

SVM + Unsupervised Learning

An Introduction to Statistical Learning

황성원

목차

최적화 방법론

1. 분석적 풀이
2. 내리막 경사법
3. 라그랑제 승수

SVM

1. 최대 마진 분류기
2. 서포트 벡터 분류기
3. 서포트 벡터 머신

Unsupervised

1. PCA
2. 클러스터링

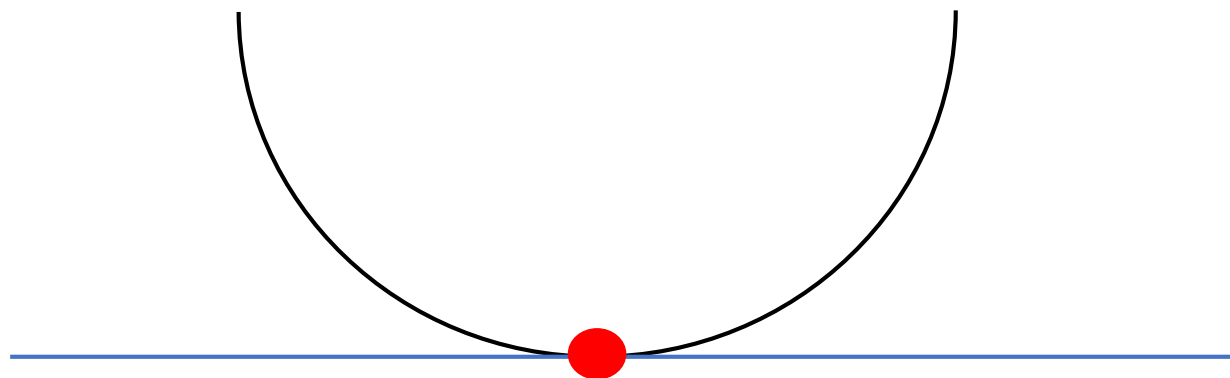
최적화 방법론

최적화 방법론

1. 분석적 풀이

2. 내리막 경사법

3. 라그랑제 승수

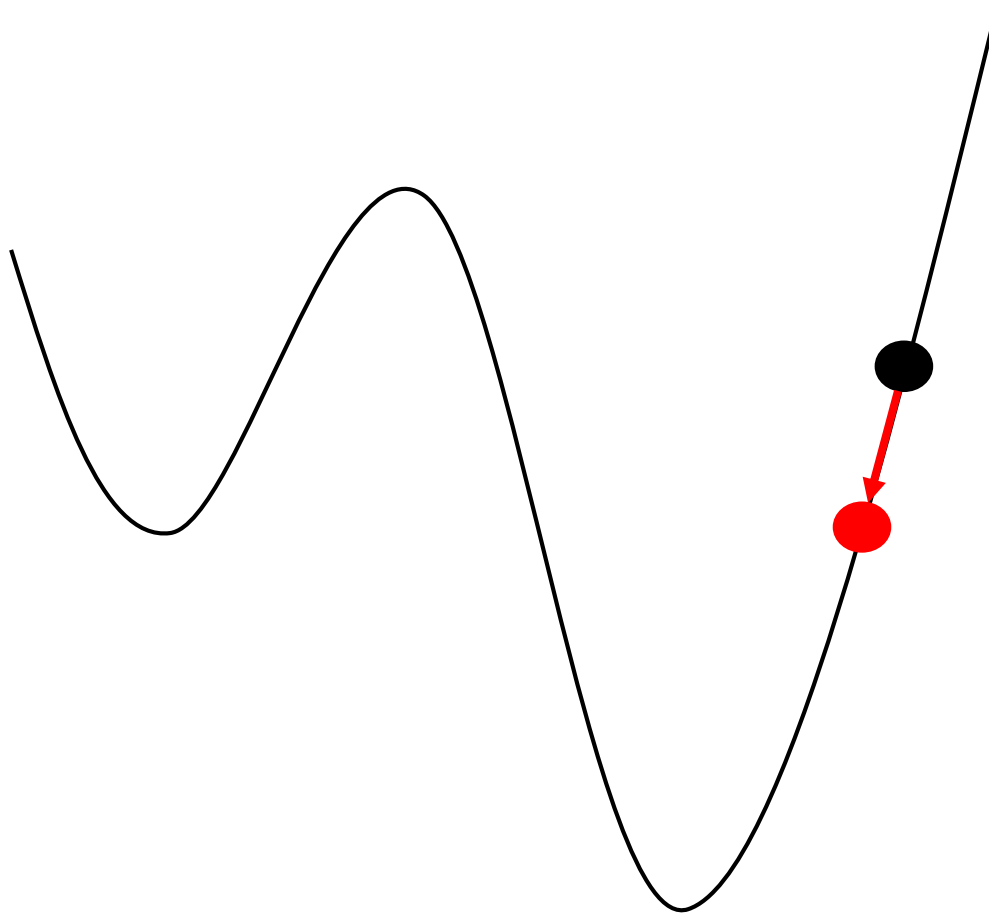


최적화 방법론

1. 분석적 풀이

2. 내리막 경사법

3. 라그랑제 승수

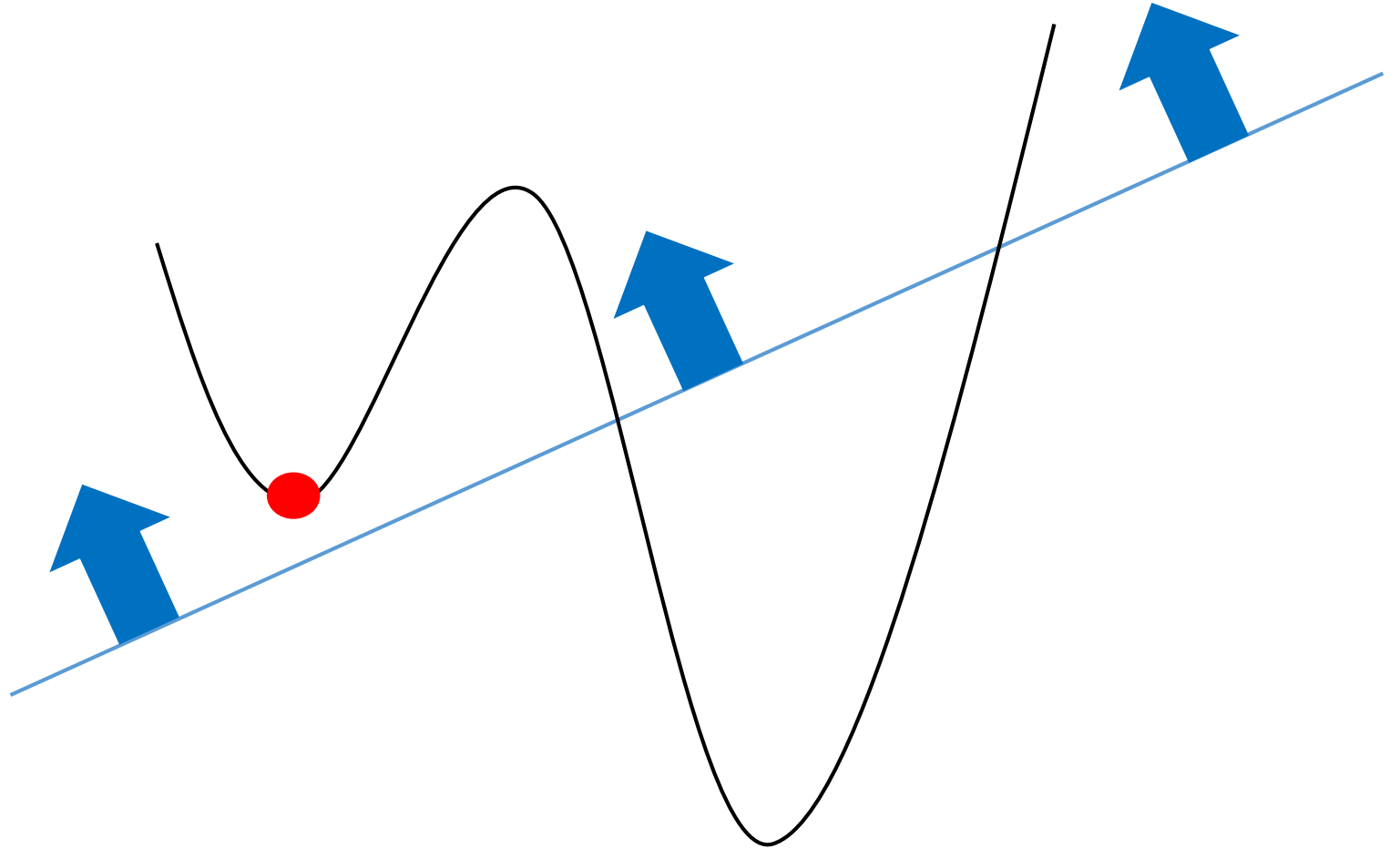


최적화 방법론

1. 분석적 풀이

2. 내리막 경사법

3. 라그랑제 승수



서포트 벡터 머신

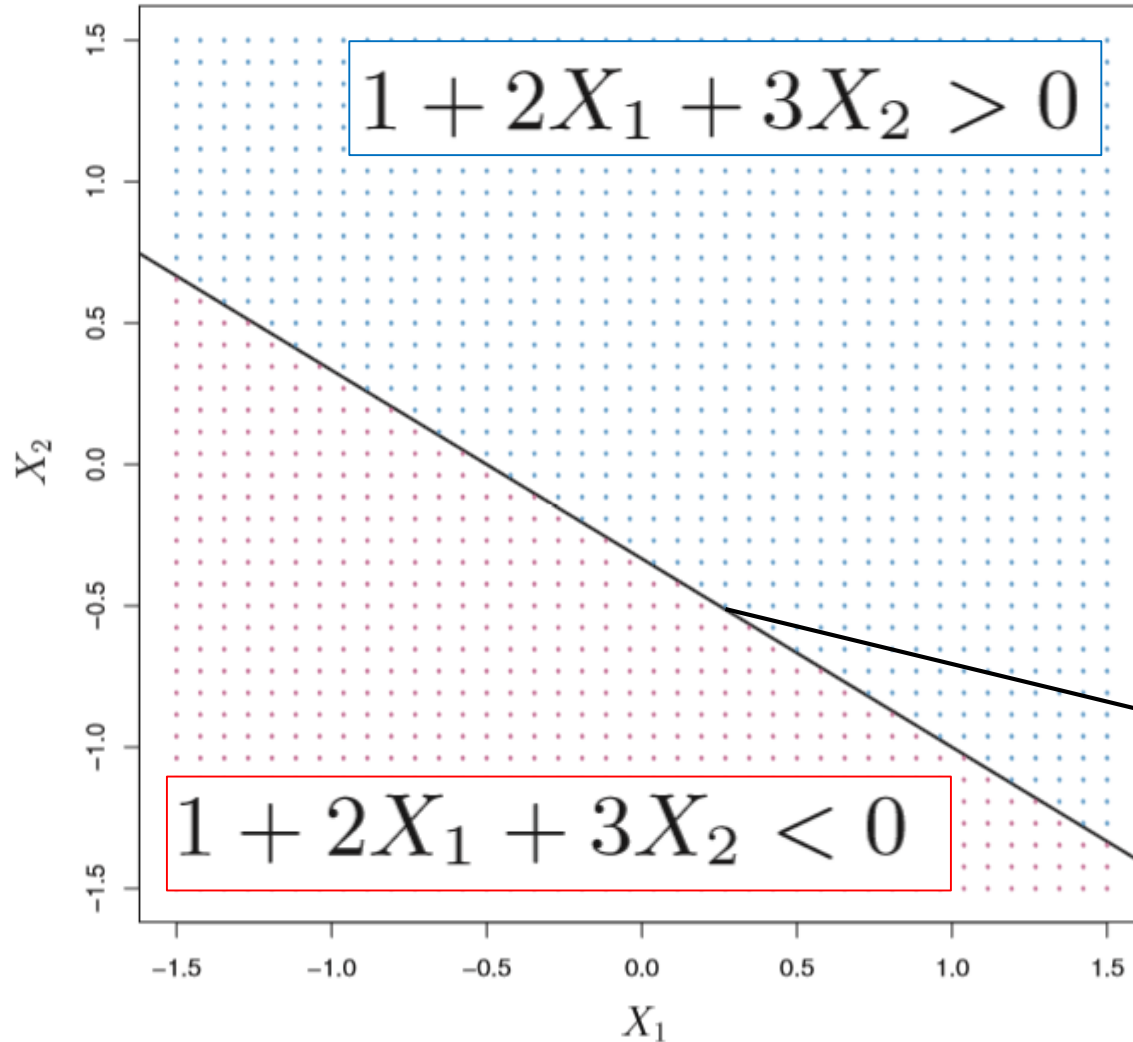
(Support Vector Machine)

1. 최대 마진 분류기
2. 서포트 벡터 분류기
3. 서포트 벡터 머신

1. 최대 마진 분류기

(Maximal Margin Classifier)

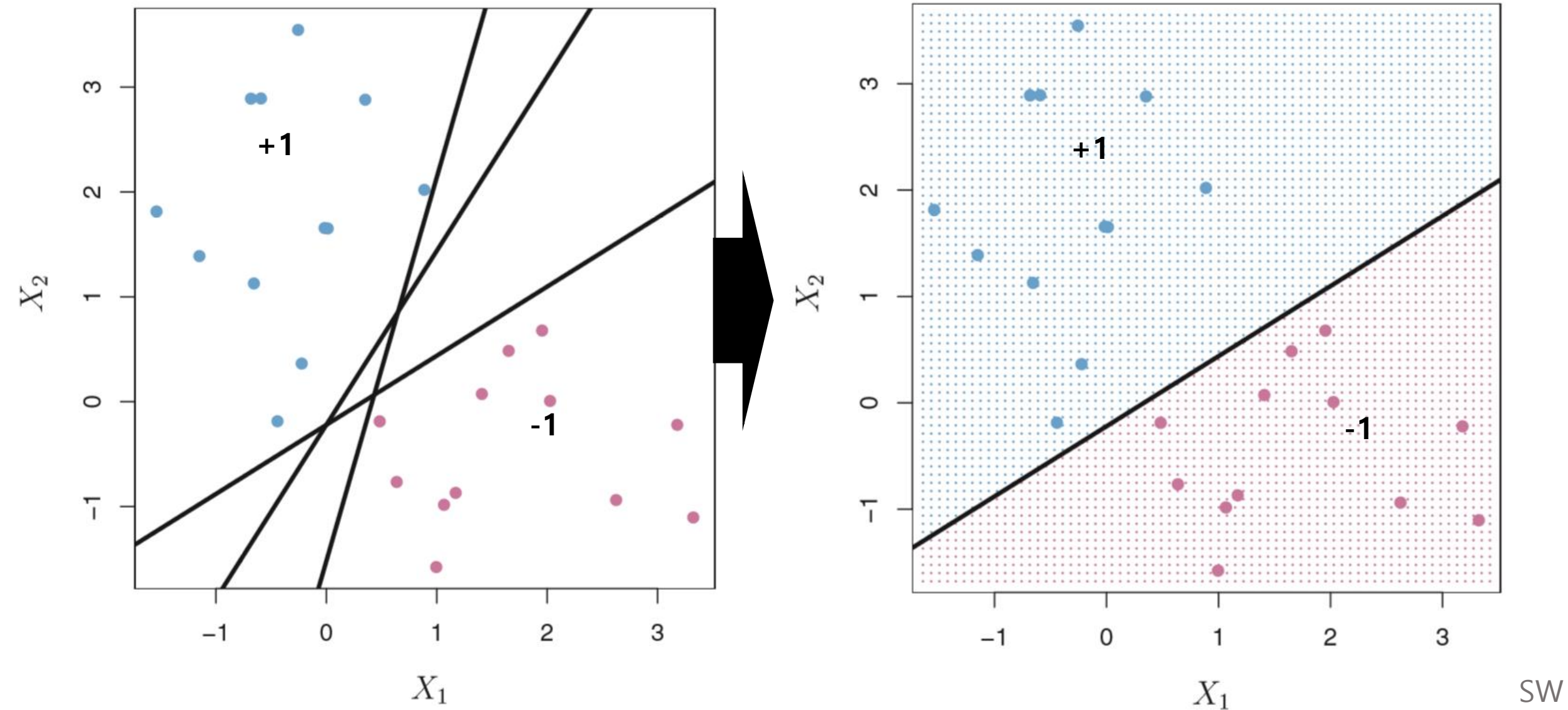
1. 최대 마진 분류기 – 초평면



$$1 + 2X_1 + 3X_2 = 0$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

분리 초평면으로 분류하기



분리 초평면 수식적 관점

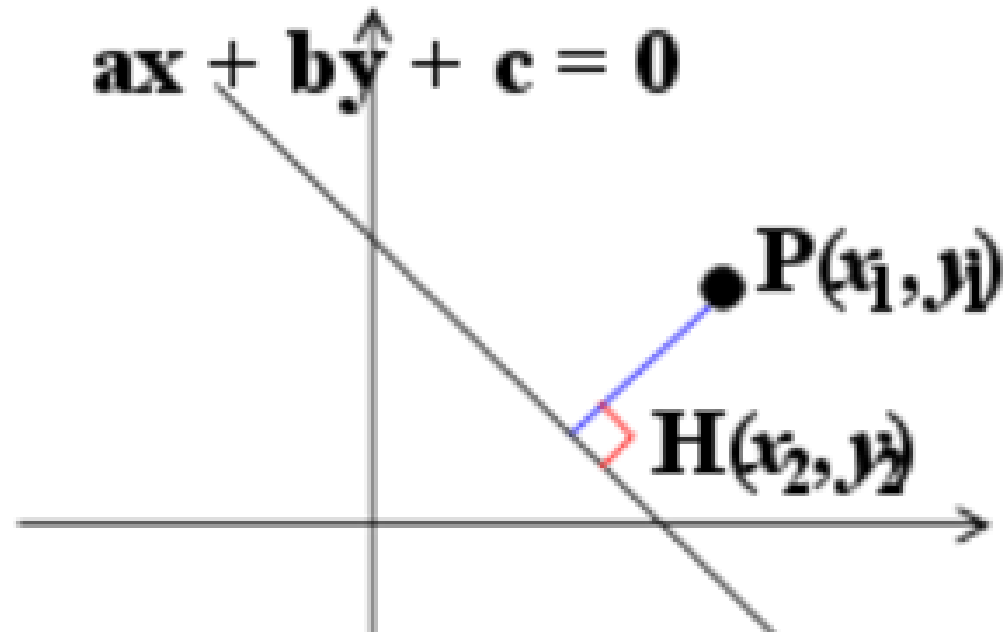
$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} > 0 \text{ if } y_i = 1$$

