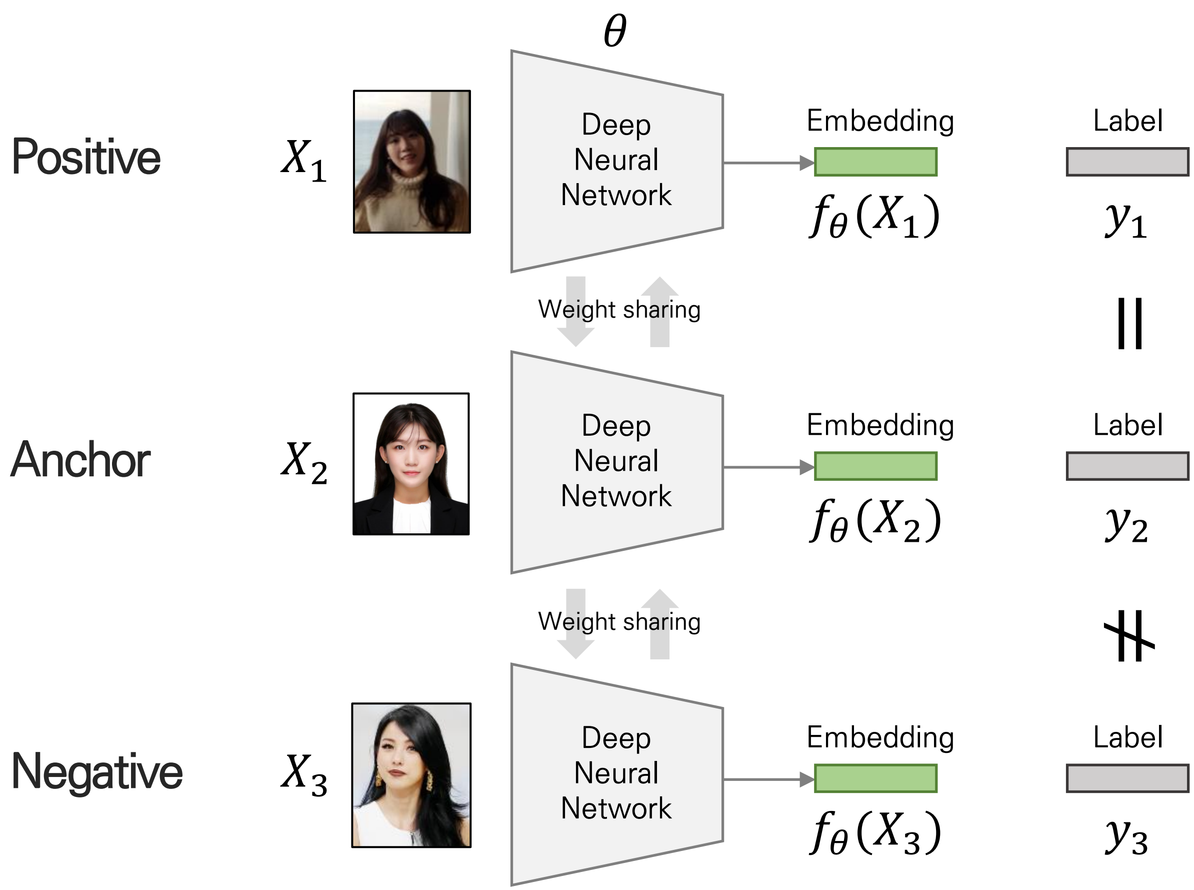

Figure 1. **Illumination and Pose invariance.** Pose and illumina-

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

# Deep metric learning 손실함수 종류

❖ Triplet loss – 가장 유명하고 널리 활용되는 손실함수

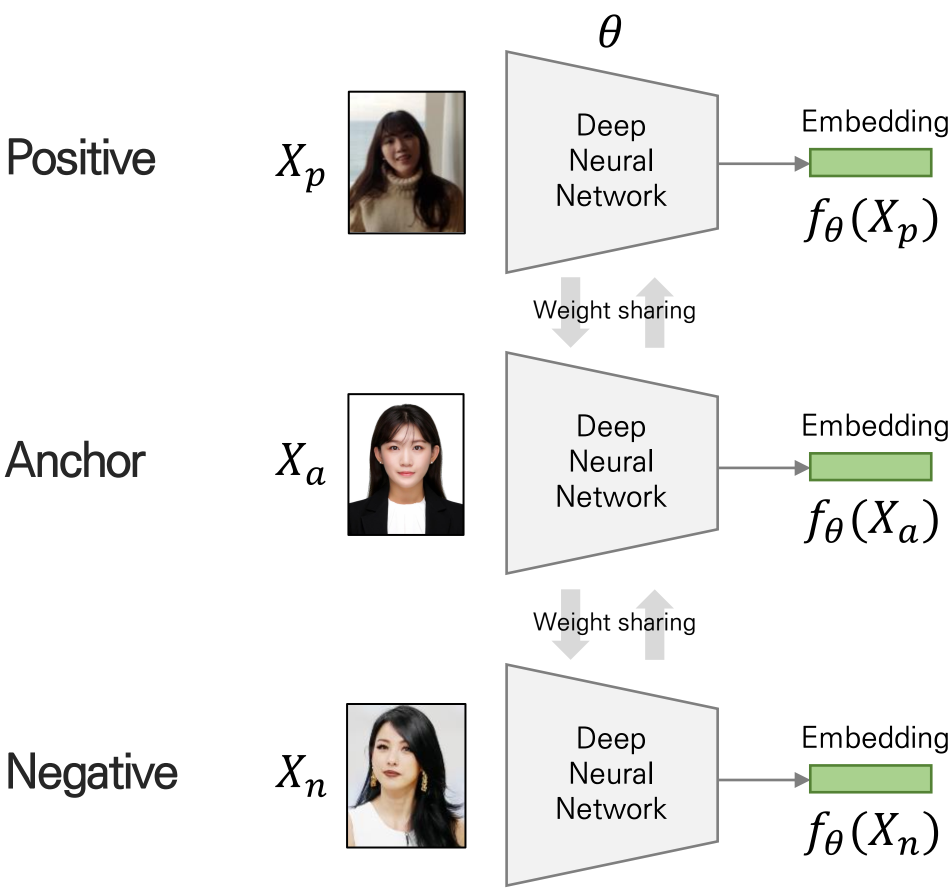
❖ 세 개의 데이터(anchor, positive, negative) 사용, positive pair보다 negative pair의 거리가 더 멀어지도록 손실함수가 고안됨



# Deep metric learning 손실함수 종류

❖ Triplet loss – 가장 유명하고 널리 활용되는 손실함수

❖ 세 개의 데이터(anchor, positive, negative) 사용, positive pair보다 negative pair의 거리가 더 멀어지도록 손실함수가 고안됨




$X_p$ → $f_\theta(X_p)$

$X_a$ → $f_\theta(X_a)$

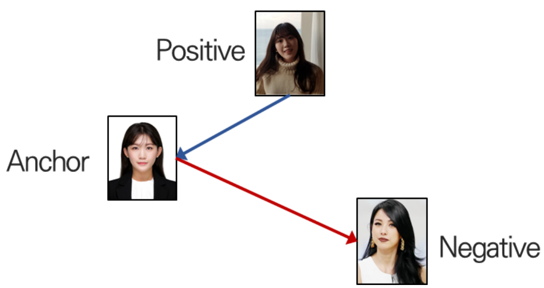
$X_n$ → $f_\theta(X_n)$

$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_p)\right) < \boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_n)\right)$$

