

# Supervised learning exploiting clustering information

서울대학교 통계학과 김동하

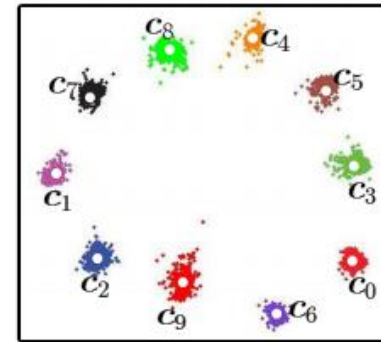
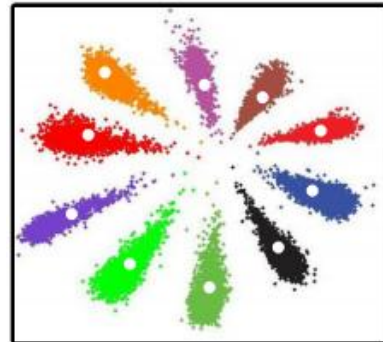
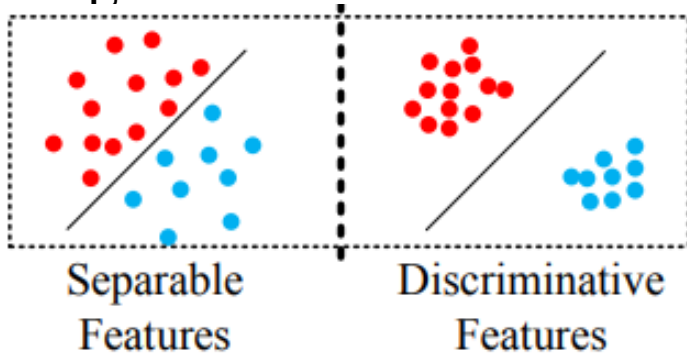
## Identification and verification problem

- Identification

- 주어진 입력 자료가 어떤 class를 갖는지 판별하는 문제. (=Classification)
- 모든 자료의 class의 가지 수를 미리 알고 있다고 가정. (=Closed set problem)
- 모형을 통해 feature vector space 상에서 자료들을 class별로 모으고, 이를 잘 나누는 초평면을 만드는 것이 중요.
- ✓ Intra-class variation과 inter-class variation의 관계는 중요하지 않음.

- Verification

- 주어진 두 입력 자료가 같은 class를 갖는지를 판별하는 문제
- 시험 자료에 unknown class가 있을 수도 있다고 가정. (=Open set problem)
- 모형을 통하여 feature vector space 상에서 같은 class를 갖는 자료들끼리 모아주고, 다른 class를 갖는 자료들끼리는 멀리 떨어지도록 하는 것이 중요.
- ✓ 즉, Intra-class variation이 inter-class variation에 비해 작도록 학습하는 것이 중요함.



## Face verification problem

---

- SAIT 측에서 궁극적으로 풀고자 하는 문제는 face verification problem
- Face verification은 feature space 상에서의 자료들 사이의 거리를 기준으로 하여 같은 class인지 아닌지를 판별하는 방식으로 이루어짐.
- 먼저 face identification problem으로 인식하여 cross-entropy loss function을 이용하여 feature space를 학습하고, 이를 그대로 사용하거나 또는 이를 초기값으로 시작하여 triplet loss 등을 이용하여 feature space를 새롭게 학습한 후에 verification problem을 풀고 있는 상황.

## Face verification problem

- 같은 사람의 이미지라도 여러 가지 변형이 있을 수 있음.  
(예 : 표정, 안경의 유무, 제스처 등)



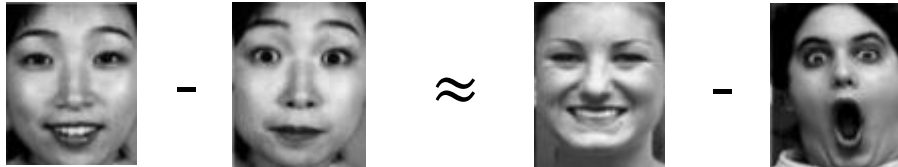
- 이러한 다양한 종류의 사진은 얼굴 인식을 어렵게 하는 주된 요인이 됨.
- Intra-class variation이 커지기 때문..
- 모형 학습을 하고 나면 한 사람의 다양한 이미지들이 feature space 상에서 안경의 유무, 표정 등으로 군집을 형성하는 것을 확인할 수 있음.
- 이러한 군집 정보를 이용하여 얼굴 인식을 한다면 더 좋은 성능을 나타낼 것으로 기대.
- Clustering based semi-supervised learning  
→ Supervised learning exploiting clustering information

### Goal

- 군집 정보를 이용하여 인식률을 높이는 방법론 개발
- 1차적으로 classification problem을 생각하고 있음. (Intra-class variation이 Inter-class variation보다 훨씬 작은 관계를 만족하여 verification problem으로의 확장이 용이하도록 하는 classification 방법론 연구 중.)

- Class 별로 군집의 분포가 비슷하도록 학습하는 것이 목표.

Ex:



- 문제점
  - SAIT 측에서는 feature space를 unit sphere가 되기를 원함. (triplet loss를 이용하여 학습을 하기 때문인 것으로 알고 있음.)
  - Unit sphere 상에서 clustering 기법을 어떻게 사용할 것인지..