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} < 0 \text{ if } y_i = -1$$



$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) > 0$$


1. 최대 마진 분류기 – 사전 지식



점 (x_1, y_1) 과 직선 $ax + by + c = 0$ 사이의 거리 d

$$d = \frac{|ax_1 + by_1 + c|}{\sqrt{a^2 + b^2}}$$

1. 최대 마진 분류기 – 수식적 관점

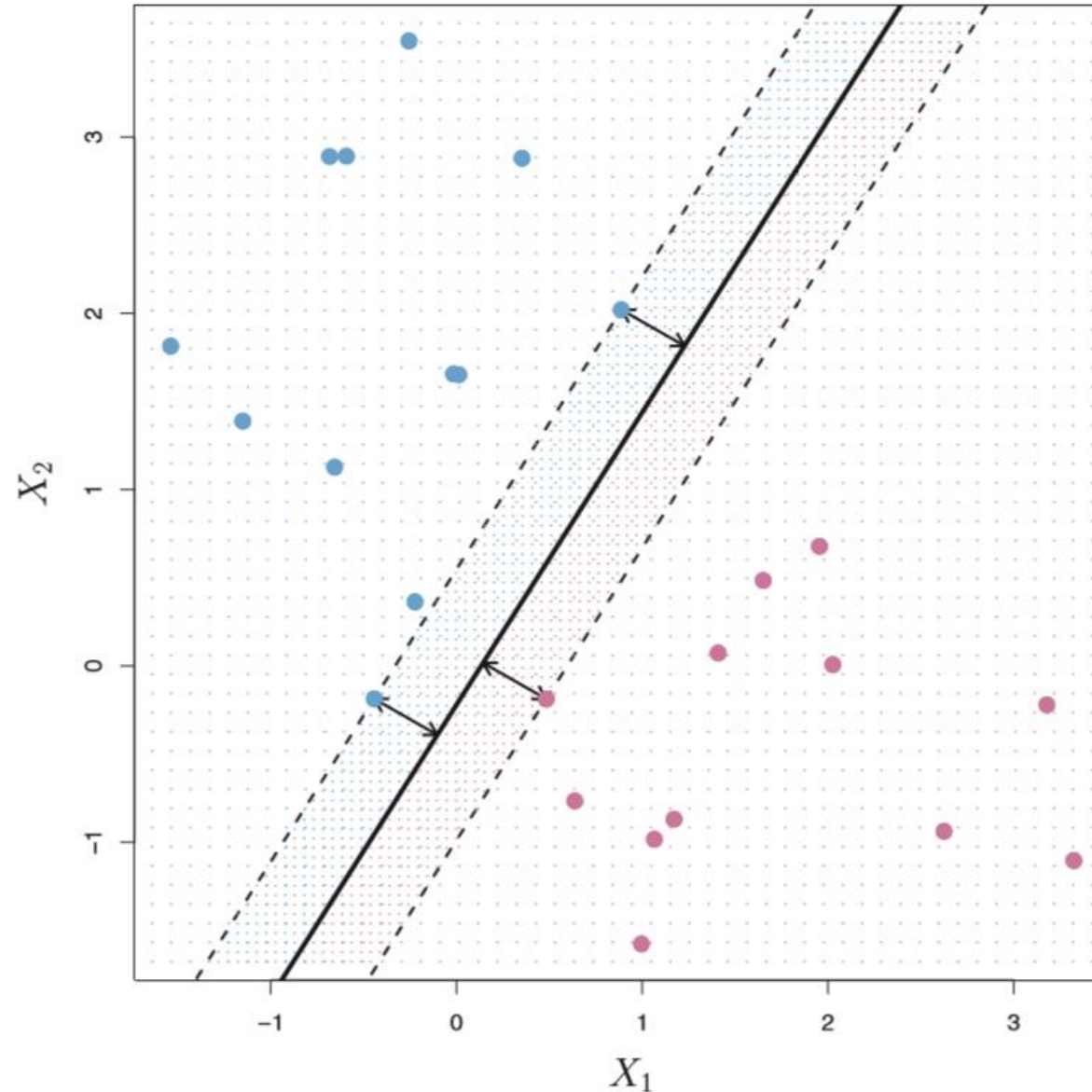

$$\text{maximize } M$$
$$\beta_0, \beta_1, \dots, \beta_p$$

$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1,$$

$$\underline{y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})} \geq M \quad \forall i = 1, \dots, n$$

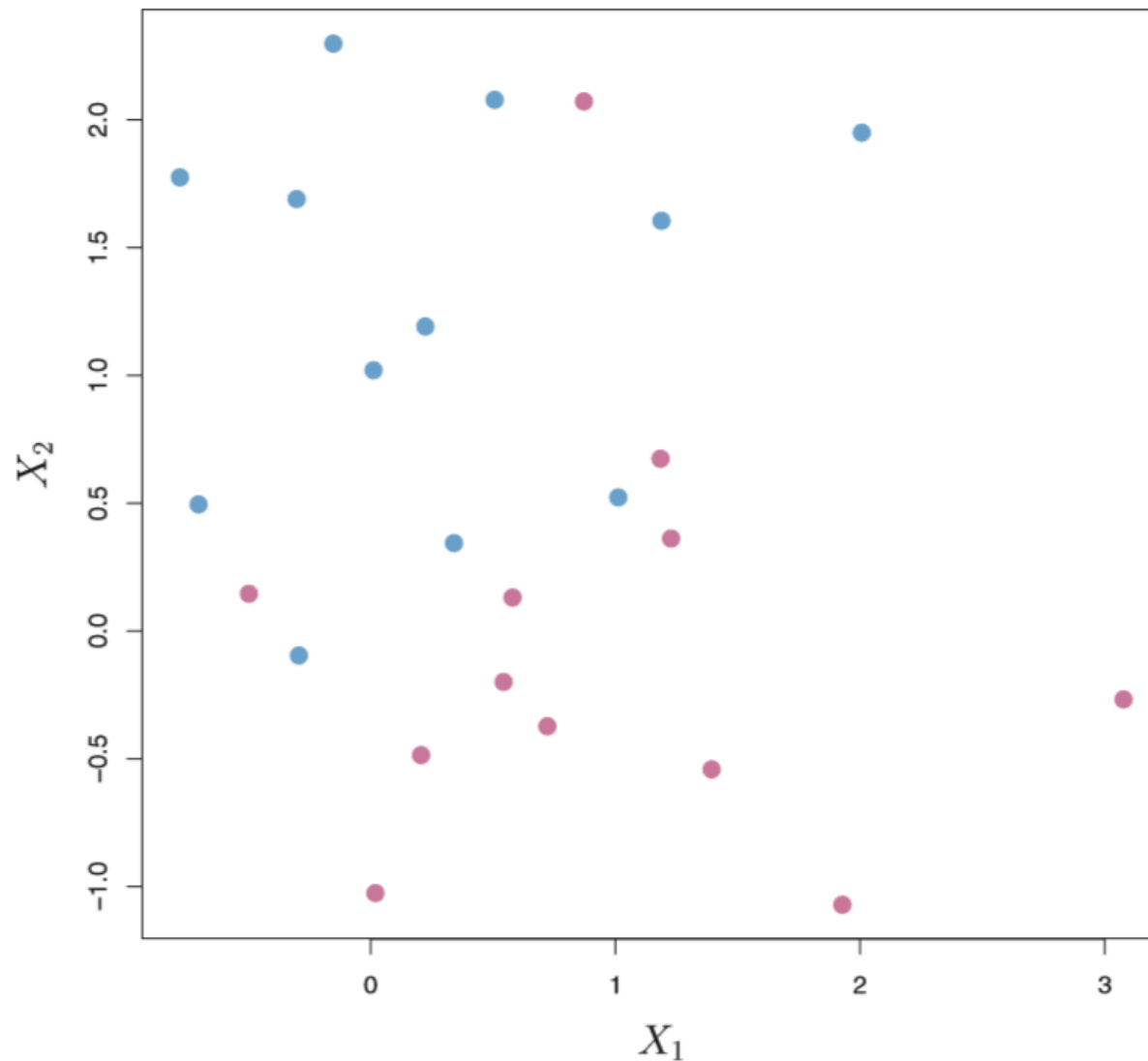
초평면과 훈련 데이터 사이의 거리!

1. 최대 마진 분류기 – 기하학적 관점

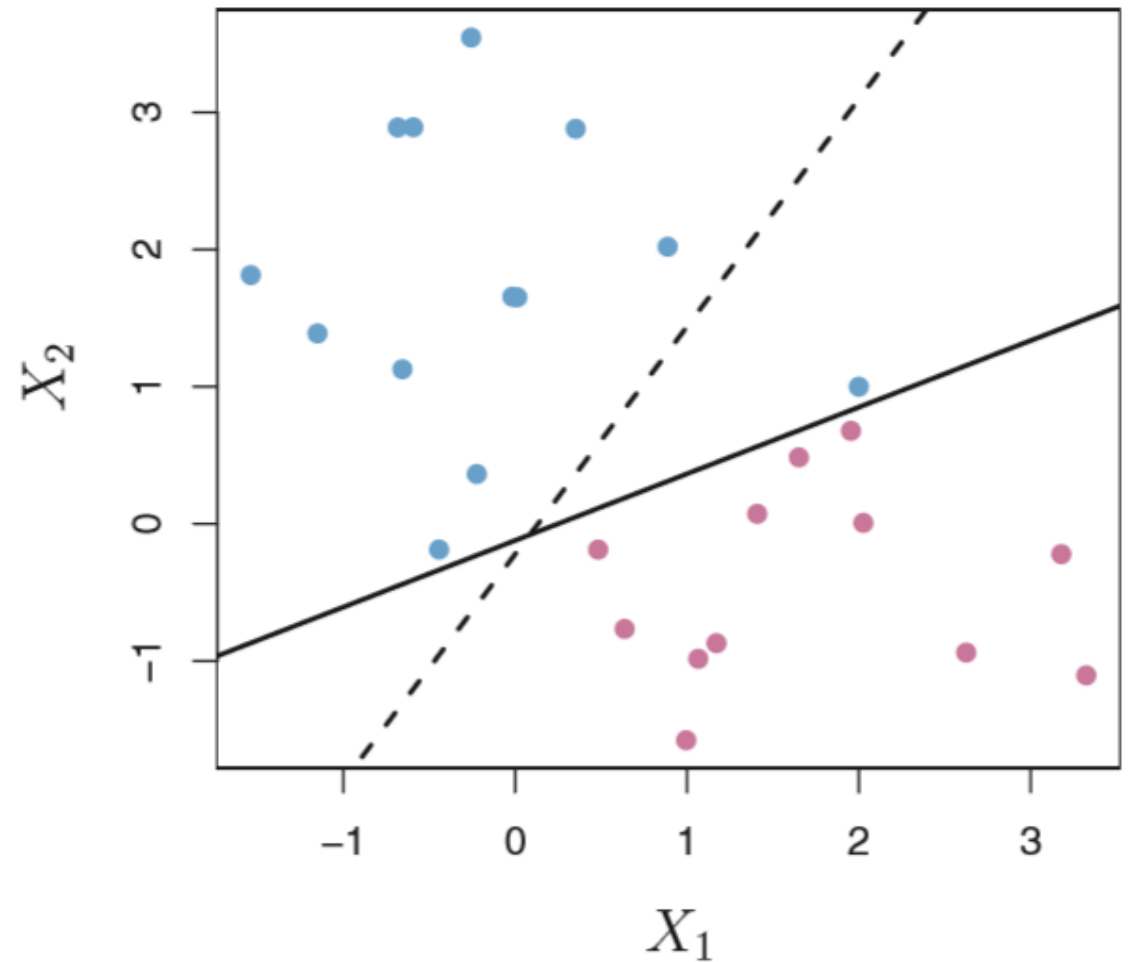
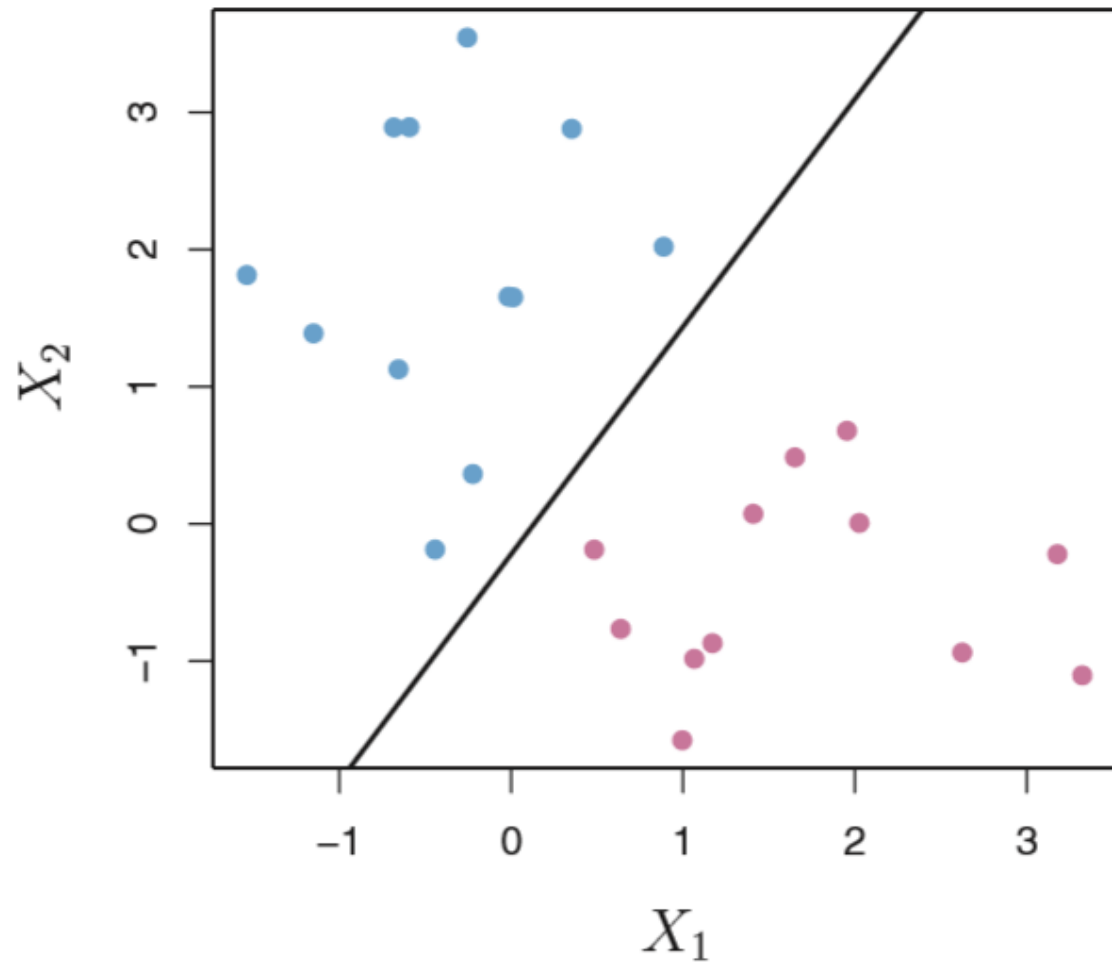


1. 최대 마진 분류기 – 한계 1. 완벽한 상태만

정확히 나뉘져 있지 않으면 못쓴다!



1. 최대 마진 분류기 – 한계 2. 너무 민감해



2. 서포트 벡터 분류기

(Support Vector Classifier)