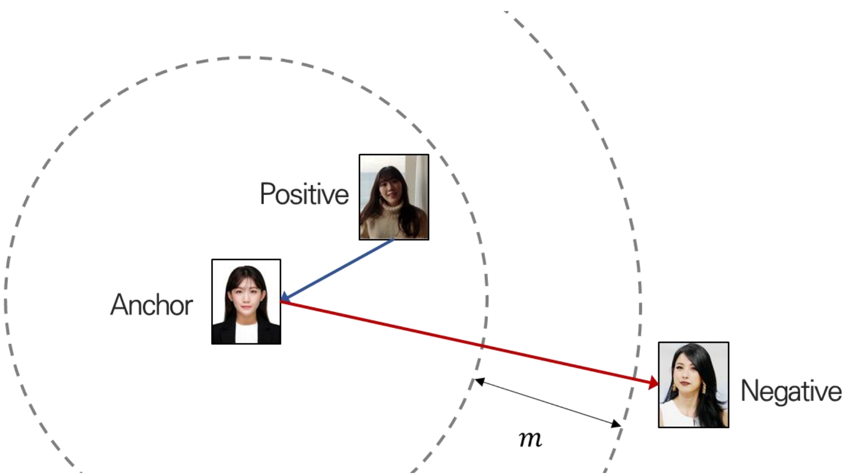
$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_p)\right) \boxed{+\, m} < \boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_n)\right)$$

# Deep metric learning 손실함수 종류

- Triplet loss – 가장 유명하고 널리 활용되는 손실함수
- 세 개의 데이터(anchor, positive, negative) 사용, positive pair보다 negative pair의 거리가 더 멀어지도록 손실함수가 고안됨




$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_p)\right) < \boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_n)\right)$$




$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_p)\right) \boxed{+\, m} < \boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_n)\right)$$

# Deep metric learning 손실함수 종류

❖ Triplet loss – 가장 유명하고 널리 활용되는 손실함수

❖ 세 개의 데이터(anchor, positive, negative) 사용, positive pair보다 negative pair의 거리가 더 멀어지도록 손실함수가 고안됨

$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_p)\right) \boxed{+\, m}$$

$$<$$

$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_n)\right)$$

$$\boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_p)\right) + m - \boldsymbol{D}\left(f_\theta(X_a), f_\theta(X_n)\right) < 0$$

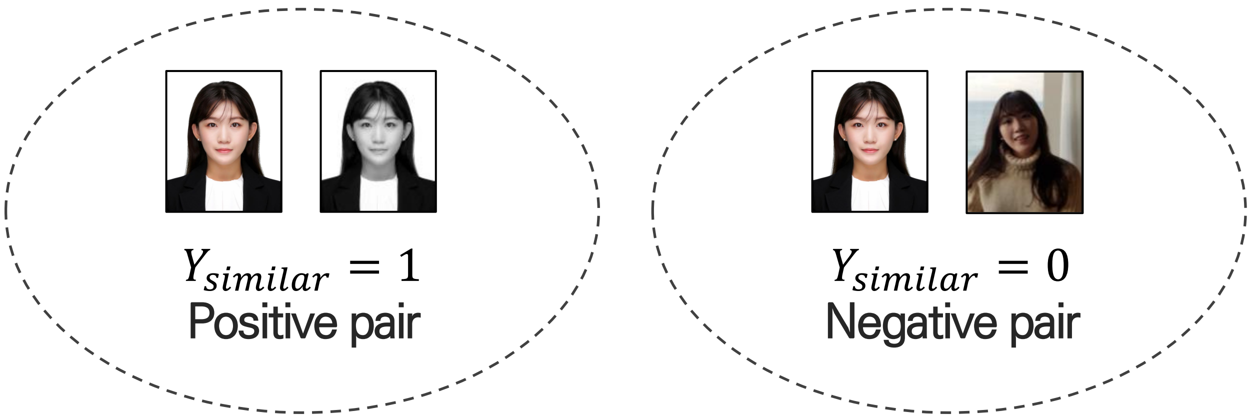
$$L_{triplet}(\theta, X_1, X_2) =$$

$$\max\left(0, \boldsymbol{D}^2\left(f_\theta(X_a), f_\theta(X_p)\right) - \boldsymbol{D}^2\left(f_\theta(X_a), f_\theta(X_n)\right) + m\right)$$

# Deep metric learning 손실함수 종류

- ❖ 유사한지 유사하지 않은지 = 같은 class에 속하는지 아닌지 → class label 필요 → supervised
- ❖ 유사한지 유사하지 않은지 = 같은 이미지에서 변형된 이미지인지 아닌지

  → class label 필요 X → self-supervised learning 자가지도학습




Self-supervised learning (자기지도학습)에서 손실함수로 활용

# Deep metric learning 손실함수 종류

- ❖ 유사한지 유사하지 않은지 = 같은 class에 속하는지 아닌지 → class label 필요 → supervised
- ❖ 유사한지 유사하지 않은지 = 같은 이미지에서 변형된 이미지인지 아닌지
  → class label 필요 X → self-supervised learning 자가지도학습

## Self-supervised learning (자기지도학습) 관련 세미나 참고


Deal with Contrastive Learning

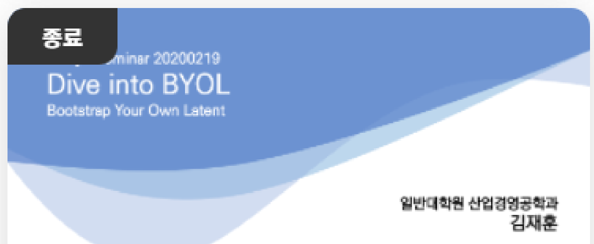
발표자: 고은성

2021년 9월 10일

오전 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶




Dive into BYOL

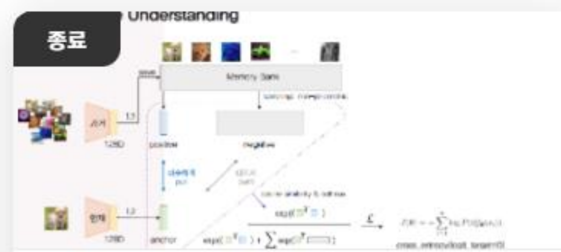
발표자: 김재훈

2021년 2월 19일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶




Towards Contrastive Learning

발표자: 곽민구

2021년 1월 29일

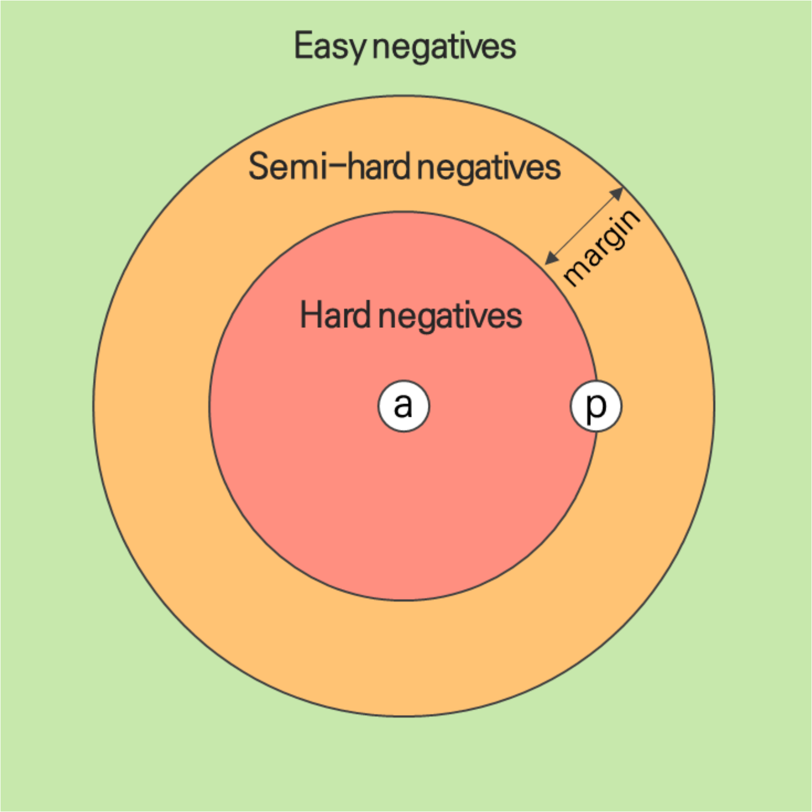
오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

# Deep metric learning 손실함수 종류

- Triplet(anchor, positive, negative) mining이 성능에 영향을 많이 줌
- 어떤 negative인지에 따라서 성능/계산량 차이가 크게 남
- FaceNet논문에서는 semi-hard negatives, positive 쌍으로 모델 학습함



# Deep metric learning 손실함수 종류

- ❖ Center loss – 2016 ECCV
- ❖ 2021년 11월 8일 기준 2592회 인용

## A Discriminative Feature Learning Approach for Deep Face Recognition

Yandong Wen[1], Kaipeng Zhang[1], Zhifeng Li[1(✉)], and Yu Qiao[1,2]

[1] Shenzhen Key Lab of Computer Vision and Pattern Recognition,
Shenzhen Institutes of Advanced Technology, CAS, Shenzhen, China
yandongw@andrew.cmu.edu, {kp.zhang,zhifeng.li,yu.qiao}@siat.ac.cn
[2] The Chinese University of Hong Kong, Sha Tin, Hong Kong

**Abstract.** Convolutional neural networks (CNNs) have been widely used in computer vision community, significantly improving the state-of-the-art. In most of the available CNNs, the softmax loss function is used as the supervision signal to train the deep model. In order to enhance the discriminative power of the deeply learned features, this paper proposes a new supervision signal, called center loss, for face recognition task. Specifically, the center loss simultaneously learns a center for deep features of each class and penalizes the distances between the deep features and their corresponding class centers. More importantly, we prove that the proposed center loss function is trainable and easy to optimize in the CNNs. With the joint supervision of softmax loss and center loss, we can train a robust CNNs to obtain the deep features with the two key learning objectives, inter-class dispension and intra-class compactness as much as possible, which are very essential to face recognition. It is encouraging to see that our CNNs (with such joint supervision) achieve the state-of-the-art accuracy on several important face recog-

Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016, October). A discriminative feature learning approach for deep face recognition. In European conference on computer vision (pp. 499–515). Springer, Cham.