2. 서포트 벡터 분류기 – 수식적 관점

$$\underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1,$$

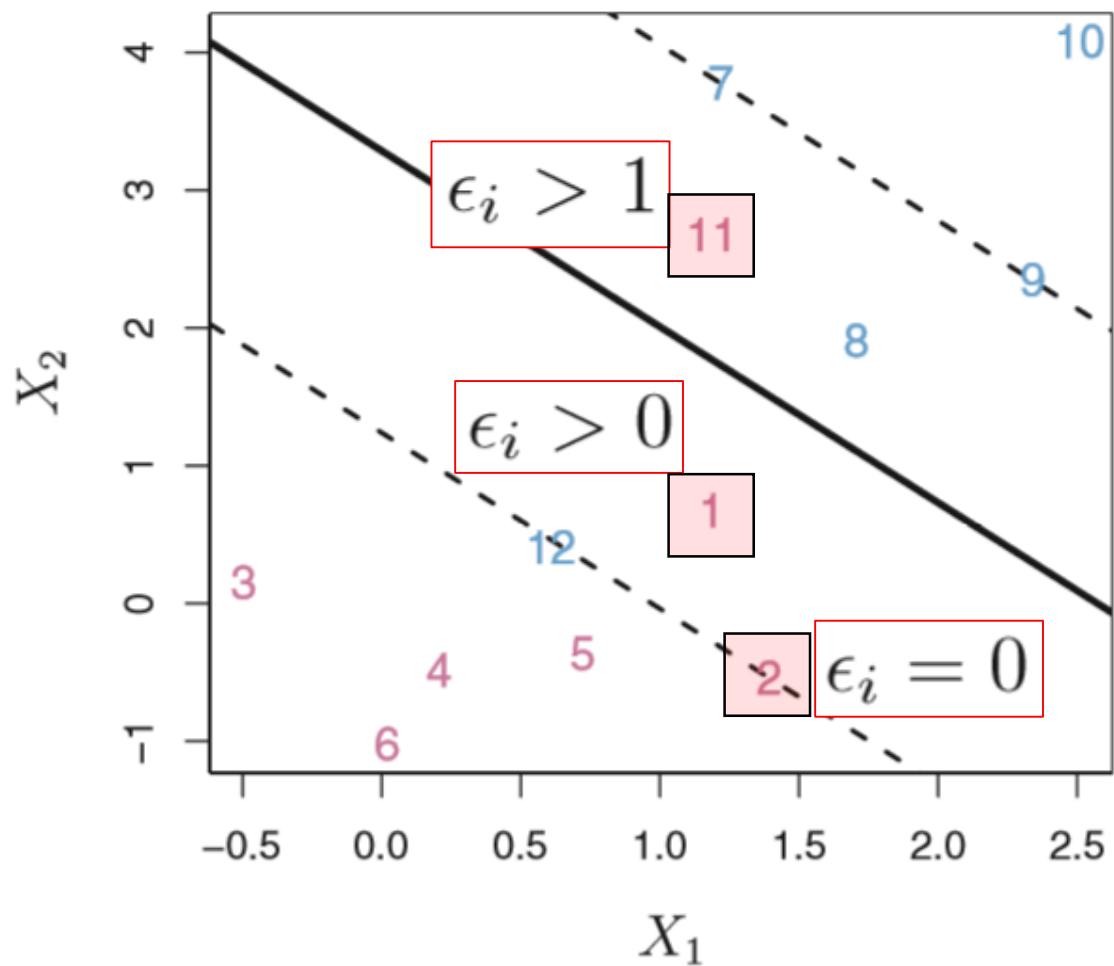
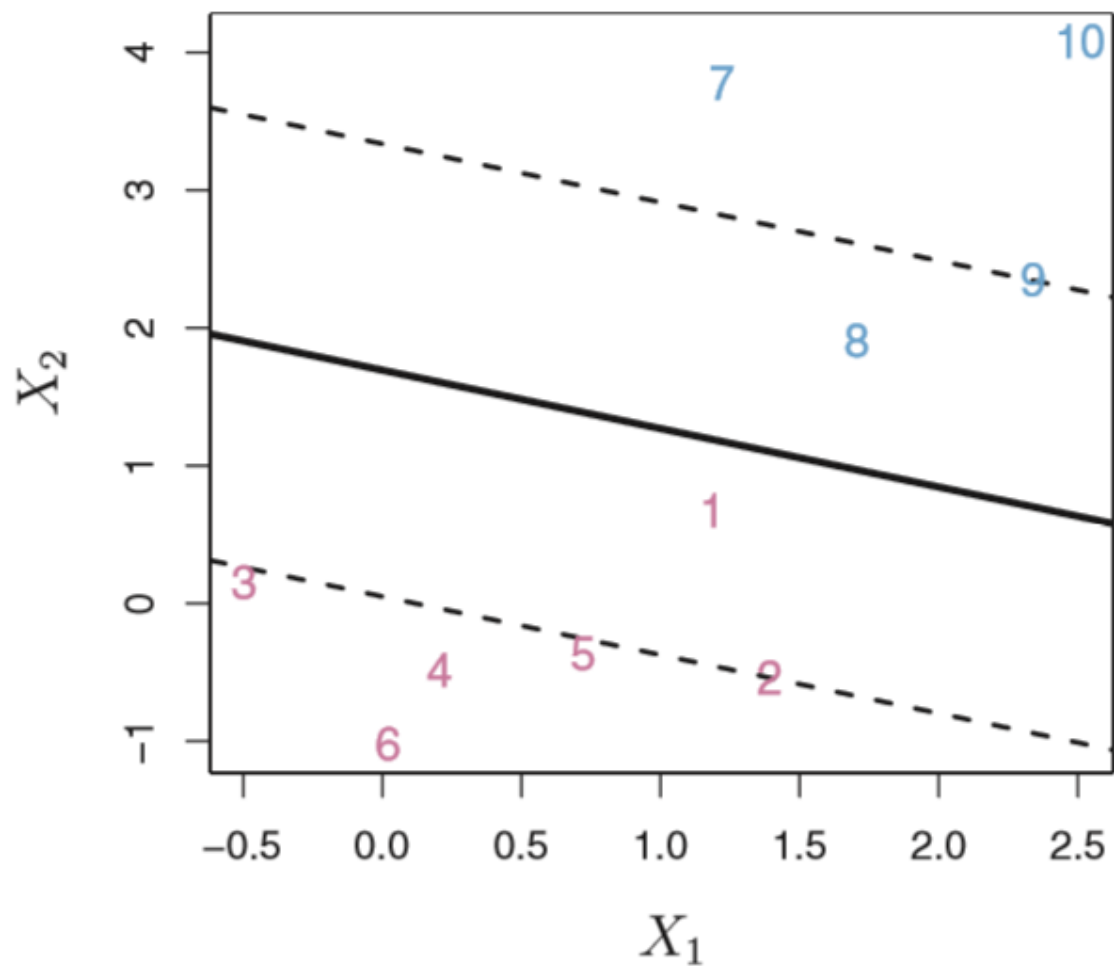
$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i),$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C,$$

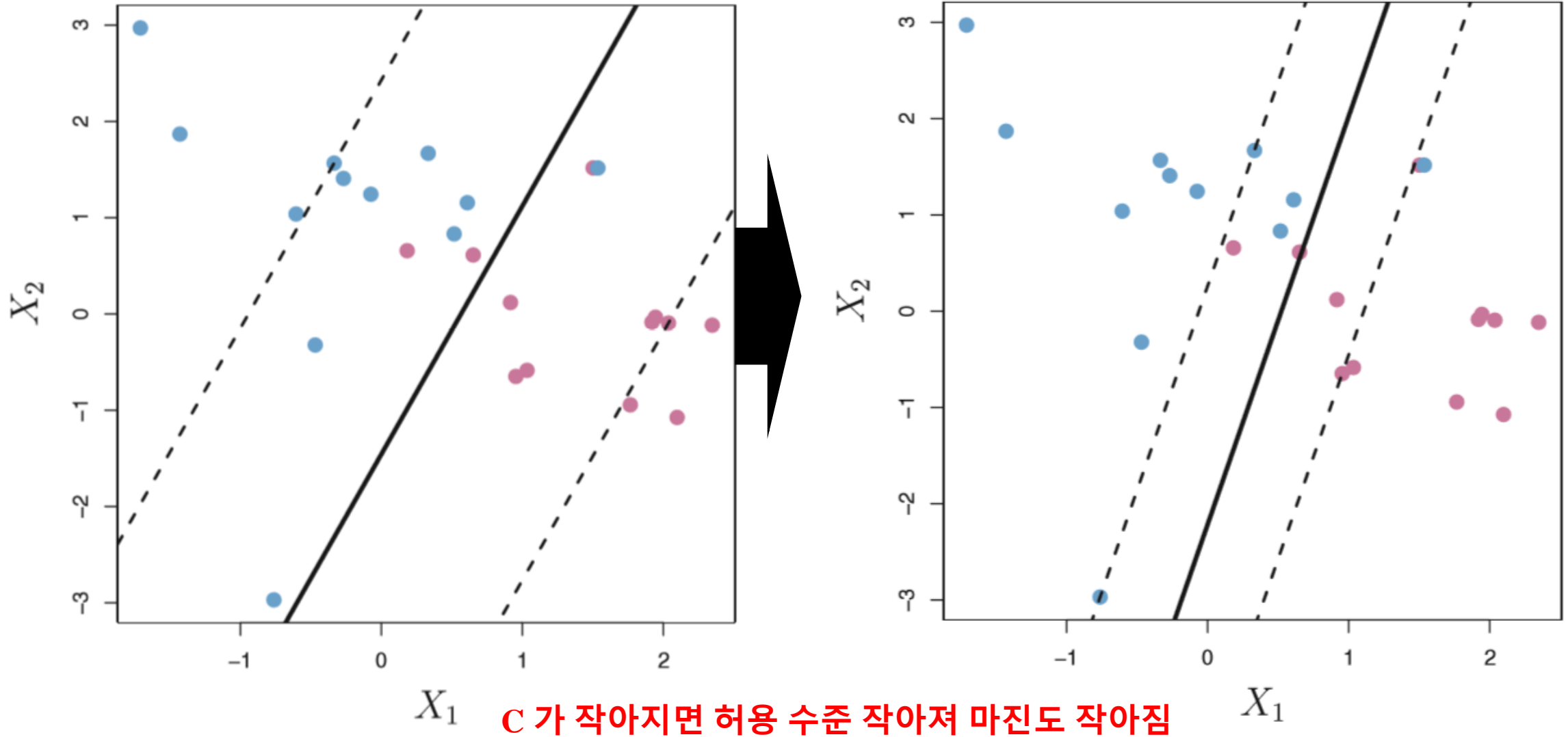
조율 파라미터: 허용될 위반의 수와 그 정도 결정

슬랙(느슨한) 변수: i번째 관측치의 상대적 위치 정보

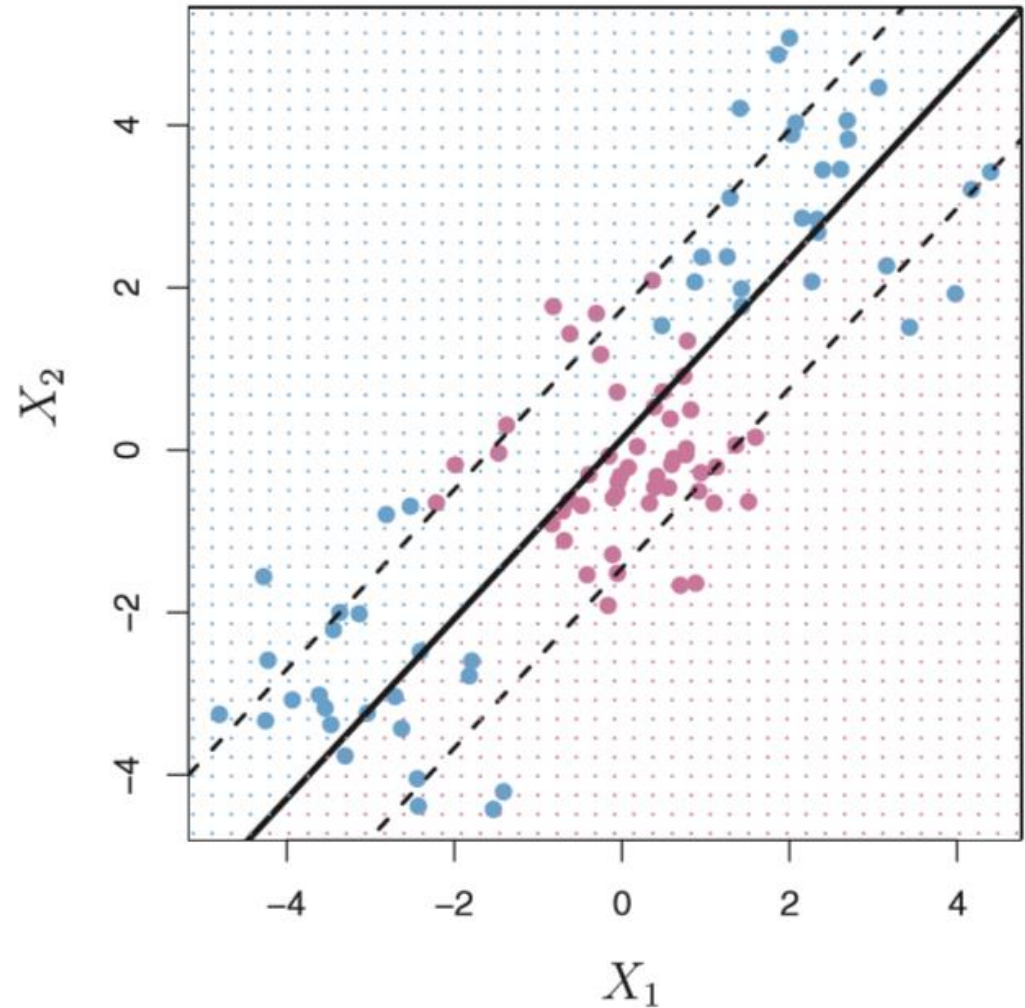
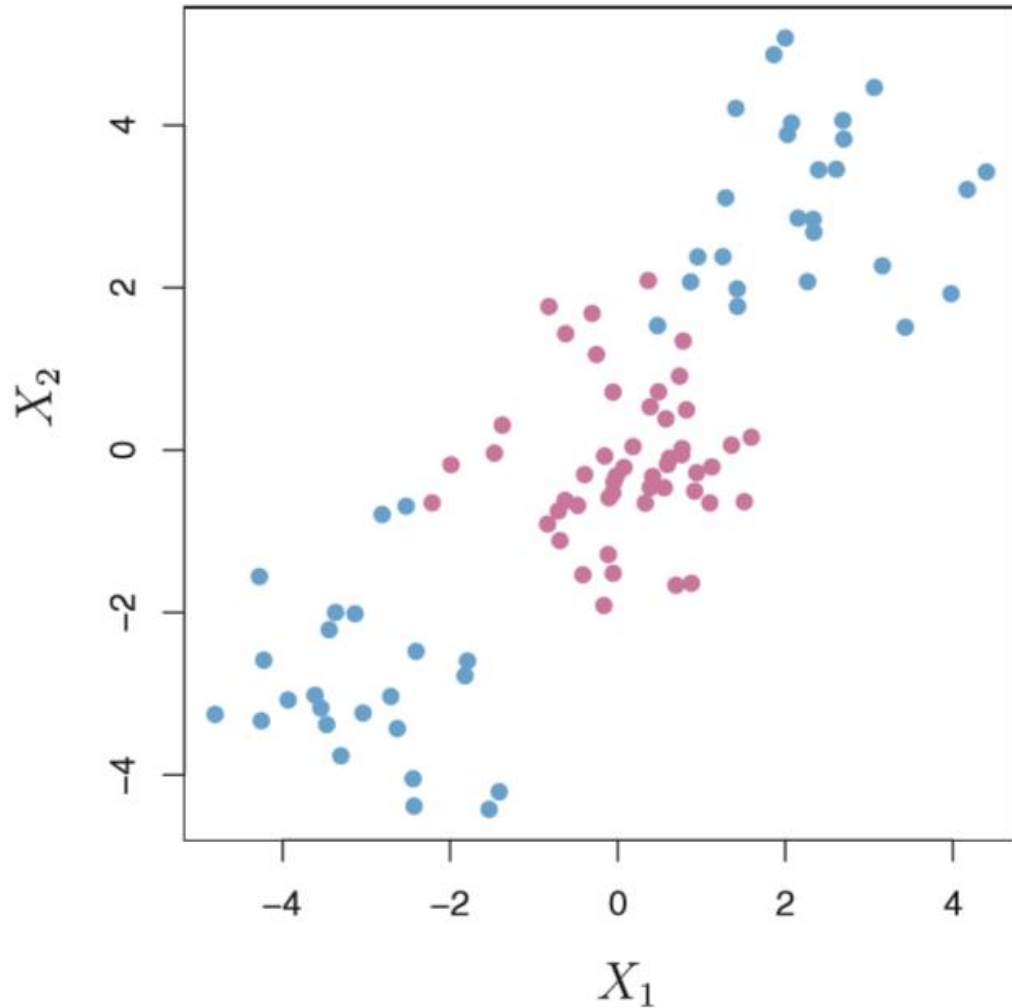
2. 서포트 벡터 분류기



2. 서포트 벡터 분류기 – 조율 파라미터 C



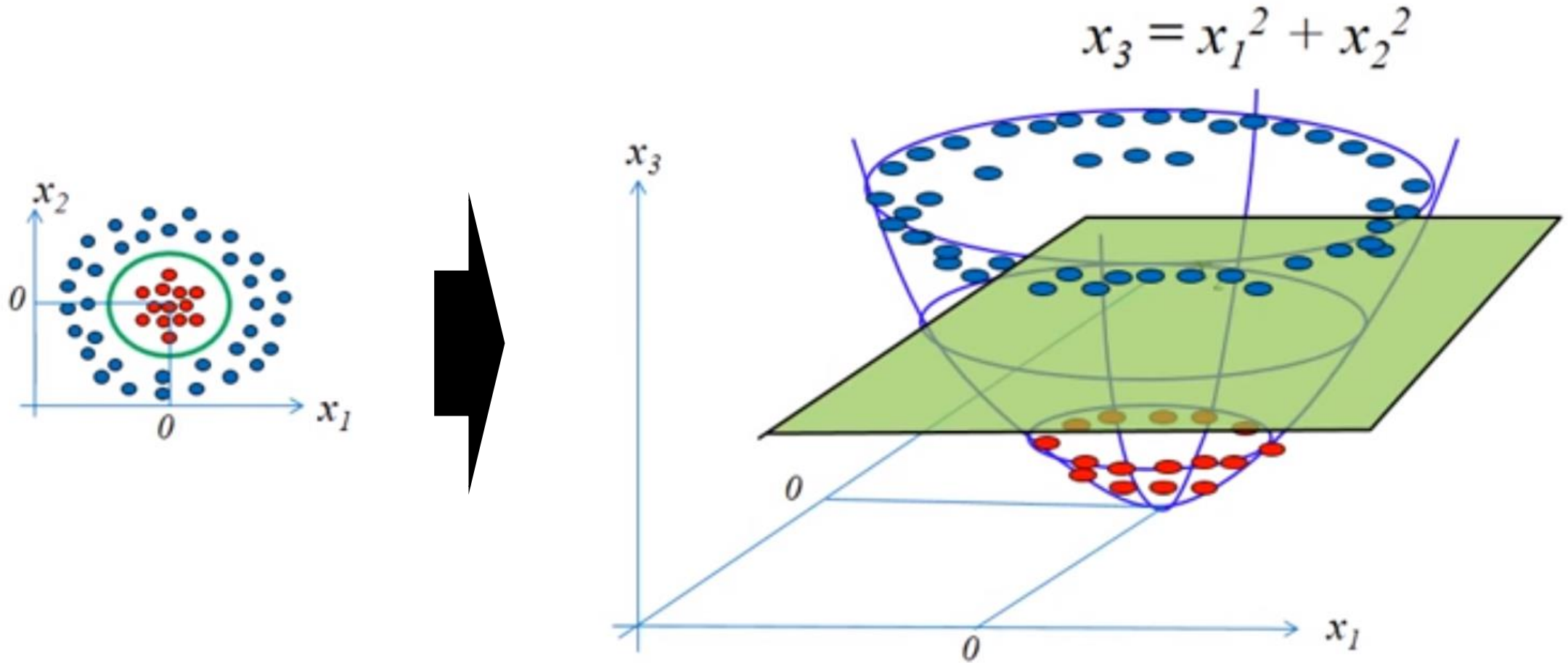
2. 서포트 벡터 분류기 – 한계: 비선형?