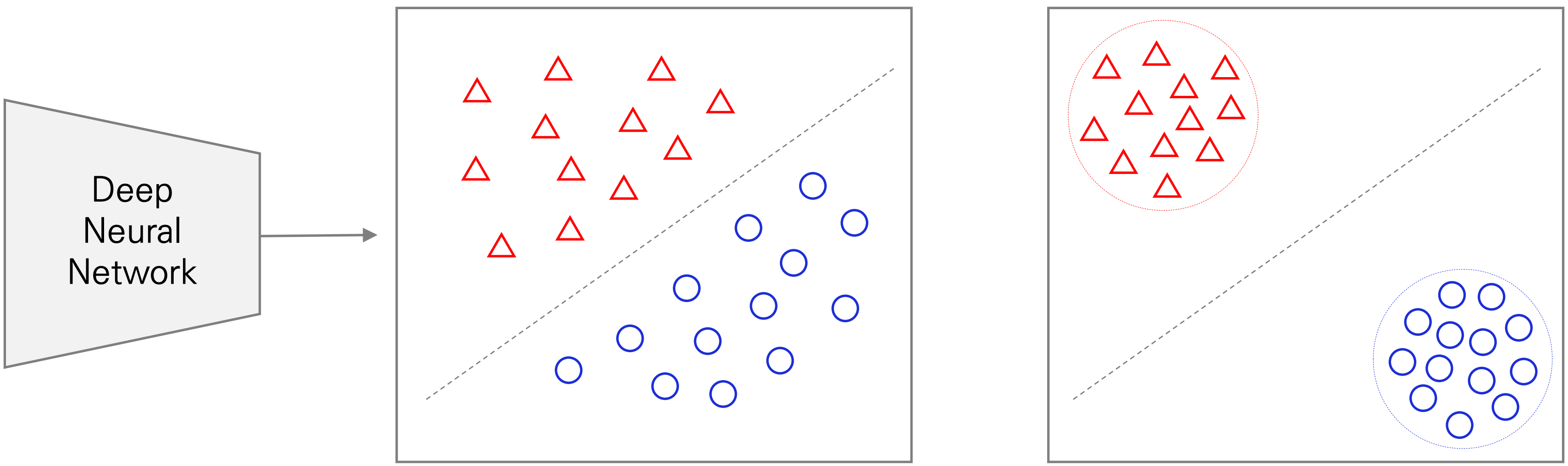
# Deep metric learning 손실함수 종류

❖ Center loss – 2016 ECCV

❖ 2021년 11월 8일 기준 2592회 인용

Softmax classifier로 형성되는 embedding space 특징

정확한 얼굴 인식을 위해 필요한 embedding space 특징




**Separable features**

적당히 구분 가능한 특징이 학습됨
– 학습에 사용되는 category의
수가 너무 많기 때문

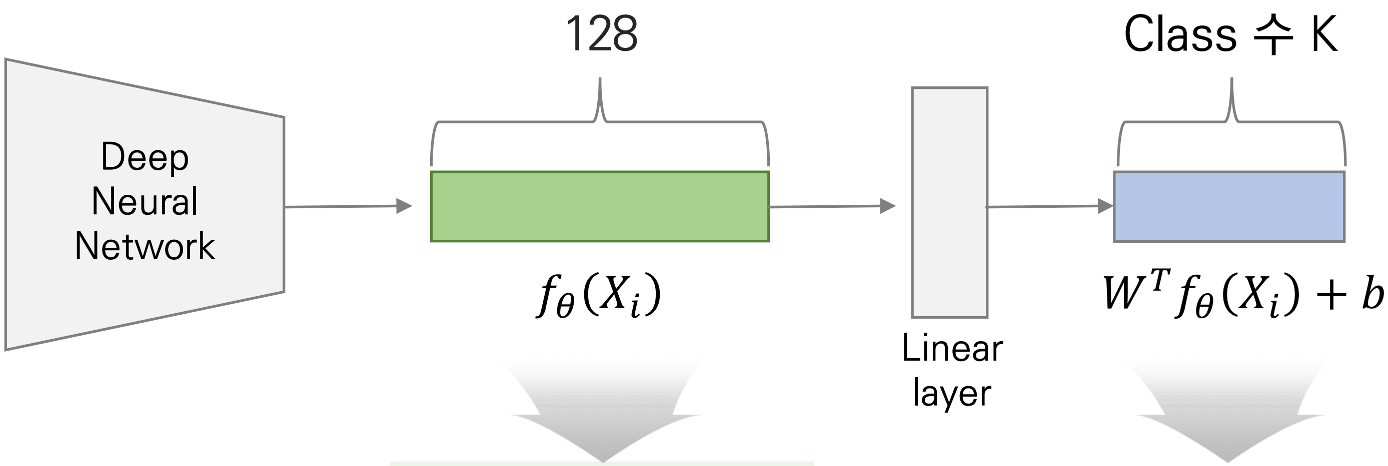
**Discriminative features**

inter-class variation 大
intra-class variation 小

# Deep metric learning 손실함수 종류

❖ Center loss – softmax(inter–class variation 大)에 정규식(intra–class variation 小) 추가

$$L_{center} = L_{softmax} + \frac{\lambda}{2}\sum_{i=1}^{N}\left\|f_\theta(X_i) - c_{y_i}\right\|_2^2$$




$$\frac{\lambda}{2}\sum_{i=1}^{N}\left\|f_\theta(X_i) - c_{y_i}\right\|_2^2$$

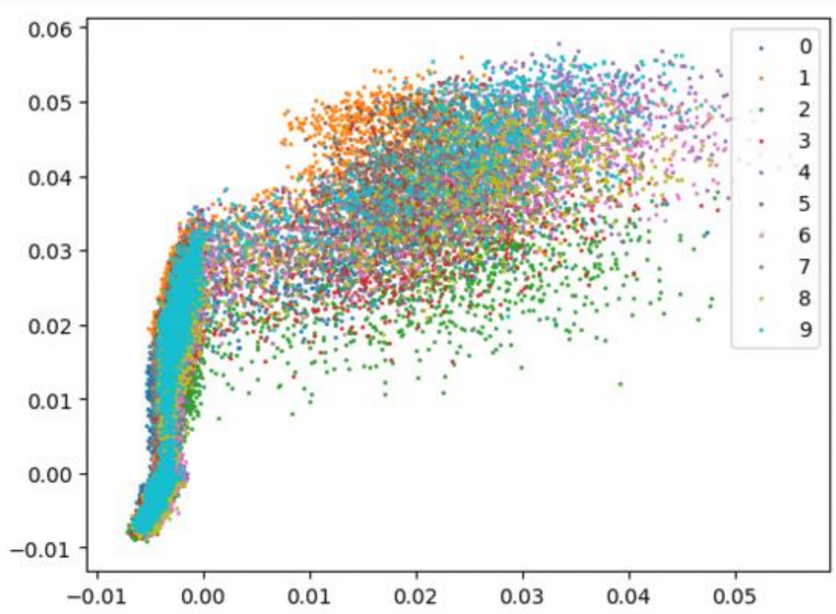
$$L_{softmax}$$

intra–class variation 小

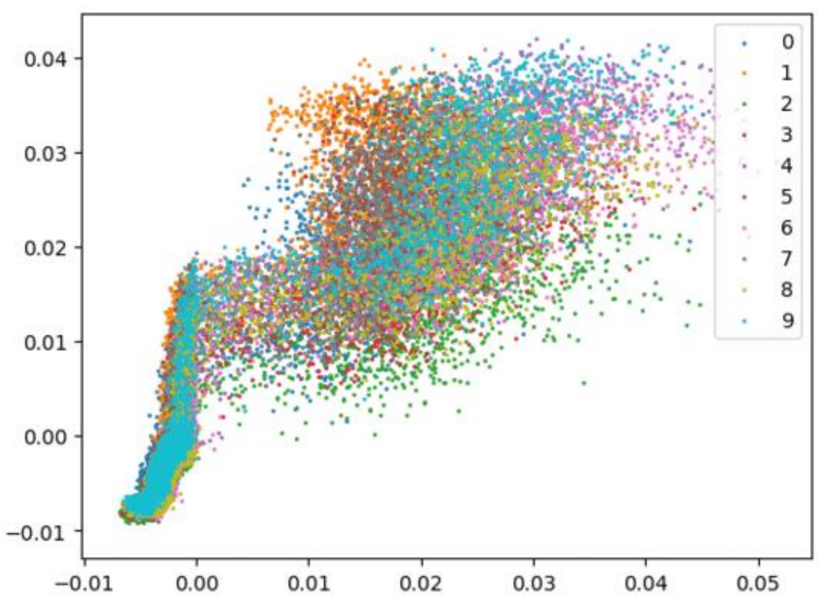
inter–class variation 大

# Deep metric learning 손실함수 종류

❖ Center loss – softmax(inter–class variation 大)에 정규식(intra–class variation 小) 추가




$L_{softmax}$




$L_{center}$

# Deep metric learning 손실함수 종류

- ❖ Additive angular margin loss – 2019 CVPR
- ❖ 2021년 11월 8일 기준 2120회 인용



**ArcFace: Additive Angular Margin Loss for Deep Face Recognition**

Jiankang Deng * [1,2,3] Jia Guo * [2] Niannan Xue[1] Stefanos Zafeiriou[1,3]
[1]Imperial College London [2]InsightFace [3]FaceSoft
{j.deng16, n.xue15, s.zafeiriou}@imperial.ac.uk , guojia@gmail.com

**Abstract**

*One of the main challenges in feature learning using Deep Convolutional Neural Networks (DCNNs) for large-scale face recognition is the design of appropriate loss functions that can enhance the discriminative power. Centre loss penalises the distance between deep features and their corresponding class centres in the Euclidean space to achieve intra-class compactness. SphereFace assumes that the linear transformation matrix in the last fully connected layer can be used as a representation of the class centres in the angular space and therefore penalises the angles between deep features and their corresponding weights in a multiplicative way. Recently, a popular line of research is to incorporate margins in well-established loss functions in order to maximise face class separability. In this paper, we propose an Additive Angular Margin Loss (ArcFace) to obtain highly discriminative features for face recognition.*

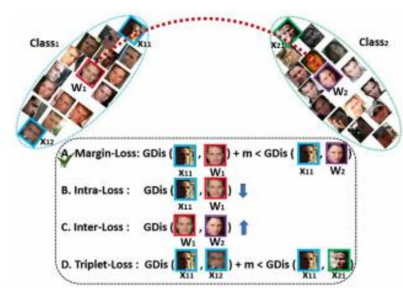
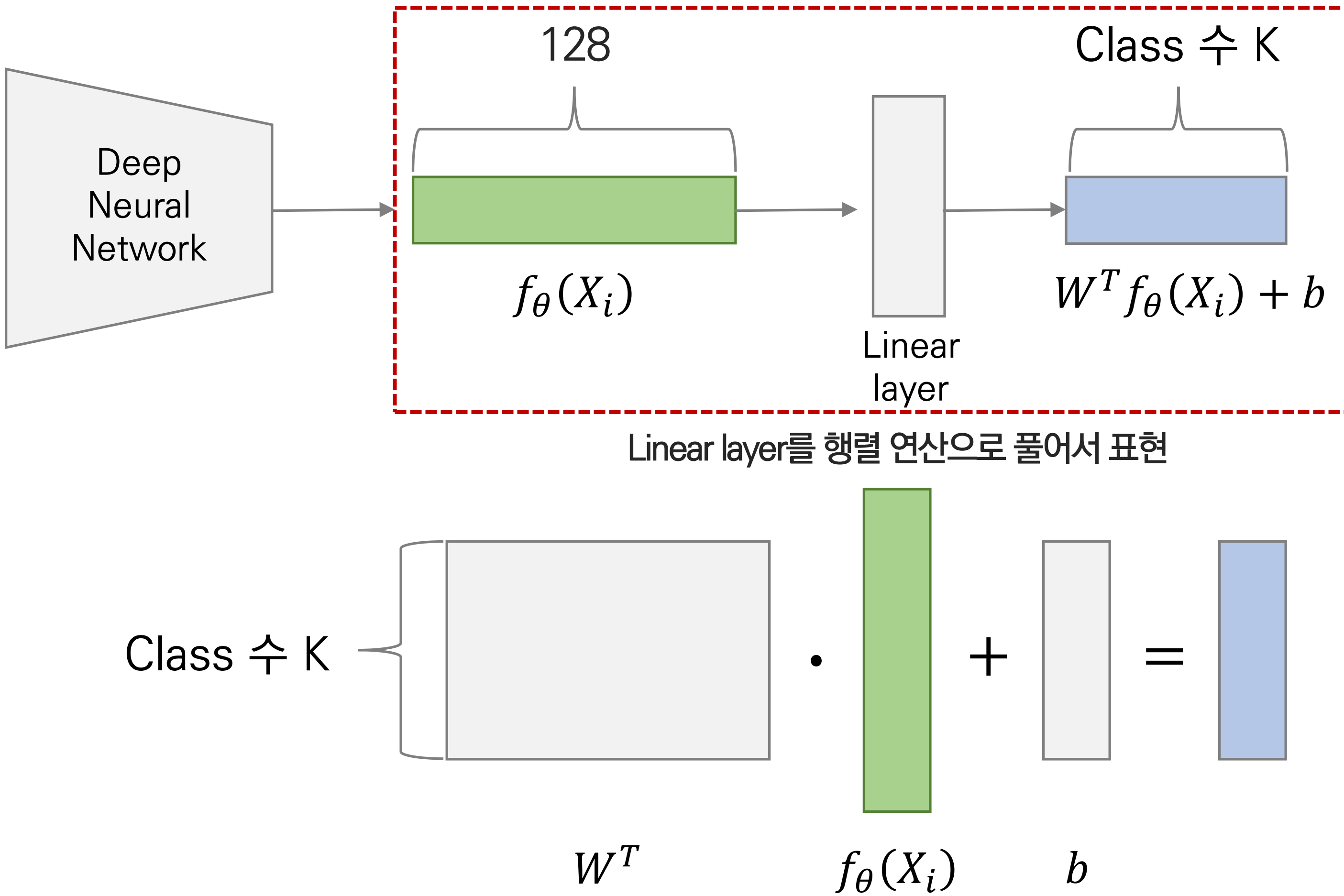



Figure 1. Based on the centre [15] and feature [35] normalisation, all identities are distributed on a hypersphere. To enhance intra-class compactness and inter-class discrepancy, we consider four kinds of Geodesic Distance (GDis) constraint. (A) Margin-Loss: insert a geodesic distance margin between the sample and centres. (B) Intra-Loss: decrease the geodesic distance between the sample and the corresponding centre. (C) Inter-Loss: increase the

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4690-4699).