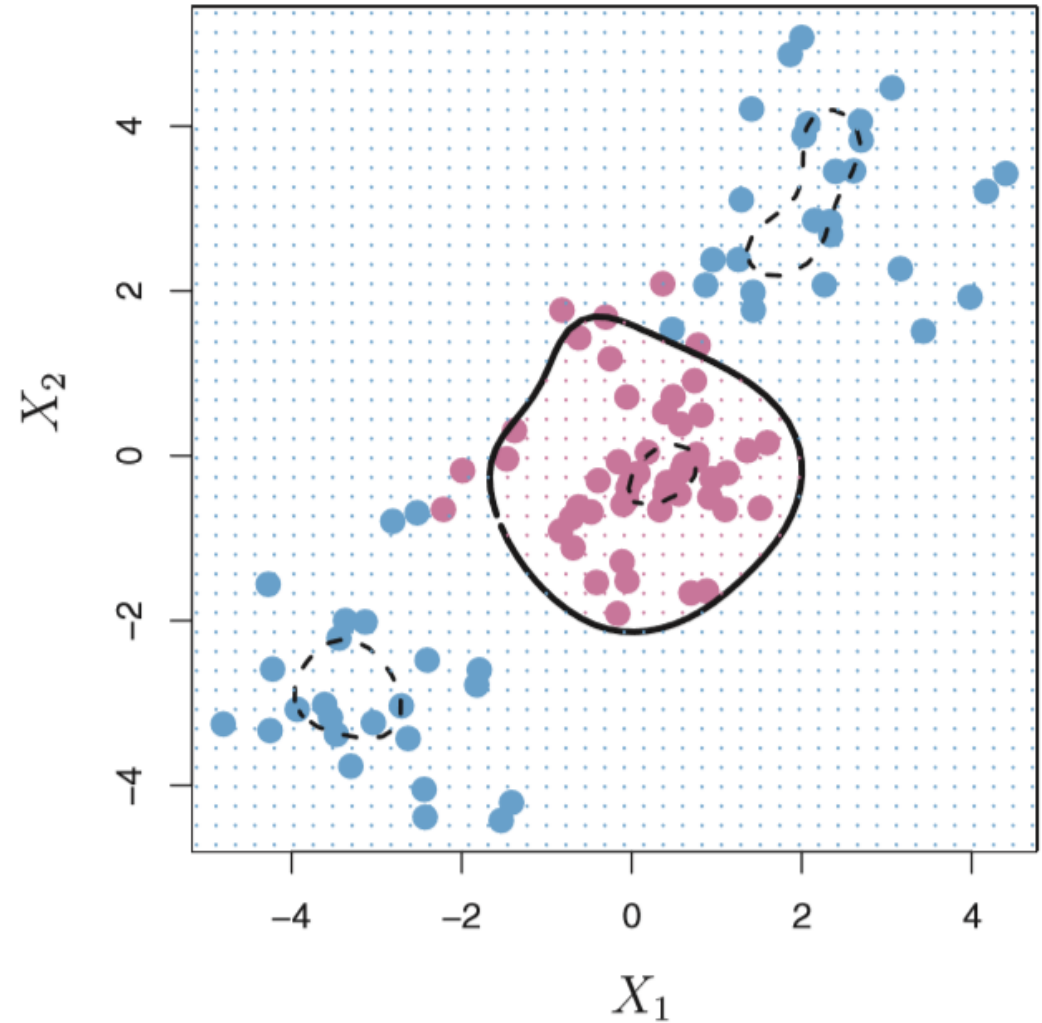
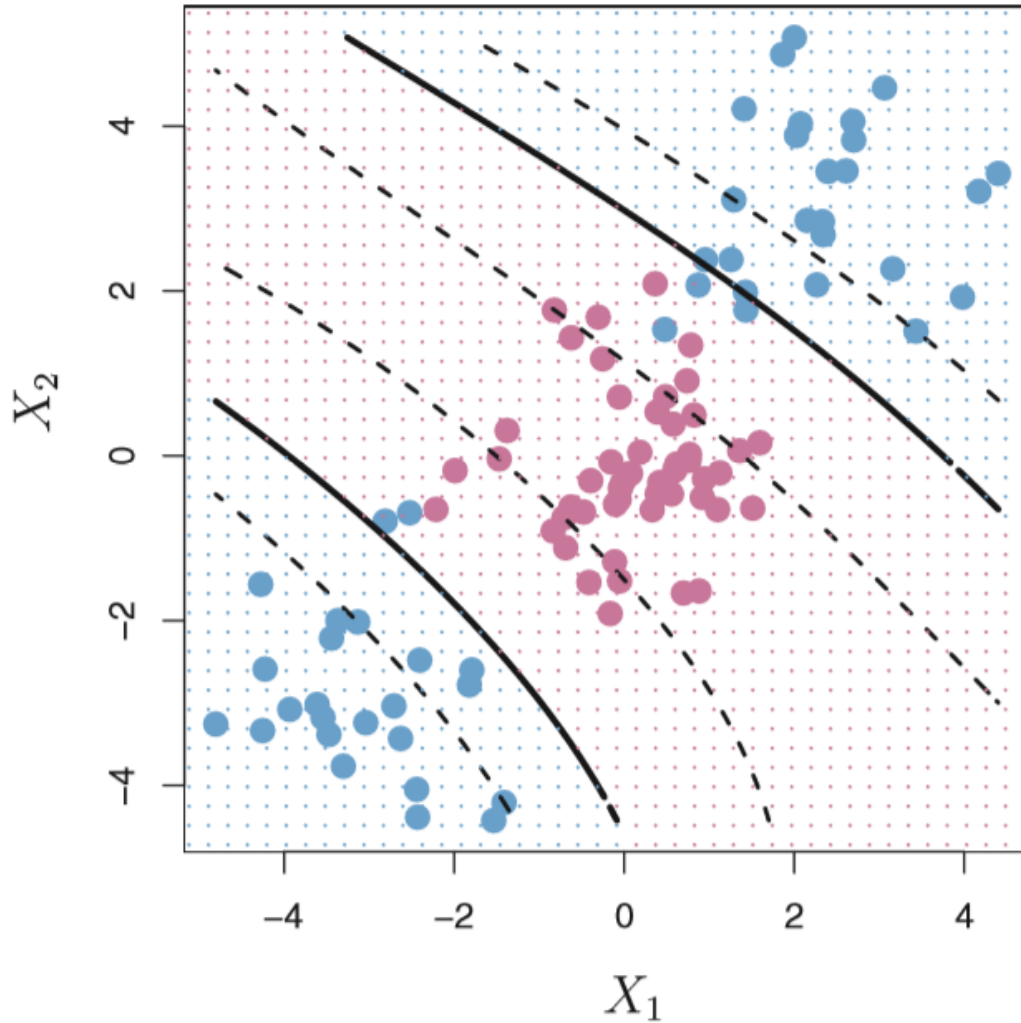
3. 서포트 벡터 머신

(Support Vector Machine)

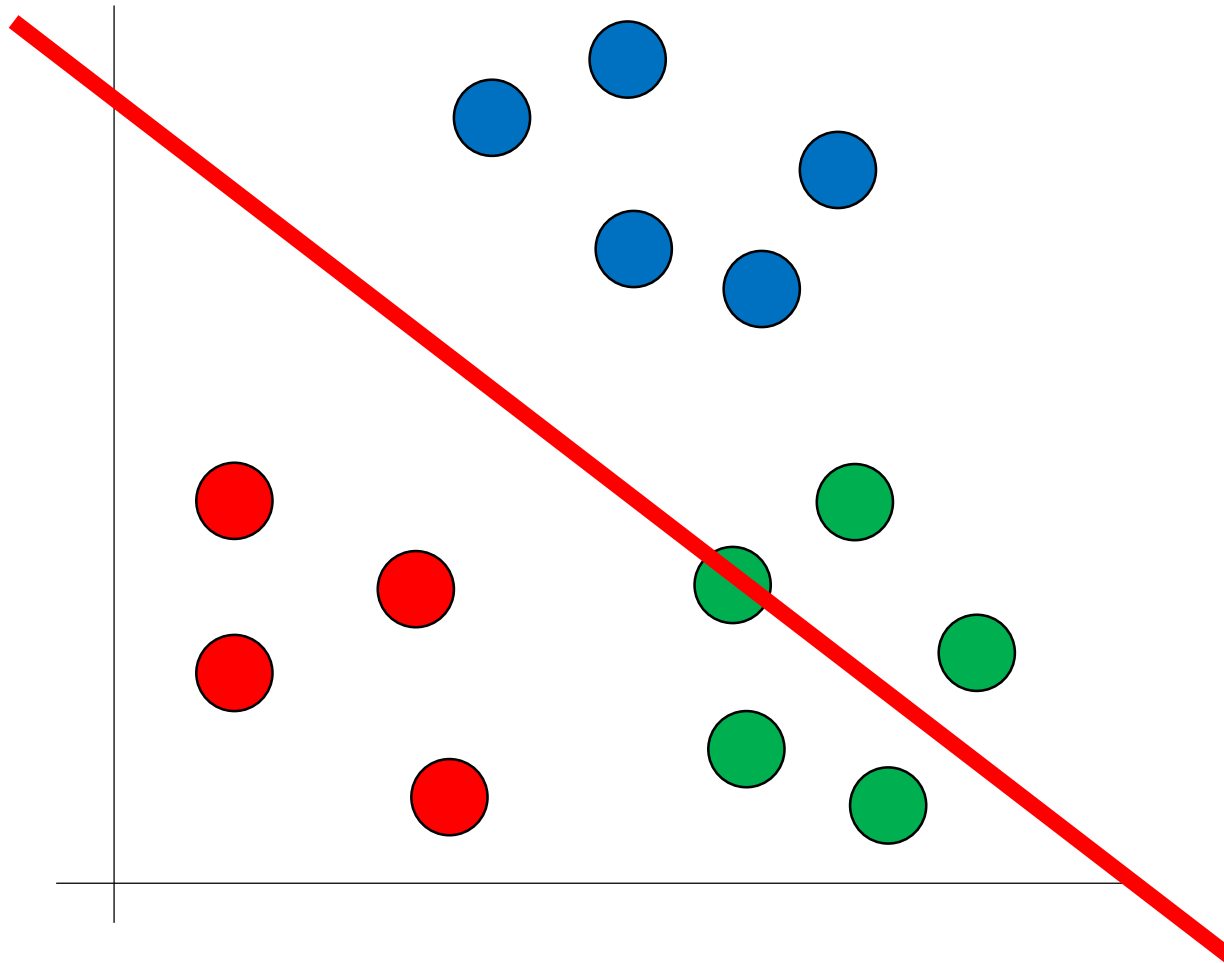
3. 서포트 벡터 머신 – Kernel \rightarrow 차원 변환



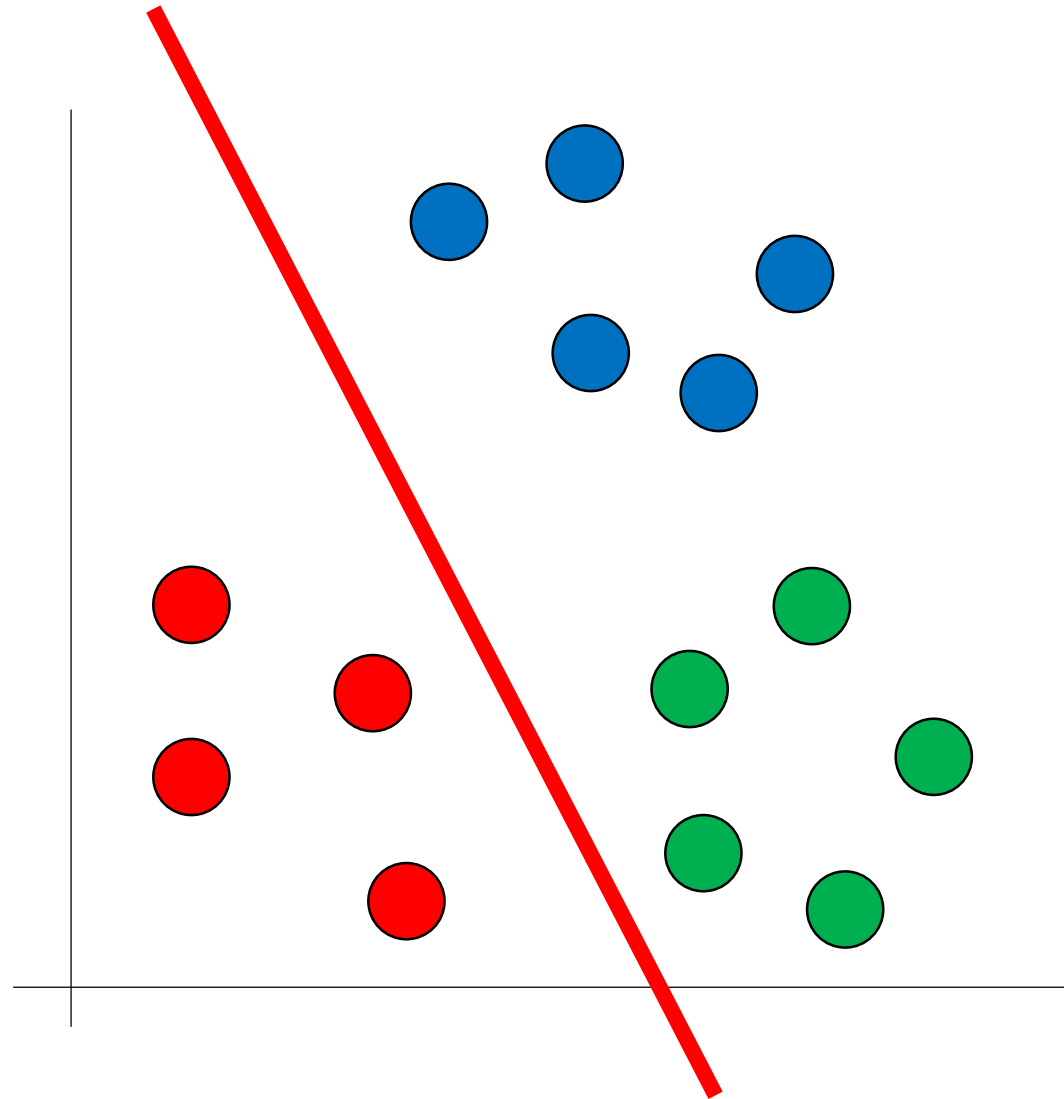
3. 서포트 벡터 머신 – 여러 가지 커널



클래스 2개 이상인 경우 – 1. 일대일 분류



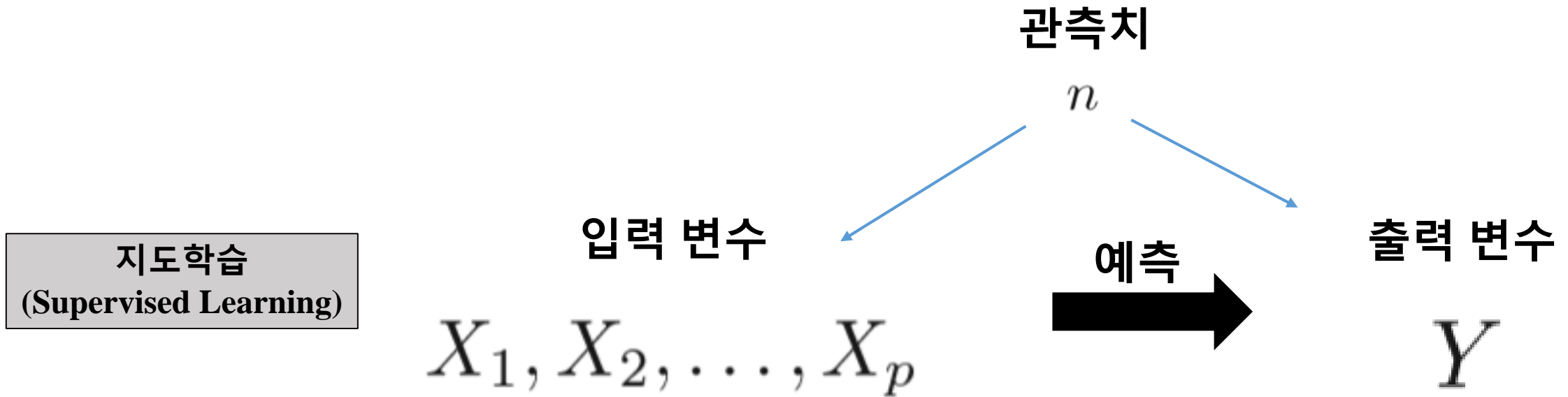
클래스 2개 이상인 경우 – 1. 일대전부 분류



비지도 학습

(Unsupervised Learning)

비지도 학습? – 입력변수도 하나의 데이터!



비지도 학습만으로 의미가 있다!



Pre-processing (전처리)

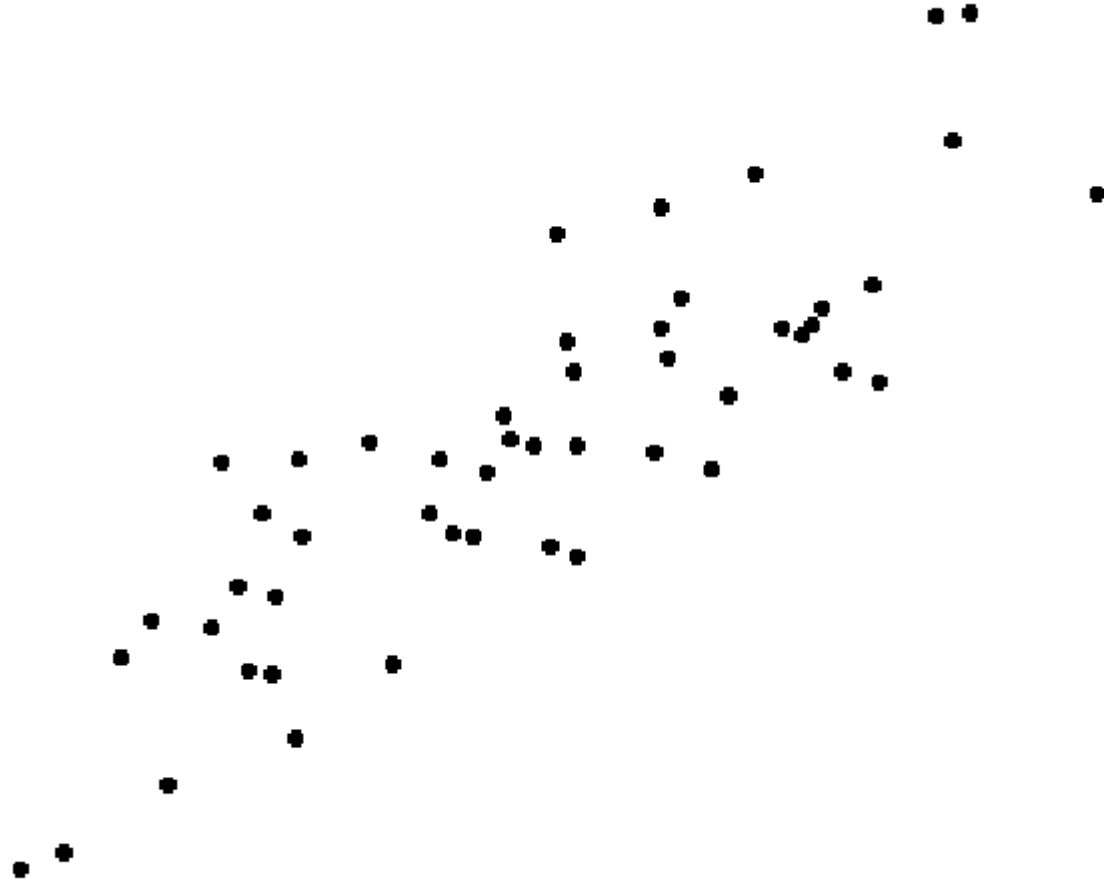
: PCA 주성분분석

서브 그룹 발견 및 특징 분석

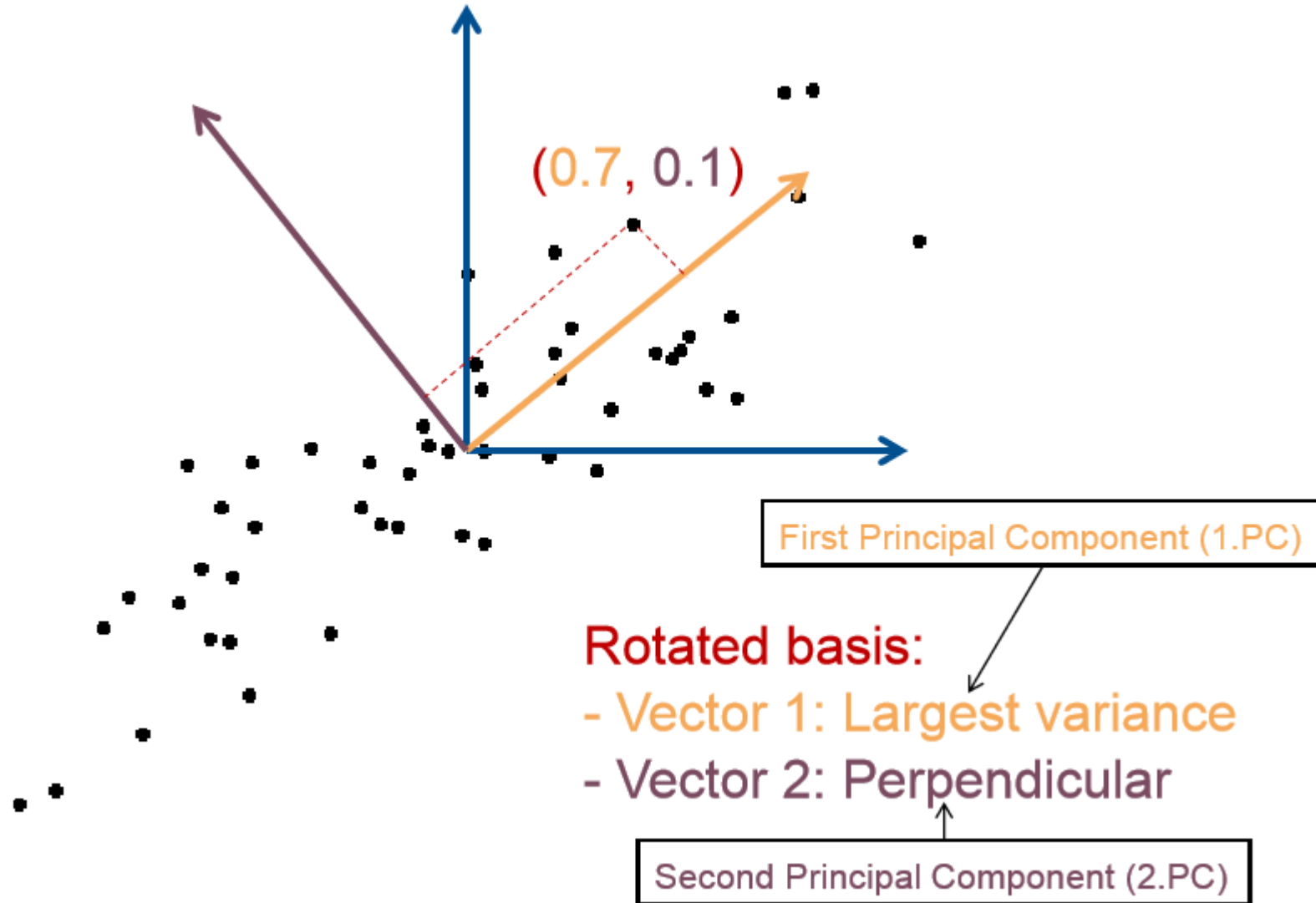
: Clustering (클러스터링)

→ 예) 온라인 쇼핑 사이트, 유방암 환자 유전자 발현 SW

1. PCA 주성분분석



1. PCA 주성분분석

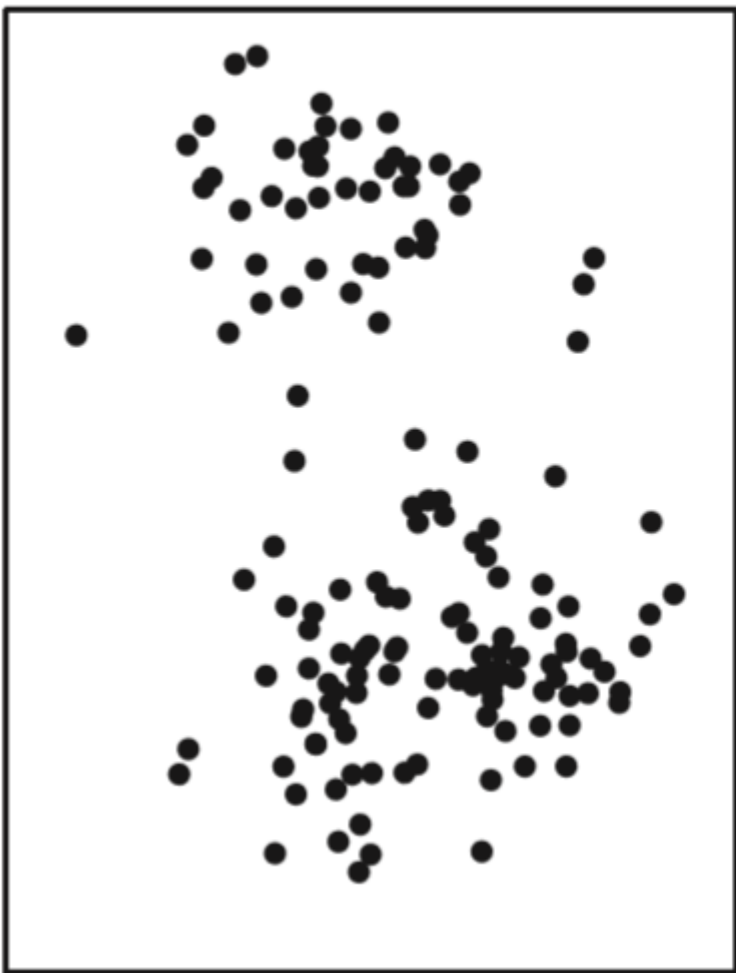


1. PCA 주성분분석

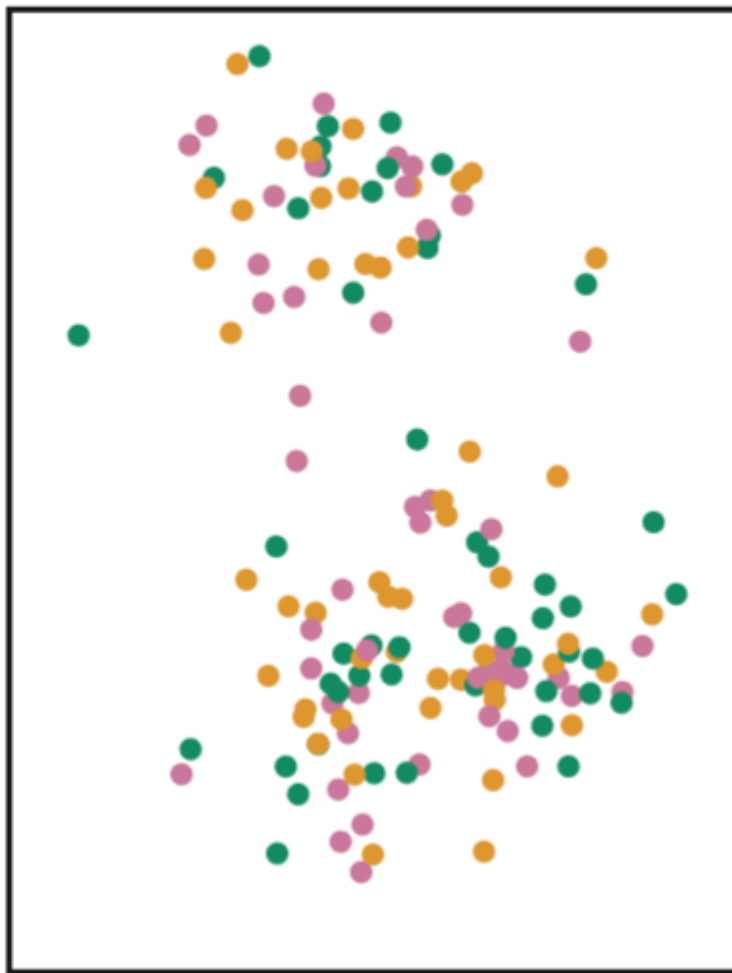
- 1) 데이터 해석력이 좋아질 수 있다.
- 2) 데이터 차원이 축소될 수 있다.