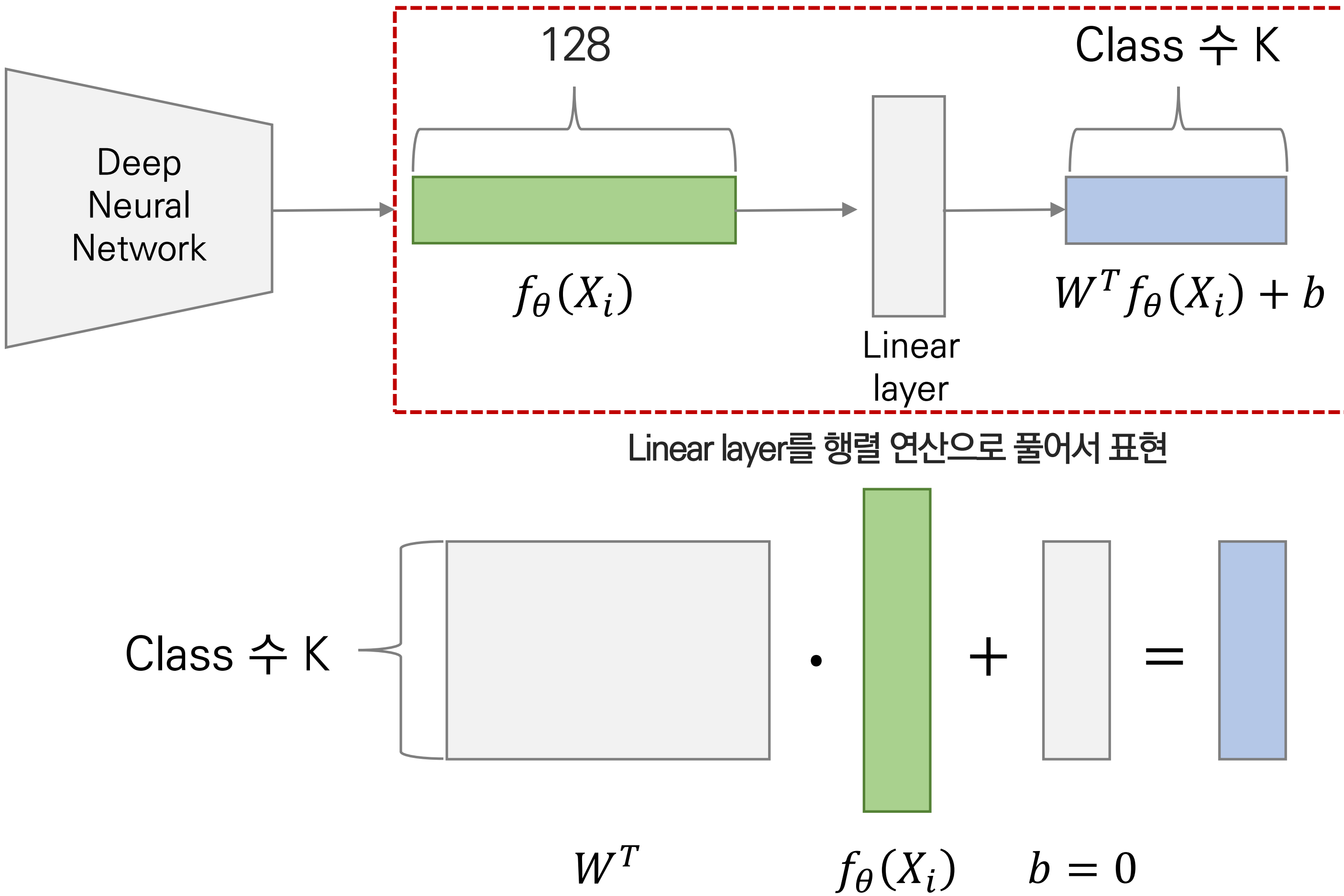
# Deep metric learning 손실함수 종류

❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시



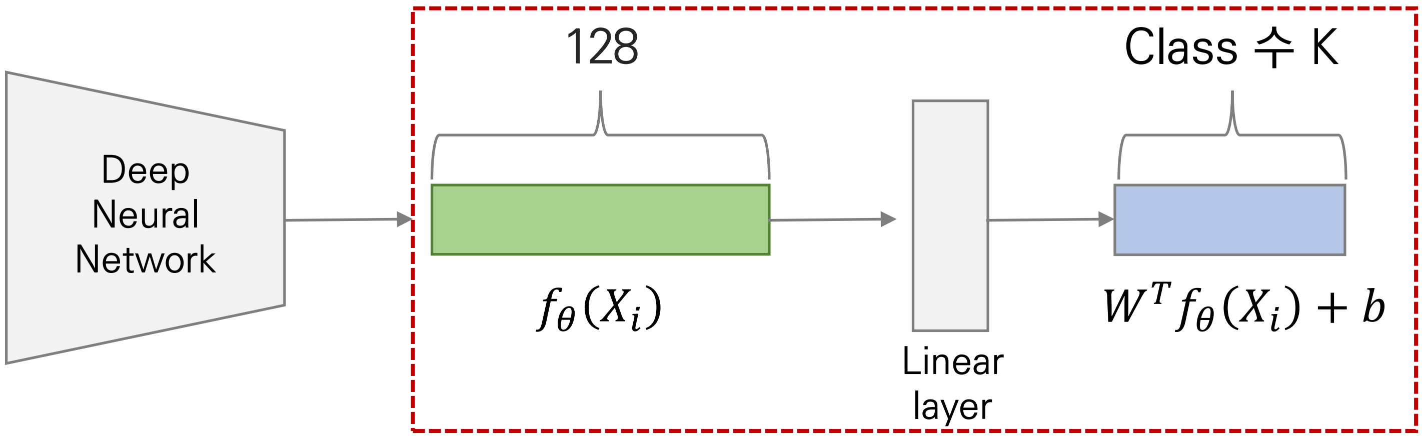
# Deep metric learning 손실함수 종류

❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시

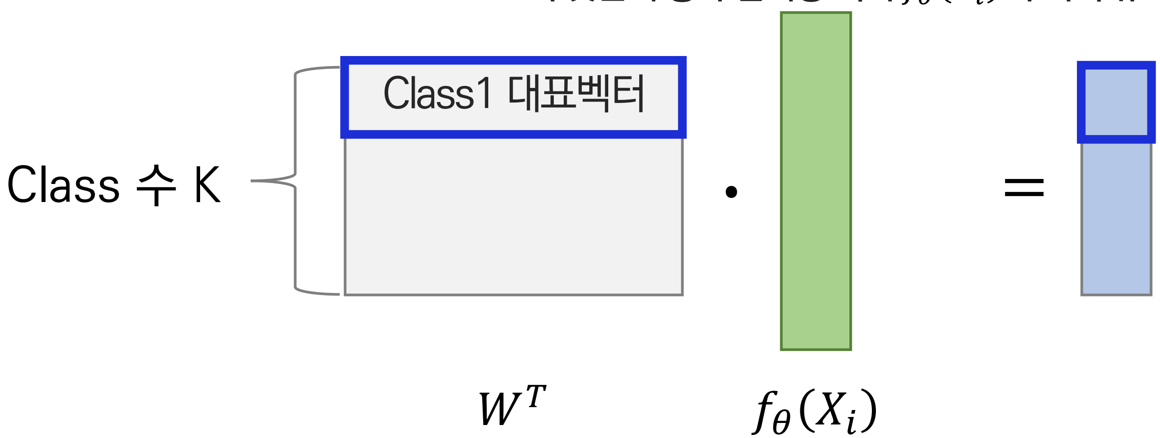


# Deep metric learning 손실함수 종류

❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시

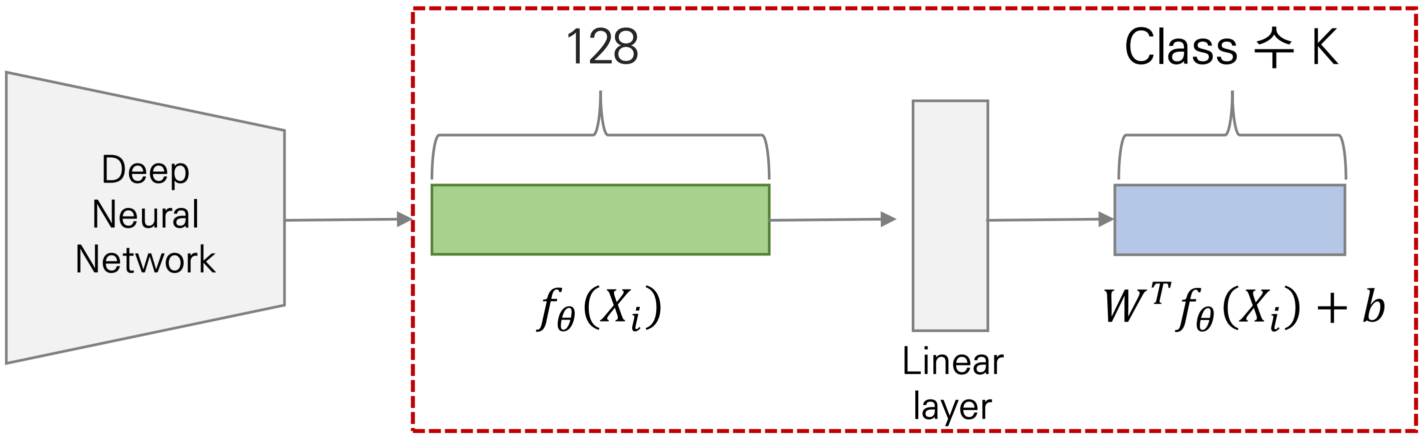



$W^T$의 첫번째 행과 임베딩 벡터 $f_\theta(X_i)$의 내적 유사도

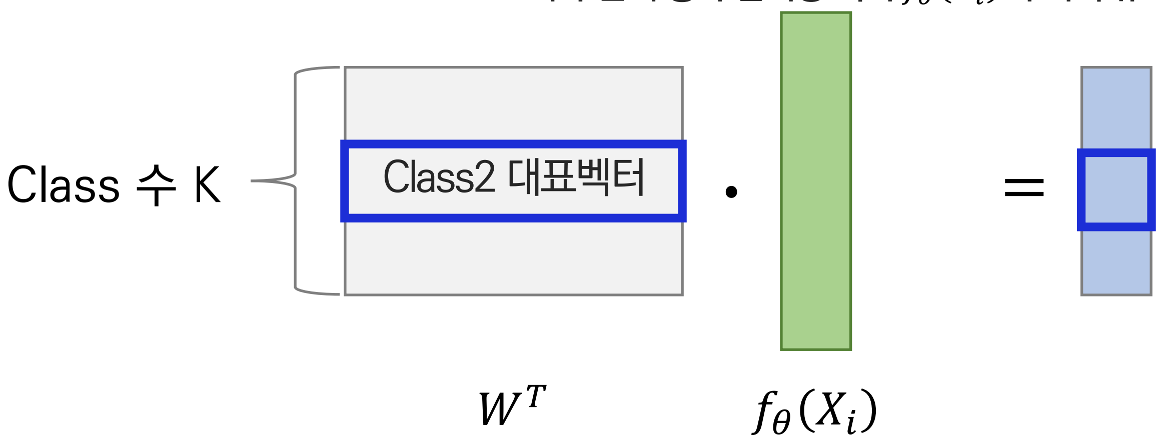


# Deep metric learning 손실함수 종류

❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시

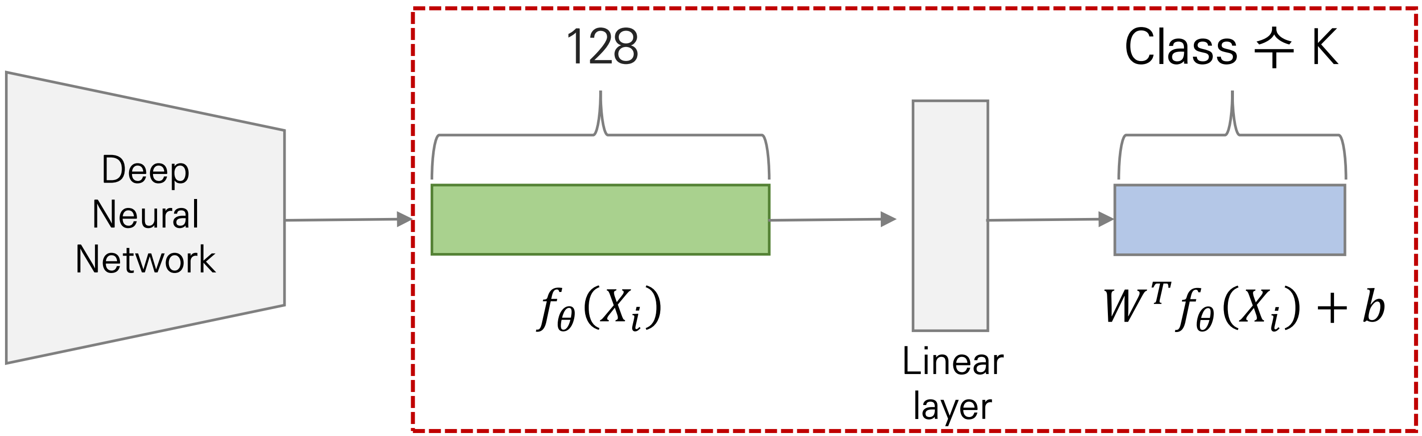



$W^T$의 두번째 행과 임베딩 벡터 $f_\theta(X_i)$의 내적 유사도