2. 클러스터링 – K-평균

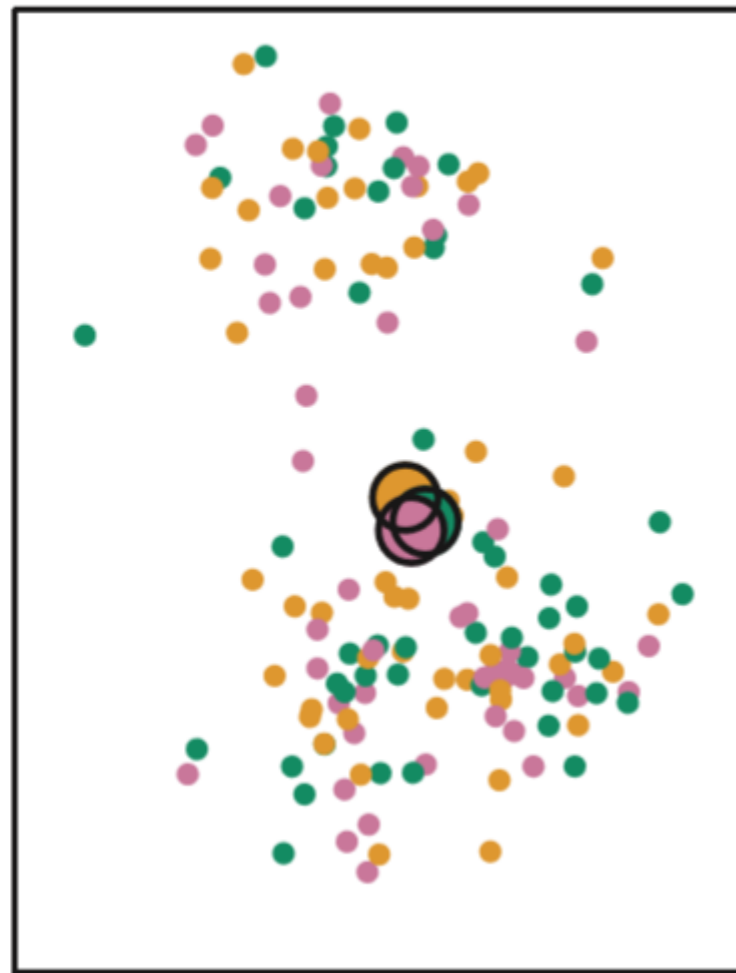
Data



Step 1

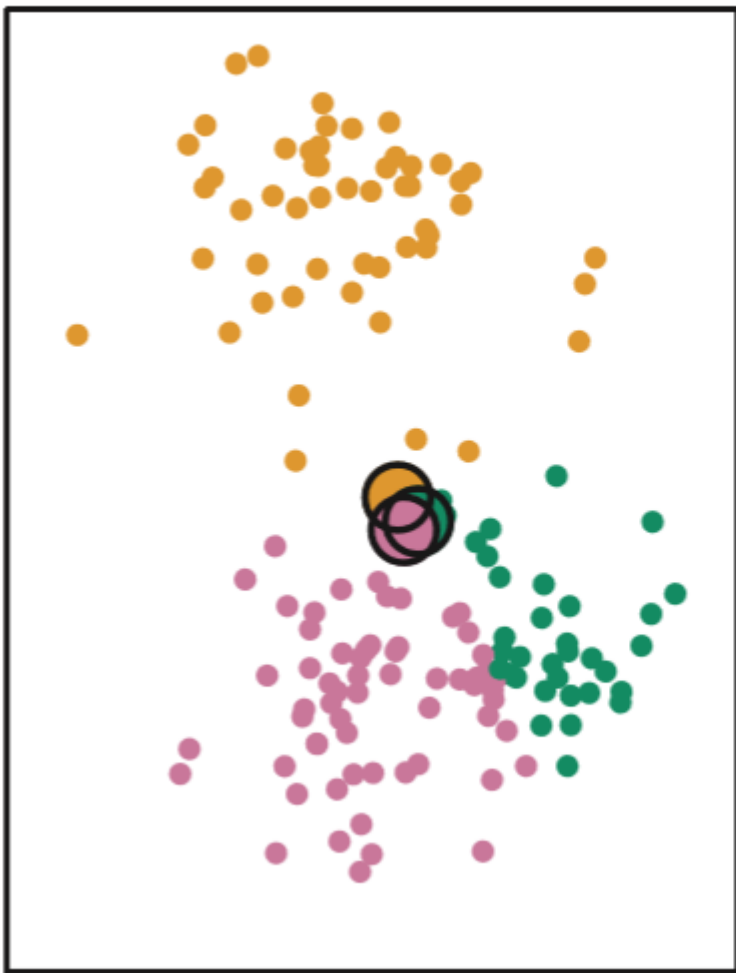


Iteration 1, Step 2a

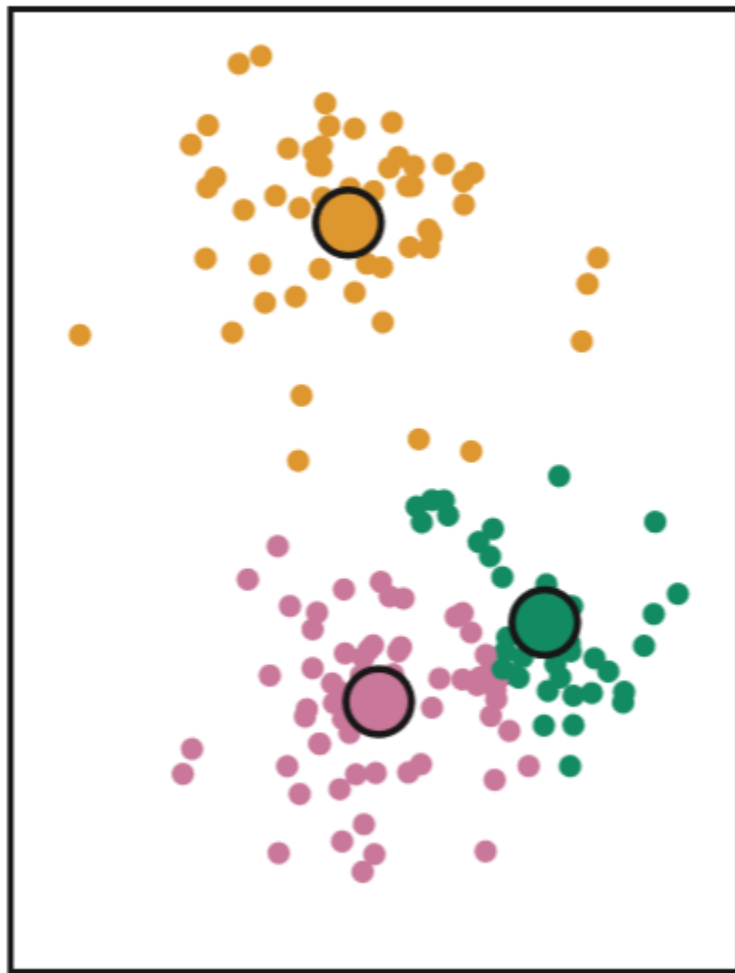


2. 클러스터링 – K-평균

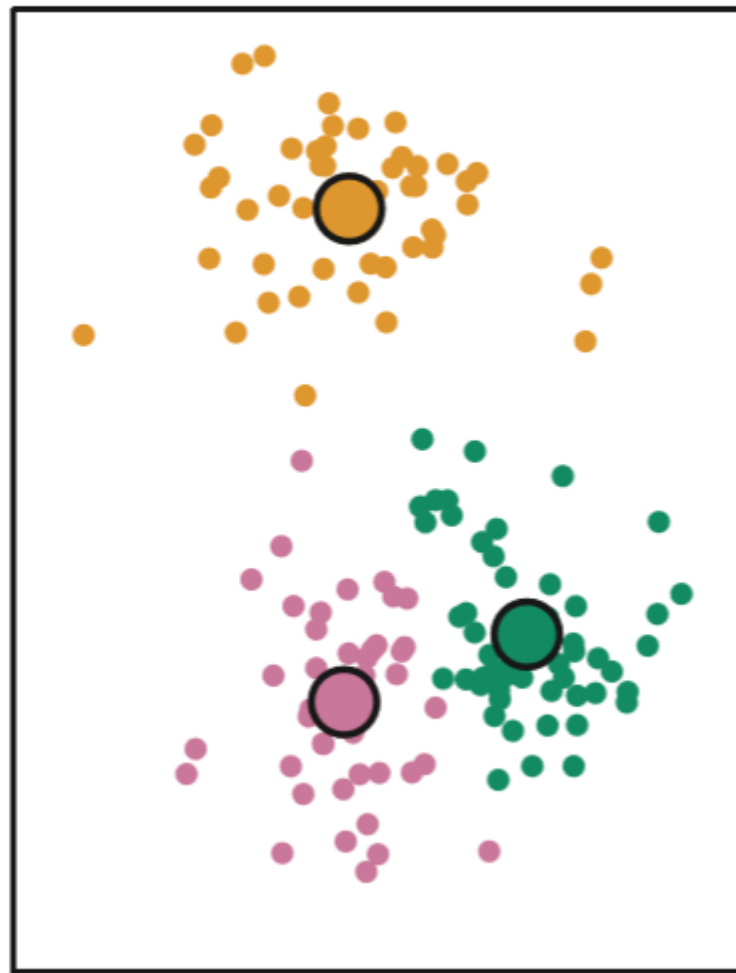
Iteration 1, Step 2b



Iteration 2, Step 2a

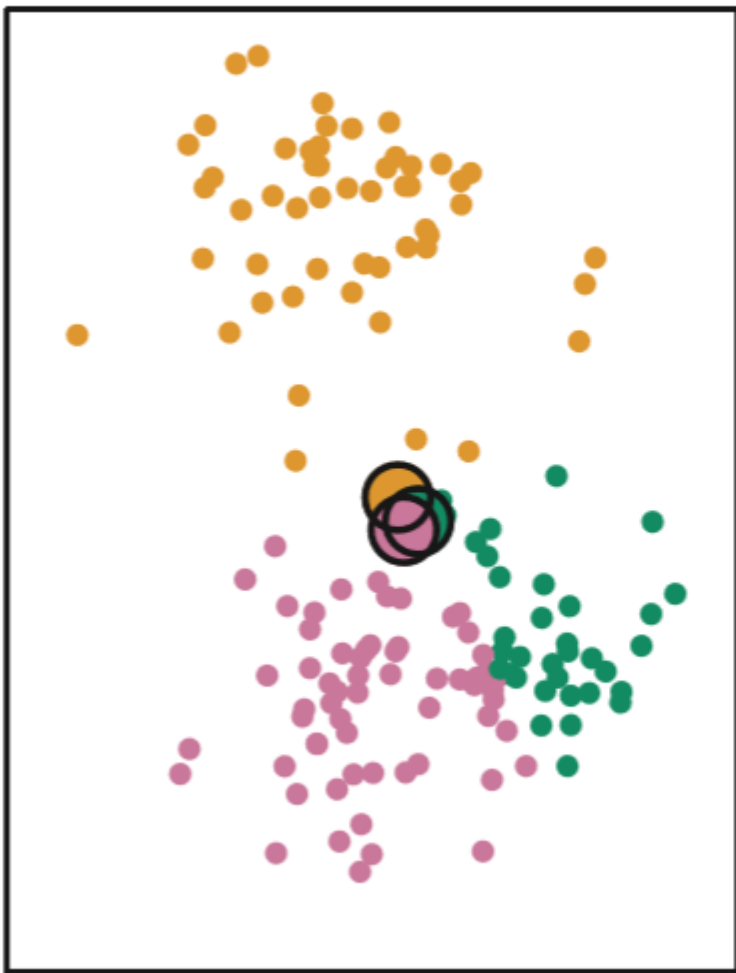


Final Results

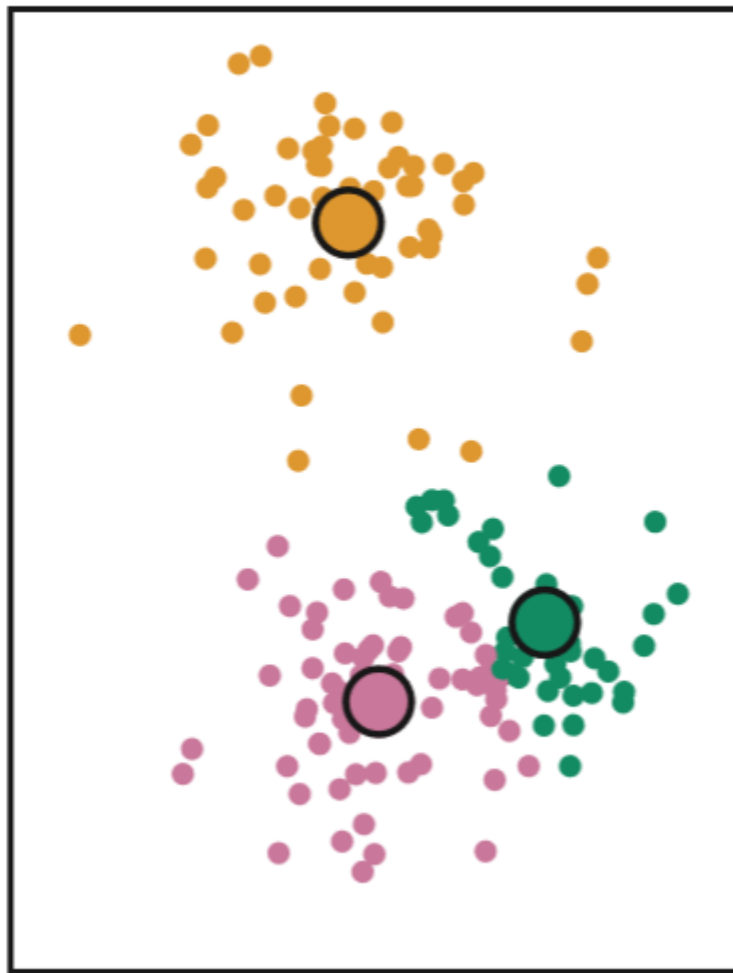


2. 클러스터링 – K-평균 단점 – K설정

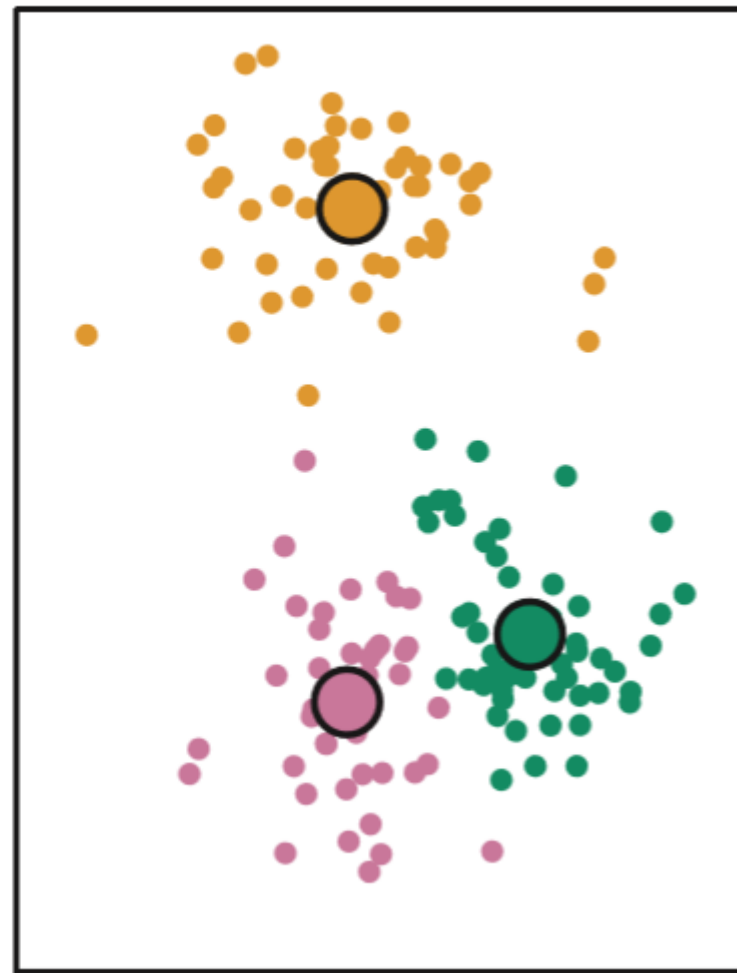
Iteration 1, Step 2b



Iteration 2, Step 2a

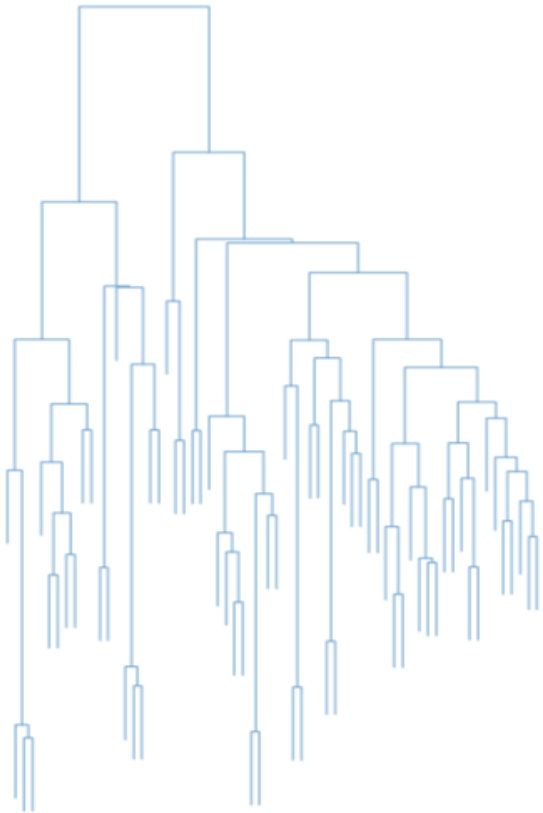
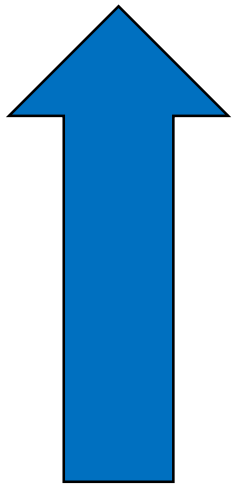


Final Results



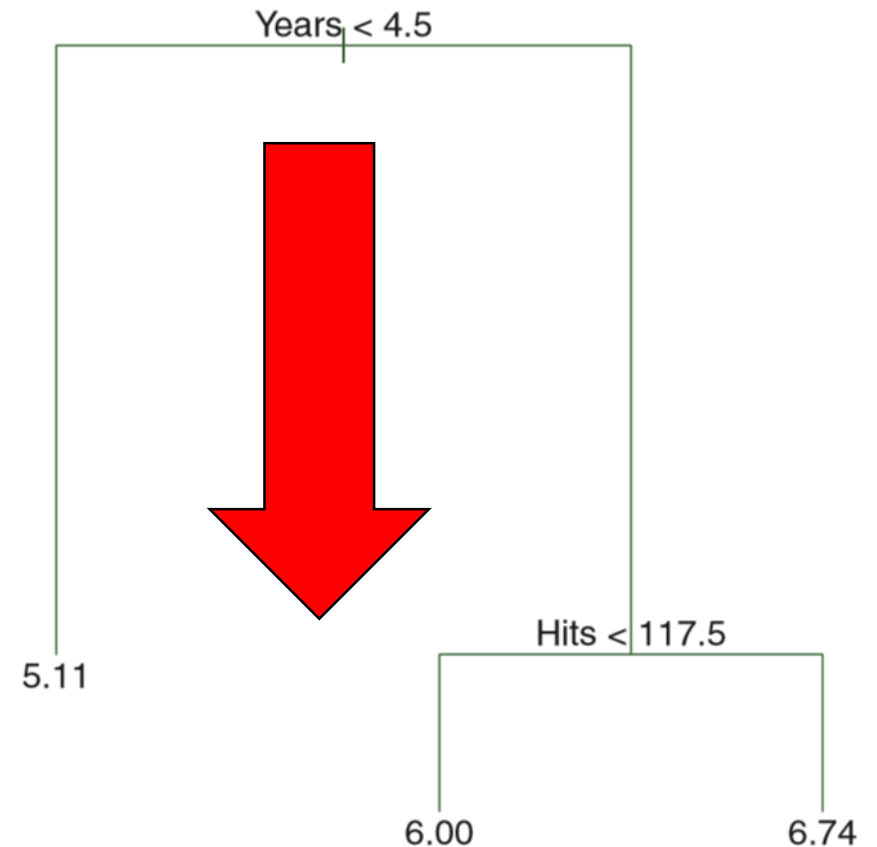
2. 클러스터링 – 계층적 클러스터링 VS. D.T

비지도 학습: 계층적 클러스터링



VS.

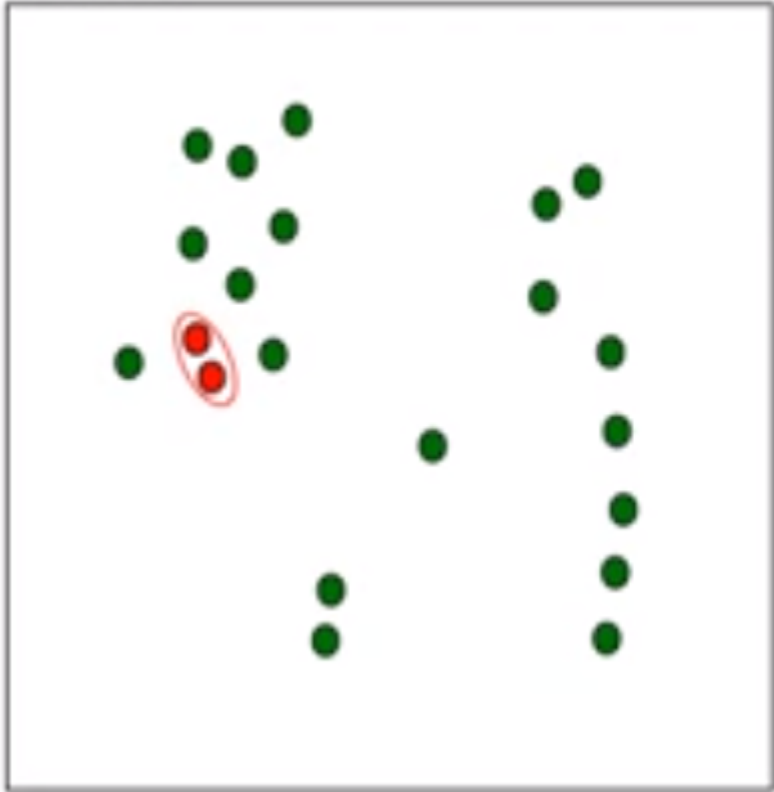
지도 학습: 의사결정트리



2. 클러스터링 – 계층적 클러스터링 Step 1

모든 쌍에 비유사성 측도 계산 후, 가장 작은 것 선택 → 융합 : 이때 덴드로그램의 높이는 비유사성

Data:



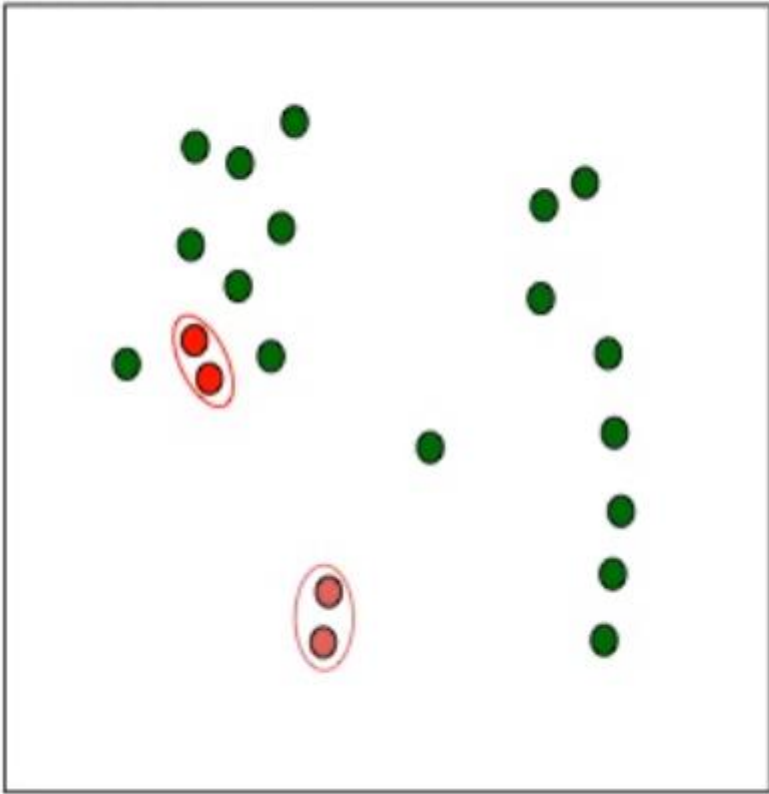
Dendrogram:



Height of the join
indicates dissimilarity

2. 클러스터링 – 계층적 클러스터링 Step 2

Data:

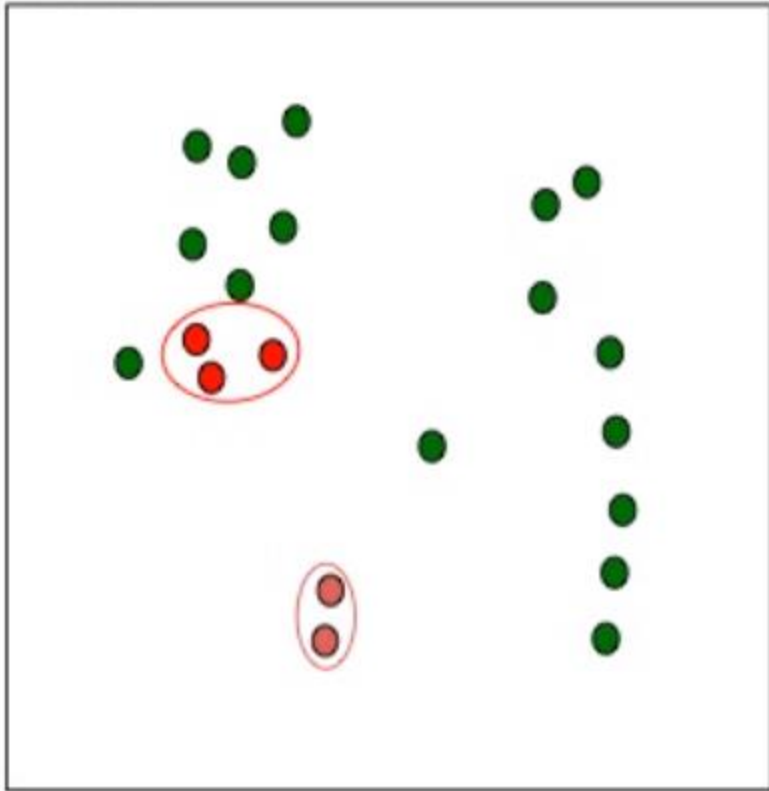


Dendrogram:

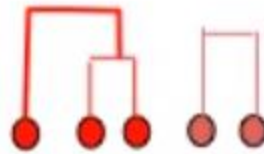


2. 클러스터링 – 계층적 클러스터링 Step 3

Data:



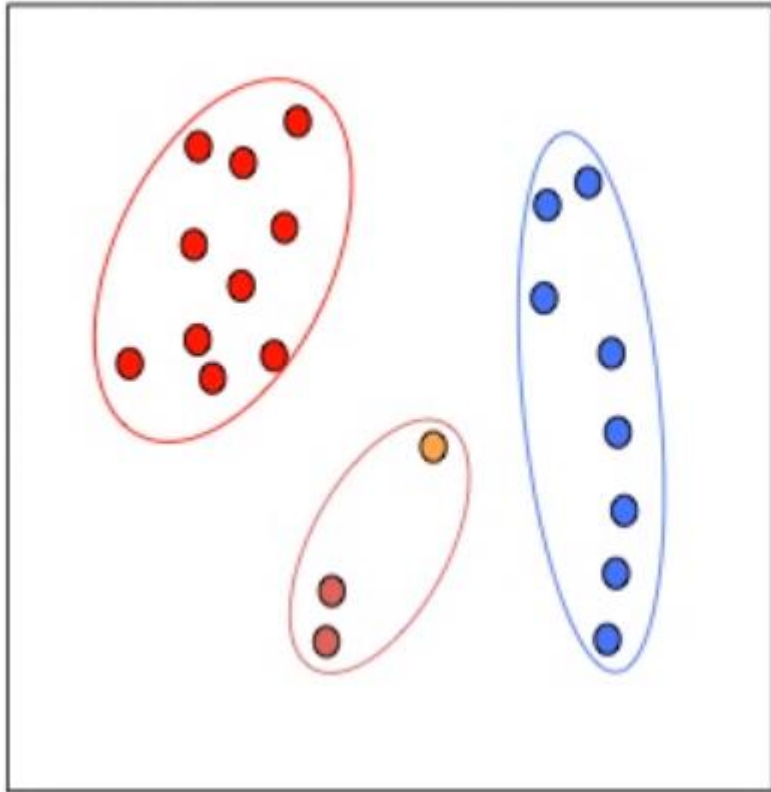
Dendrogram:



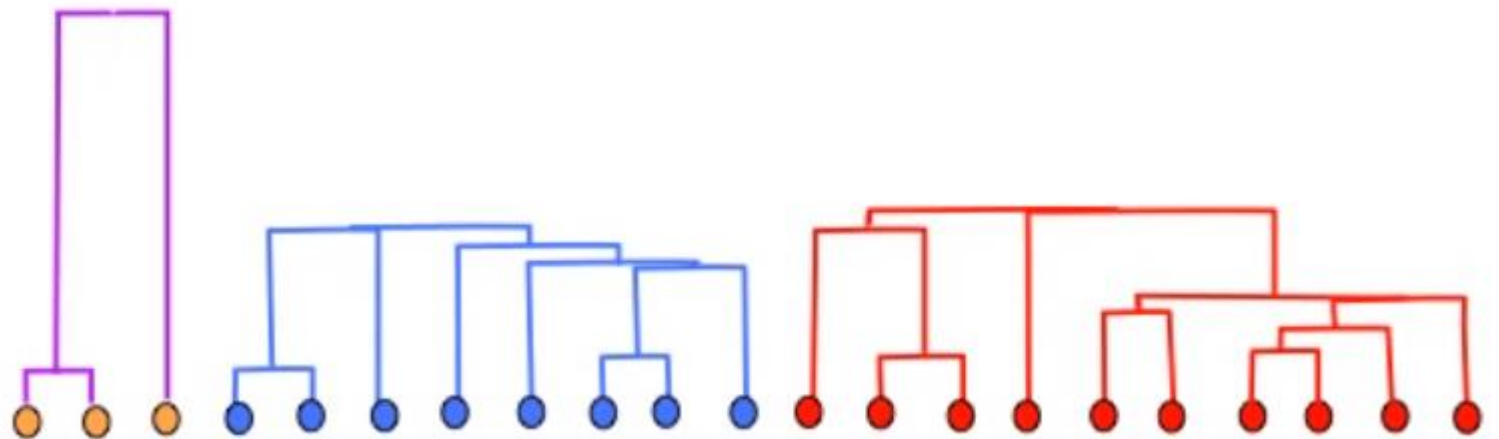
Height of the join
indicates dissimilarity

2. 클러스터링 – 계층적 클러스터링 Step m-3

Data:

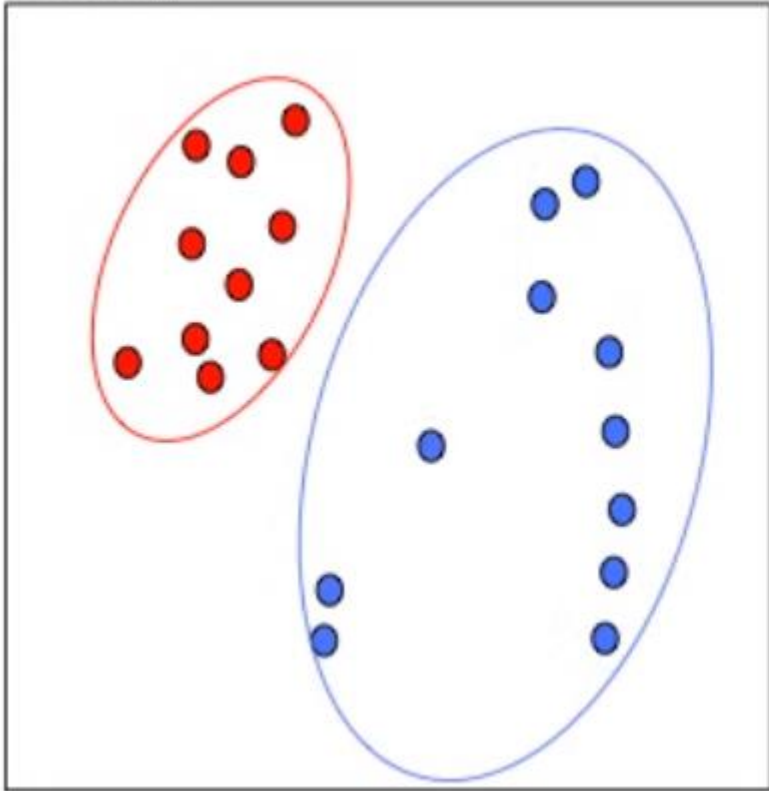


Dendrogram:

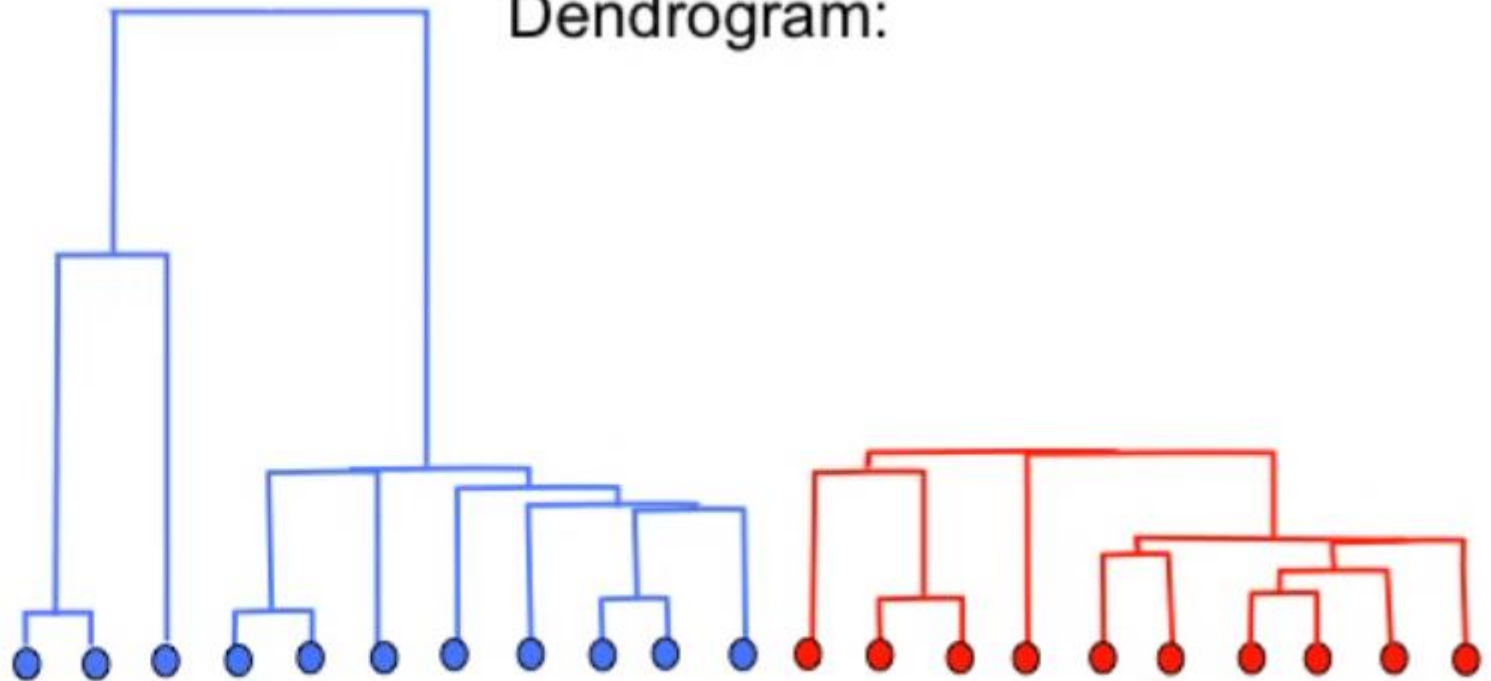


2. 클러스터링 – 계층적 클러스터링 Step m-2

Data:

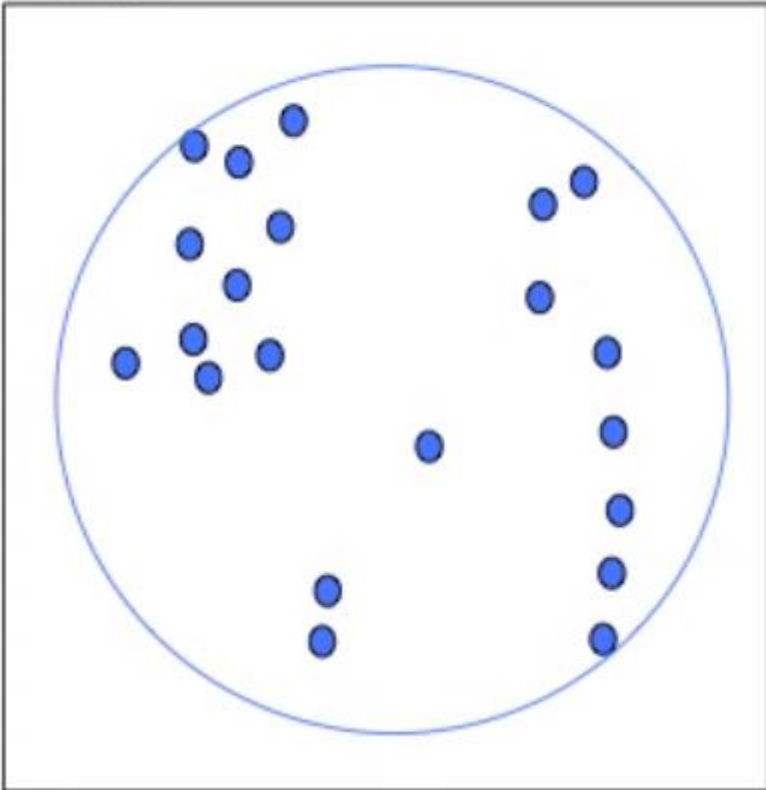


Dendrogram:

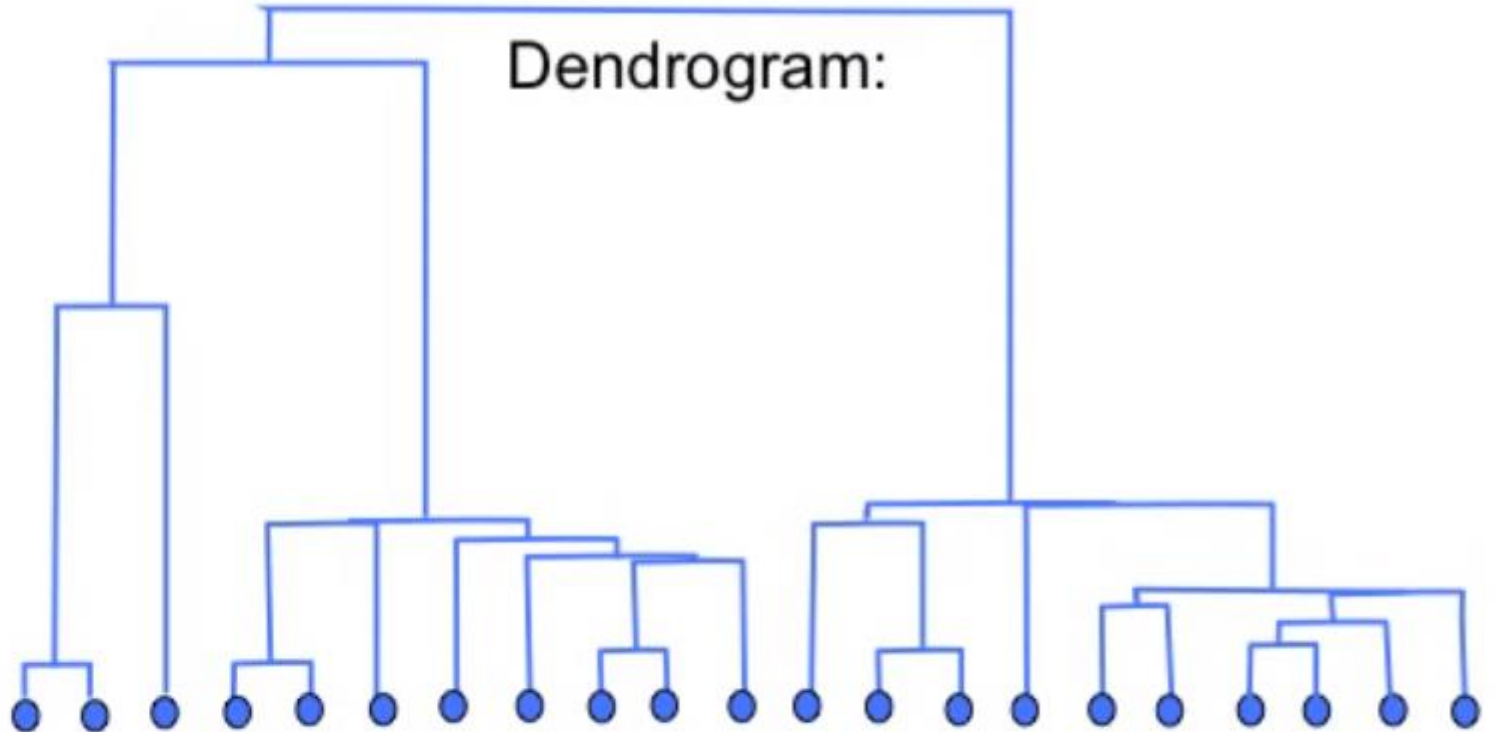


2. 클러스터링 – 계층적 클러스터링 Step m-1

Data:



Dendrogram:

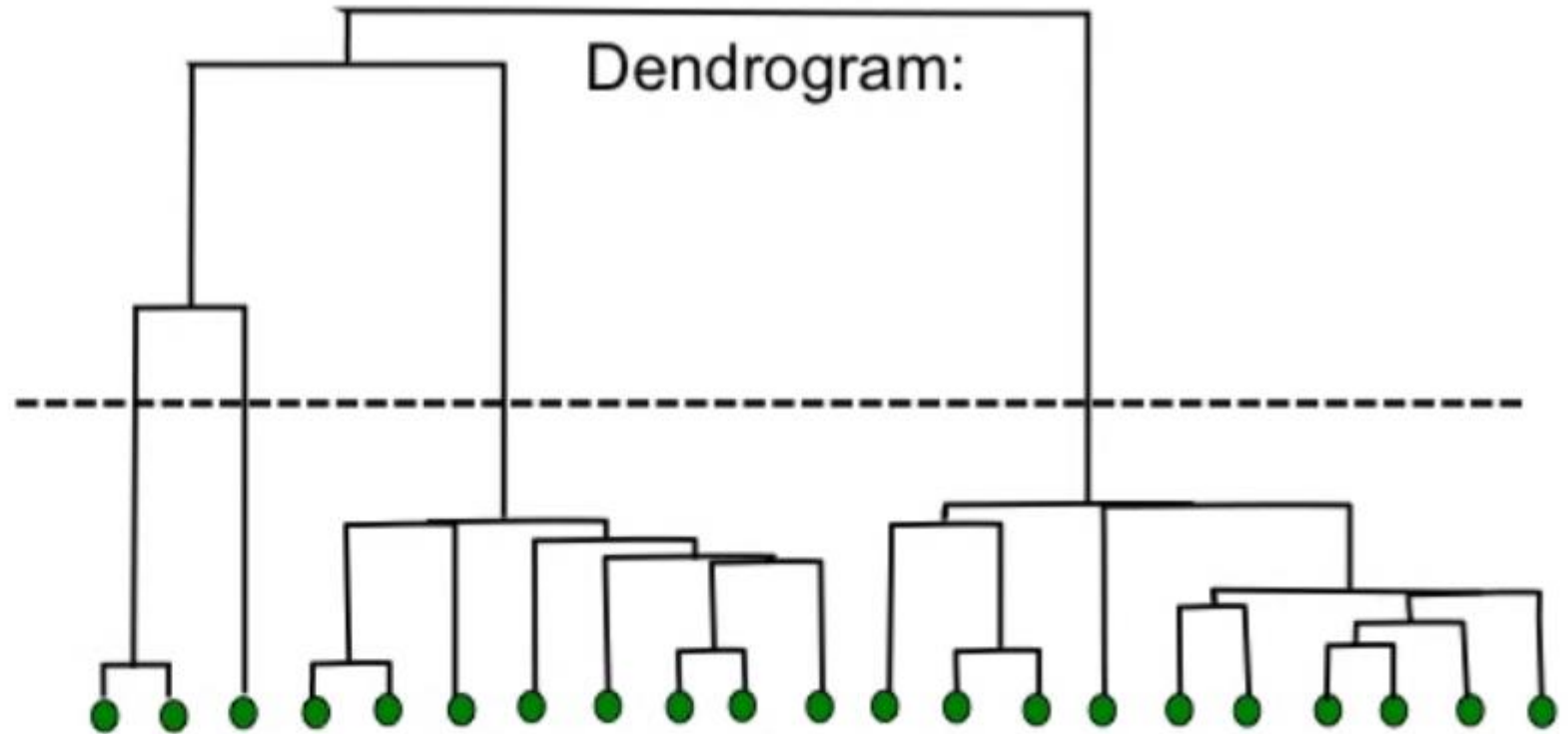
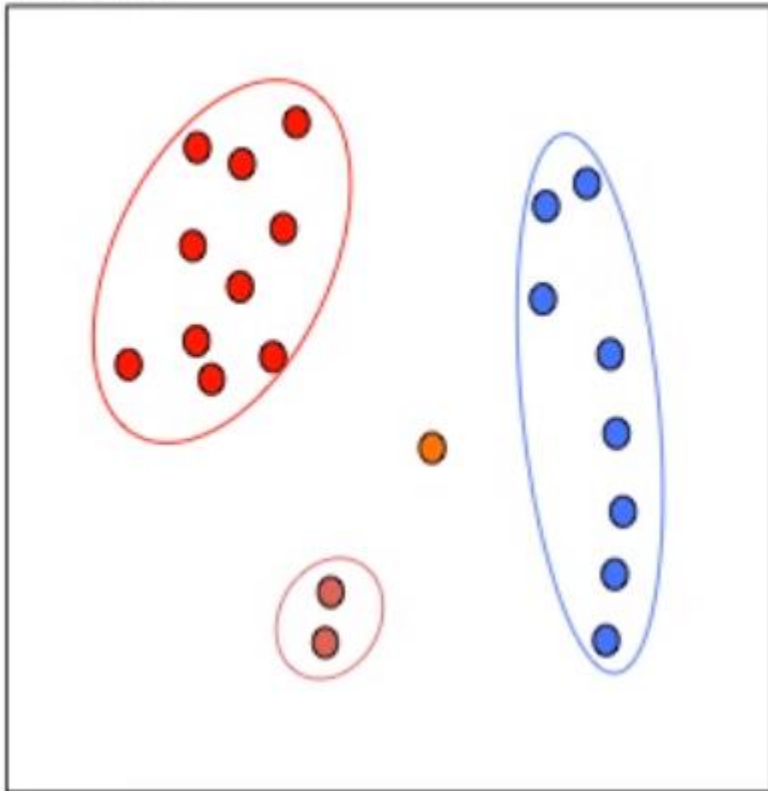


2. 클러스터링 – 계층적 클러스터링 **Cut!**

K-평균에서의 K와 같은 역할을 하지만, 장점으로 작용 가능! → 처음에 지정하지 않고, 나중에 조절 가능하므로!