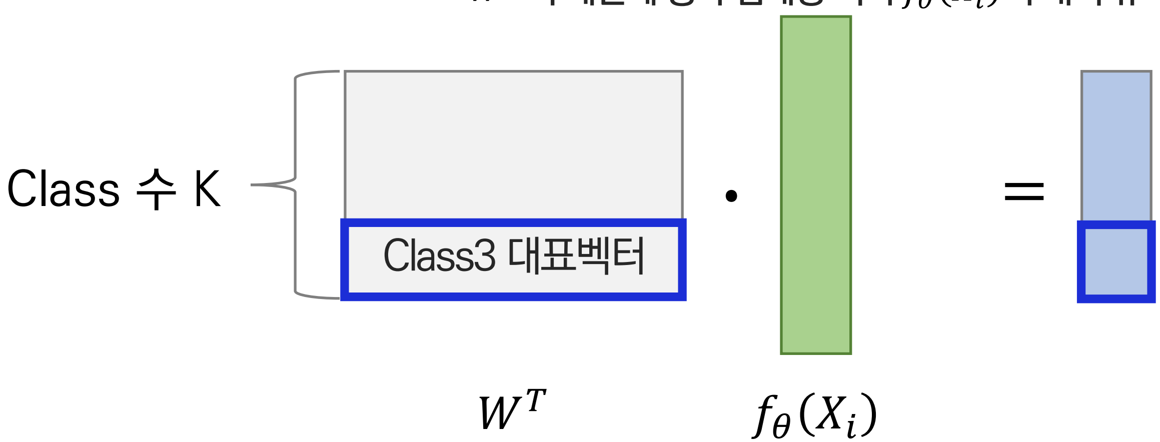


# Deep metric learning 손실함수 종류

❖ Angular softmax – softmax 함수를 변형한 새로운 해석 제시

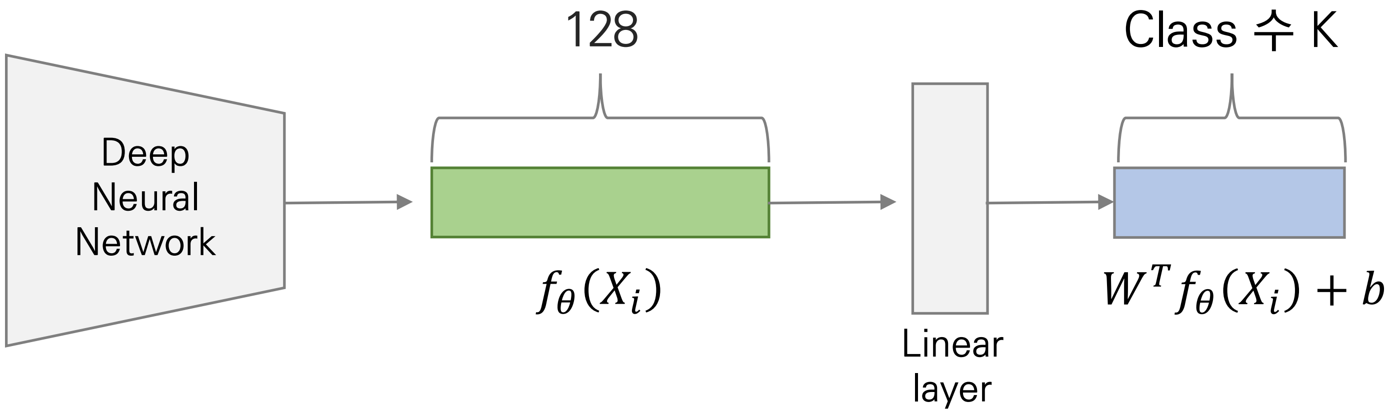



$W^T$의 세번째 행과 임베딩 벡터 $f_\theta(X_i)$의 내적 유사도



# Deep metric learning 손실함수 종류

❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시




$$L_{softmax}$$

$$= -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{W_{yi}^T f_\theta(X_i) + b\}}{\sum_{j=1}^{K} \exp\{W_j^T f_\theta(X_i) + b\}}$$

각 class 대표벡터와 유사도 측정
실제 정답 class 대표벡터와의 유사도가
가장 크도록 학습하는 metric learning

$$= -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{W_{y_i} \cdot f_\theta(X_i)\}}{\sum_{j=1}^{K} \exp\{W_j \cdot f_\theta(X_i)\}}$$