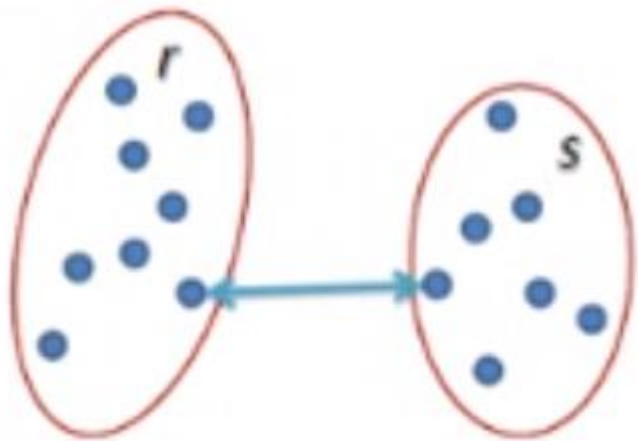
기준은 명확하지 않음!

Data:



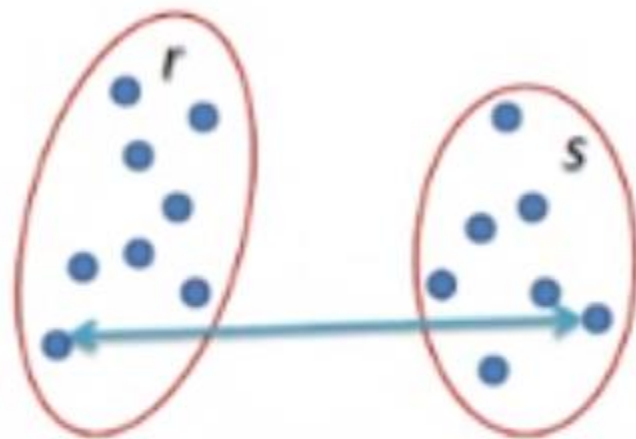
2. 클러스터링 – 계층적 클러스터링 연결

단일 연결



$$L(r, s) = \min(D(x_{ri}, x_{sj}))$$

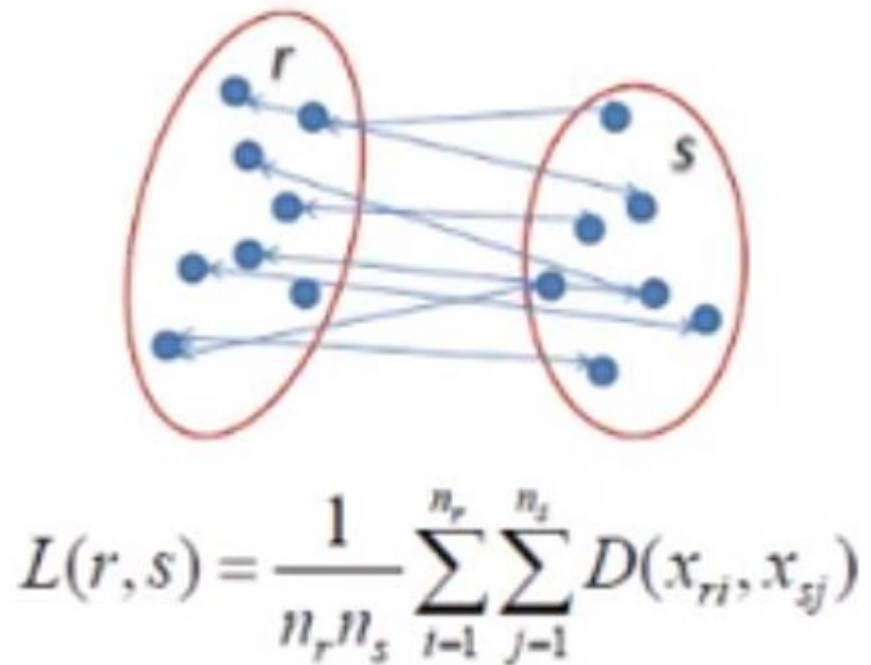
완전 연결



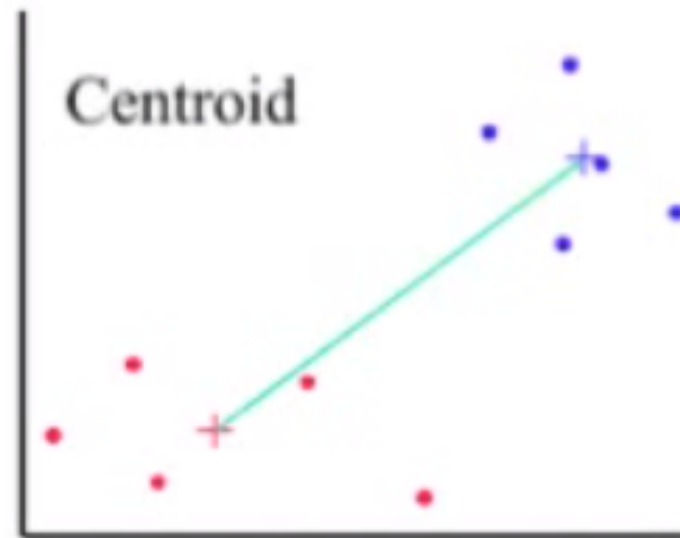
$$L(r, s) = \max(D(x_{ri}, x_{sj}))$$

2. 클러스터링 – 계층적 클러스터링 연결

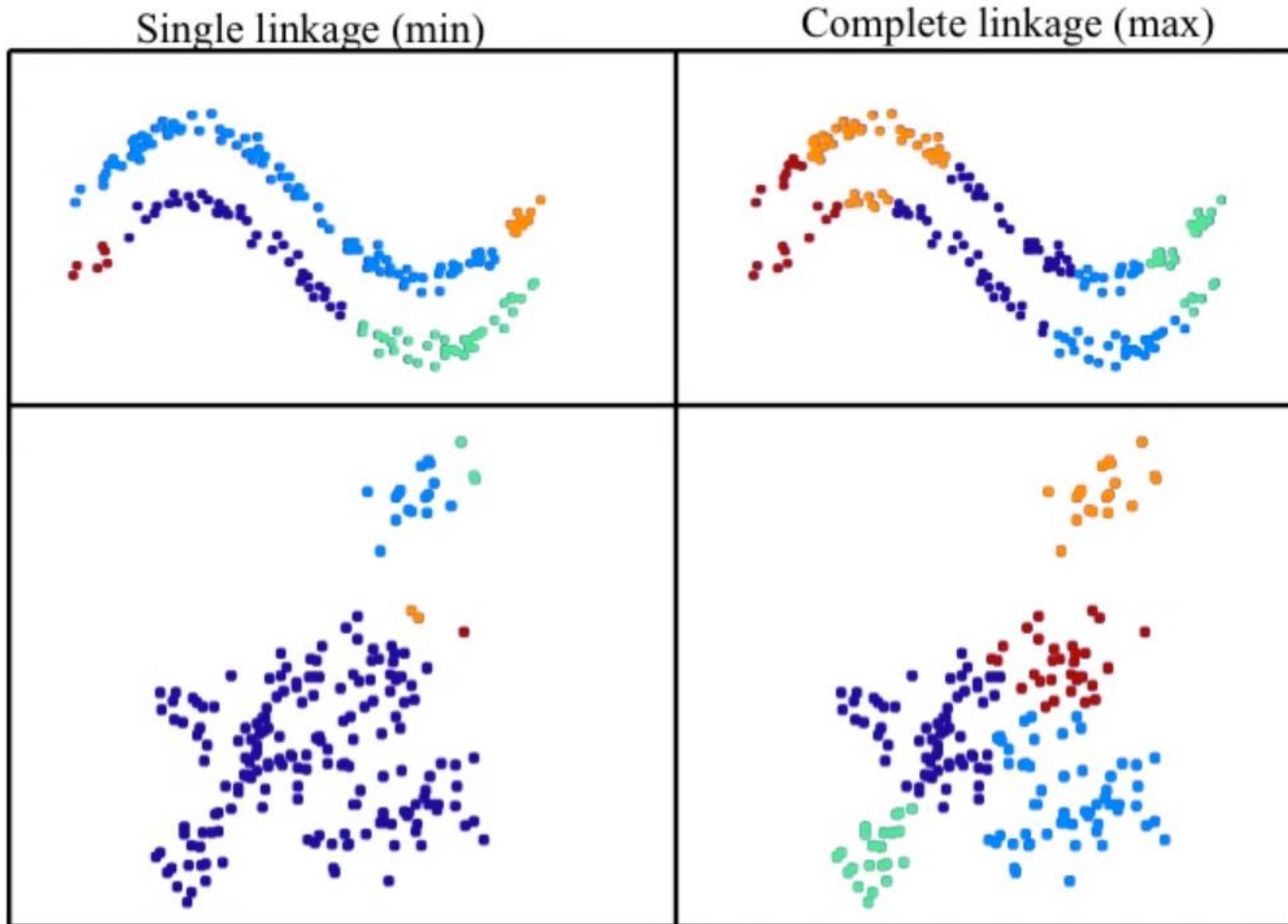
평균 연결



무게중심 연결



2. 클러스터링 – 계층적 클러스터링 연결 비교



정리

최적화 방법론

1. 분석적 풀이
2. 내리막 경사법
3. 라그랑제 승수

SVM

1. 최대 마진 분류기
2. 서포트 벡터 분류기
3. 서포트 벡터 머신

Unsupervised

1. PCA
2. 클러스터링

수식 관련 문제 완전한 이해 기준

1. 수식적 풀이

2. 기하학적 상상

3. 물리적인 직관

2. 서포트 벡터 분류기 – 수식적 관점

$$\underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1,$$

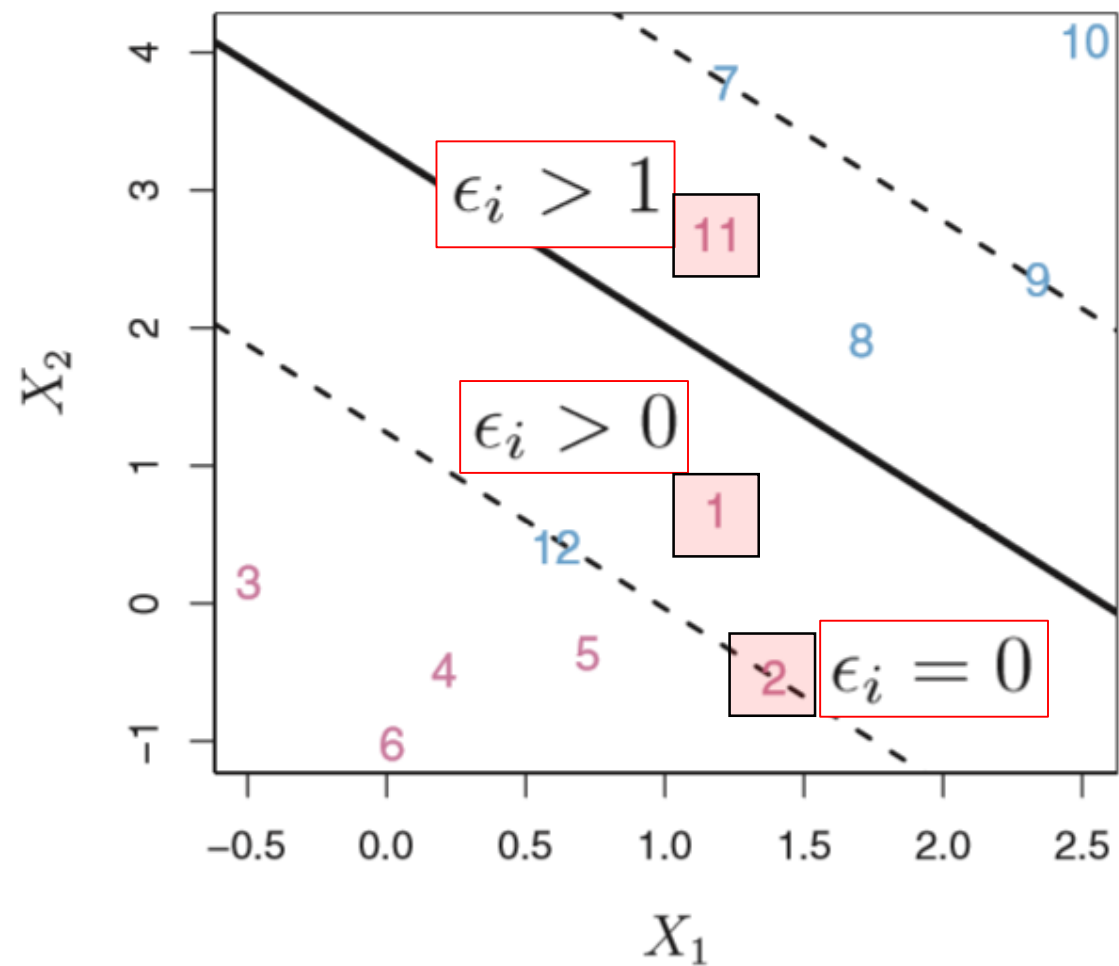
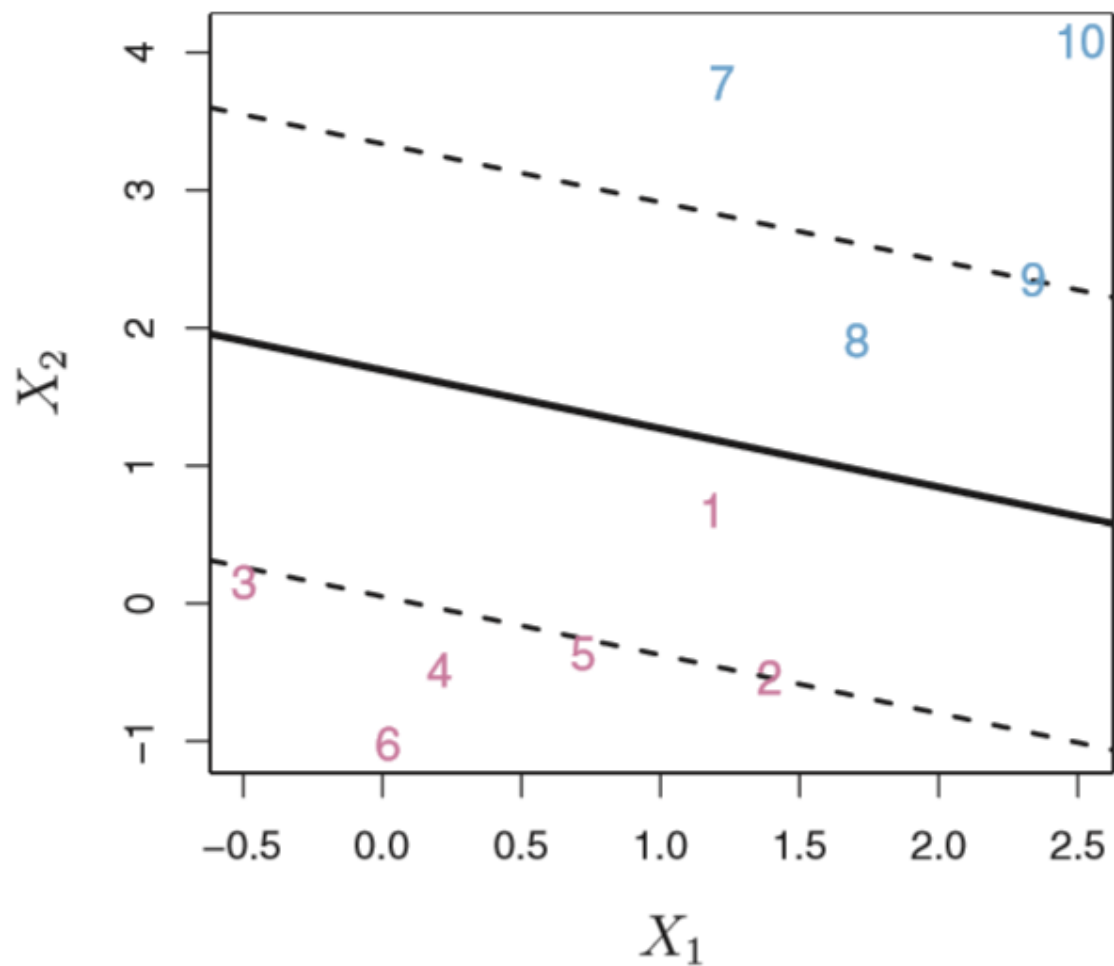
$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i),$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C,$$

조율 파라미터: 허용될 위반의 수와 그 정도 결정

슬랙(느슨한) 변수: i번째 관측치의 상대적 위치 정보

2. 서포트 벡터 분류기



1. PCA 주성분분석

- 1) 데이터 해석력이 좋아질 수 있다.
- 2) 데이터 차원이 축소될 수 있다.

Thank you!