# Deep metric learning 손실함수 종류

❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시

$$L_{softmax} = -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{W_{yi}^{T} f_{\theta}(X_i) + b\}}{\sum_{j=1}^{K} \exp\{W_j^{T} f_{\theta}(X_i) + b\}} = -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{W_{y_i} \cdot f_{\theta}(X_i)\}}{\sum_{j=1}^{K} \exp\{W_j \cdot f_{\theta}(X_i)\}}$$

내적의 기하학적 정의 사용

$$= -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{\|W_{y_i}\| \|f_{\theta}(X_i)\| \cos(\theta_{y_i,i})\}}{\sum_{j=1}^{K} \exp\{\|W_j\| \|f_{\theta}(X_i)\| \cos(\theta_{j,i})\}}$$

$\|W_j\| = 1, j \in 1, \ldots K$ class 대표벡터 크기를 모두 1로 정규화

$\|f_{\theta}(X_i)\|$ 크기를 s로 정규화

$$= -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{s \cdot \cos(\theta_{y_i,i})\}}{\sum_{j=1}^{K} \exp\{s \cdot \cos(\theta_{j,i})\}}$$

additive angular margin penalty m

inter-class variation 大

intra-class variation 小

$$= -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{s \cdot \cos(\theta_{y_i,i} \boxed{+ m})\}}{\exp\{s \cdot \cos(\theta_{y_i,i} \boxed{+ m})\} + \sum_{j=1, j \neq y_i}^{K} \exp\{s \cdot \cos(\theta_{j,i})\}}$$

# Deep metric learning 손실함수 종류

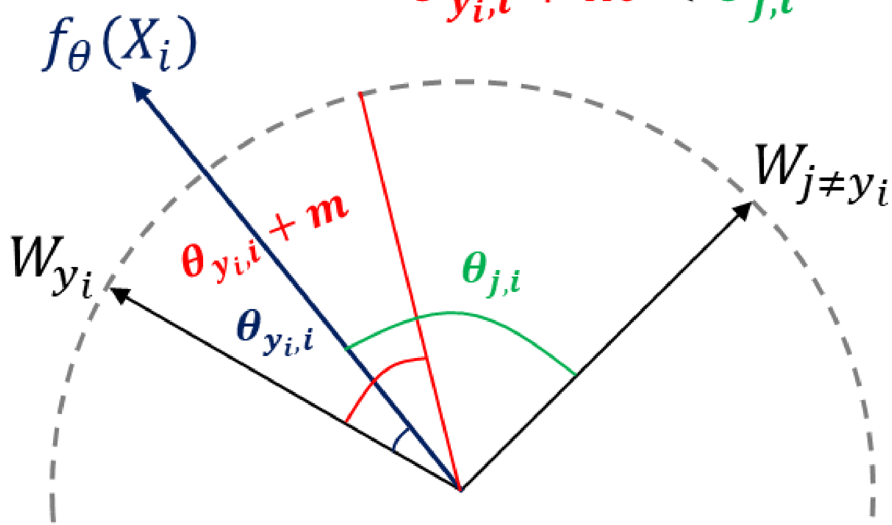
❖ Additive angular margin loss – softmax 함수를 변형한 새로운 해석 제시

additive angular margin penalty m

inter–class variation 大
intra–class variation 小

$$= -\frac{1}{N}\sum_{i=1}^{N} log \frac{\exp\{s \cdot \cos(\theta_{y_i,i} \boxed{+\, m})\}}{\exp\{s \cdot \cos(\theta_{y_i,i} \boxed{+\, m})\} + \sum_{j=1,j\neq y_i}^{K} \exp\{s \cdot \cos(\theta_{j,i})\}}$$

$$\theta_{y_i,i} + m < \theta_{j,i}$$



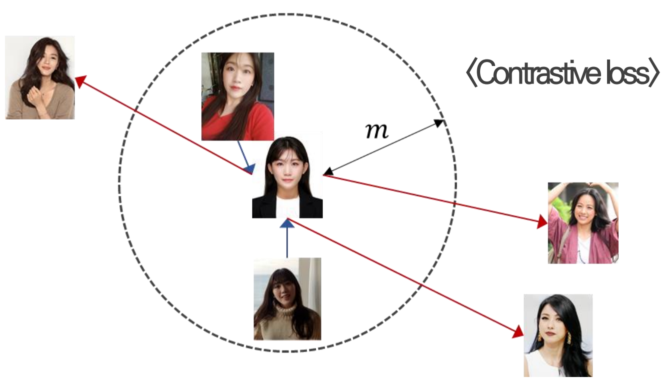
# 결론

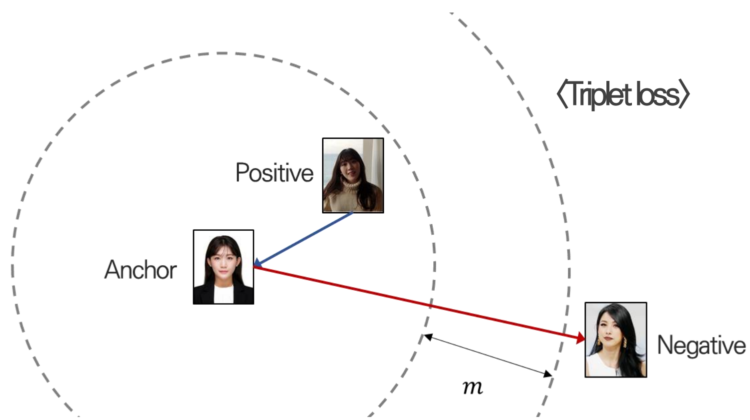
❖ 딥메트릭러닝(Deep metric learning)은 심층인공신경망(Deep neural network) 모델로 데이터 간 유사도가 잘 수치화된 잠재/임베딩 공간(Latent/embedding space)을 학습하는 방법론

❖ 최근 컴퓨터 비전 얼굴 인식 태스크에서 softmax classifier보다 우수한 성능을 보이며, 다양한 손실함수가 연구되고 있음

❖ 본 세미나에서는 다음 네 가지의 손실함수를 간략히 소개하였음

- Contrastive loss
- Triplet loss
- Center loss
- Addictive angular margin loss



# 참조


- https://en.wikipedia.org/wiki/Universal_approximation_theorem
- https://ko.wikipedia.org/wiki/%EA%B1%B0%EB%A6%AC
- https://tech.kakaoenterprise.com/63
- https://ko.wikipedia.org/wiki/K-%EC%B5%9C%EA%B7%BC%EC%A0%91_%EC%9D%B4%EC%9B%83_%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98
- https://ko.wikipedia.org/wiki/K-%ED%8F%89%EA%B7%A0_%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98
- https://hav4ik.github.io/articles/deep-metric-learning-survey
- https://medium.com/slyce-engineering/digging-deeper-into-metric-learning-loss-functions-29d89edfe200

- Chopra, S., Hadsell, R., & LeCun, Y. (2005, June). Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 539-546). IEEE.
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).
- Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016, October). A discriminative feature learning approach for deep face recognition. In European conference on computer vision (pp. 499-515). Springer, Cham.
- Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4690-4699).

Data Mining
Quality Analytics

# 감사합니다

2021. 11. 12.

Korea Univ, Data Mining & Quality Analytics Lab.

**이민정**
leemj2520@korea.ac.kr