

13

주차



로지스틱 회귀

머신러닝과 빅데이터분석(R)
박길식 교수




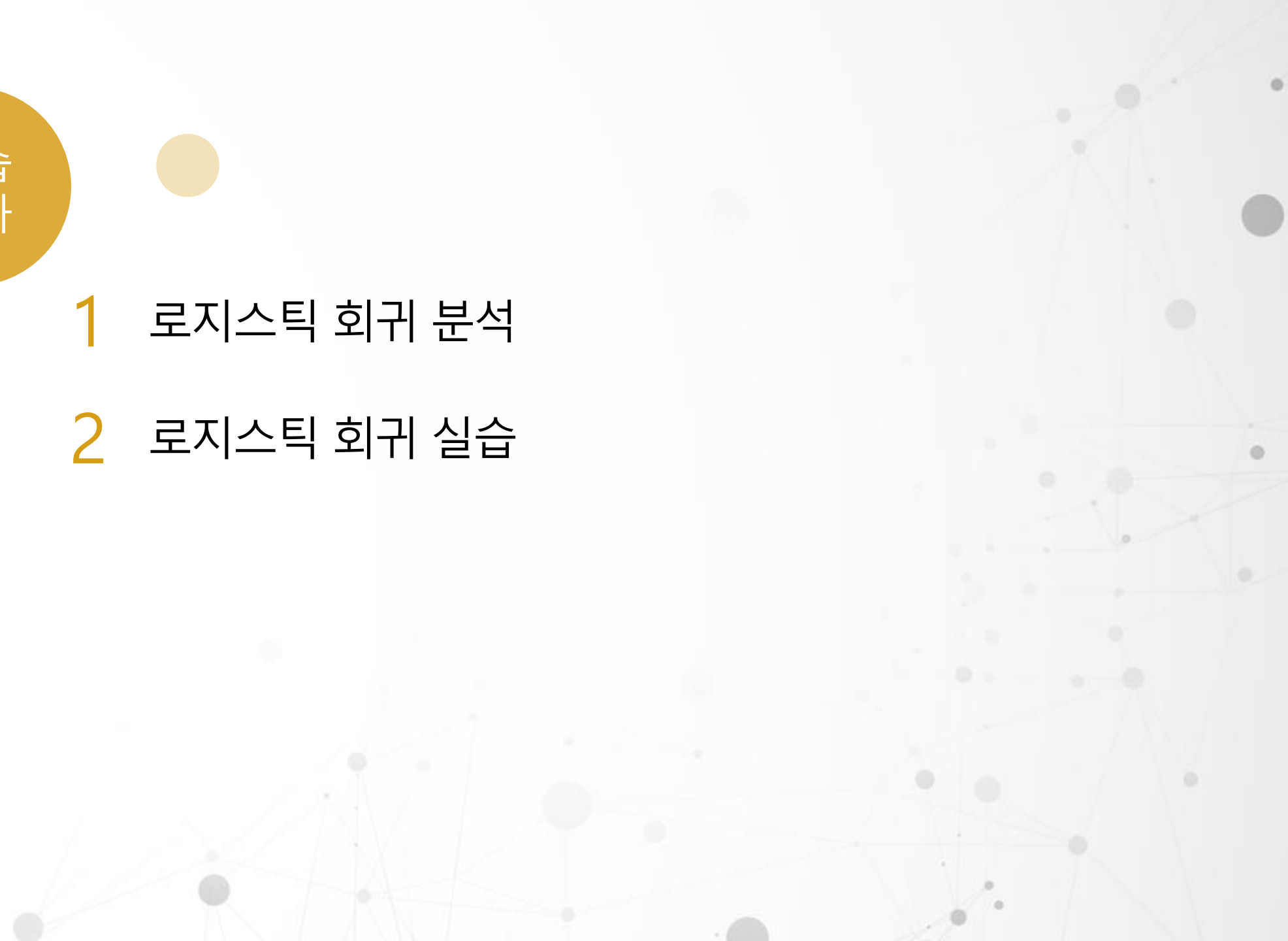


학습 목표

-  로지스틱 회귀분석의 원리를 이해하고 구현할 수 있다.
-  활성화 함수를 설명할 수 있다.



학습 목차

- 
- 1 로지스틱 회귀 분석
 - 2 로지스틱 회귀 실습
- 

CHAPTER

「
01
」

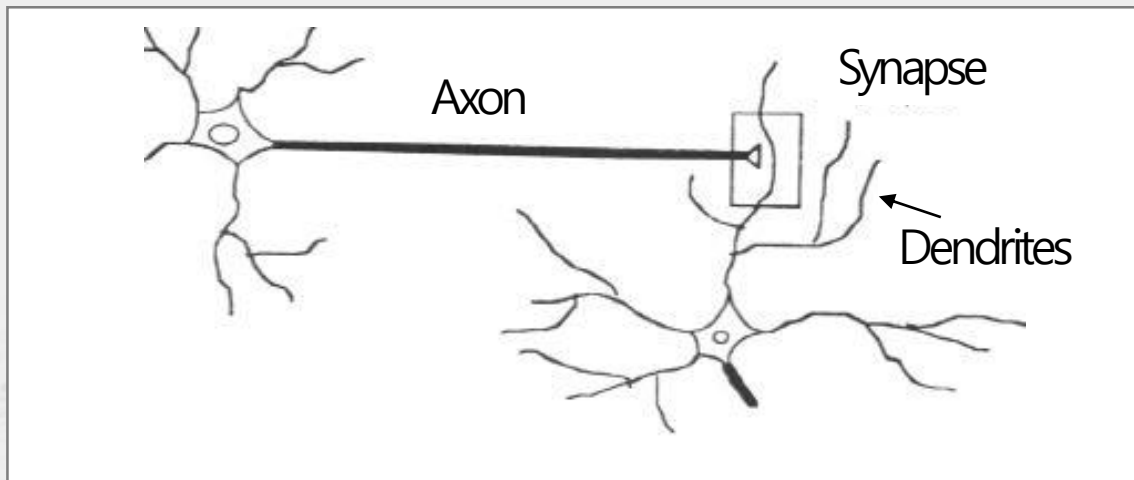
|

신경망과 SVM의 이해

인공 신경망 모형

인간의 뉴런의 자극전달 과정에 아이디어를 착안하여
발생한 머신러닝 알고리즘

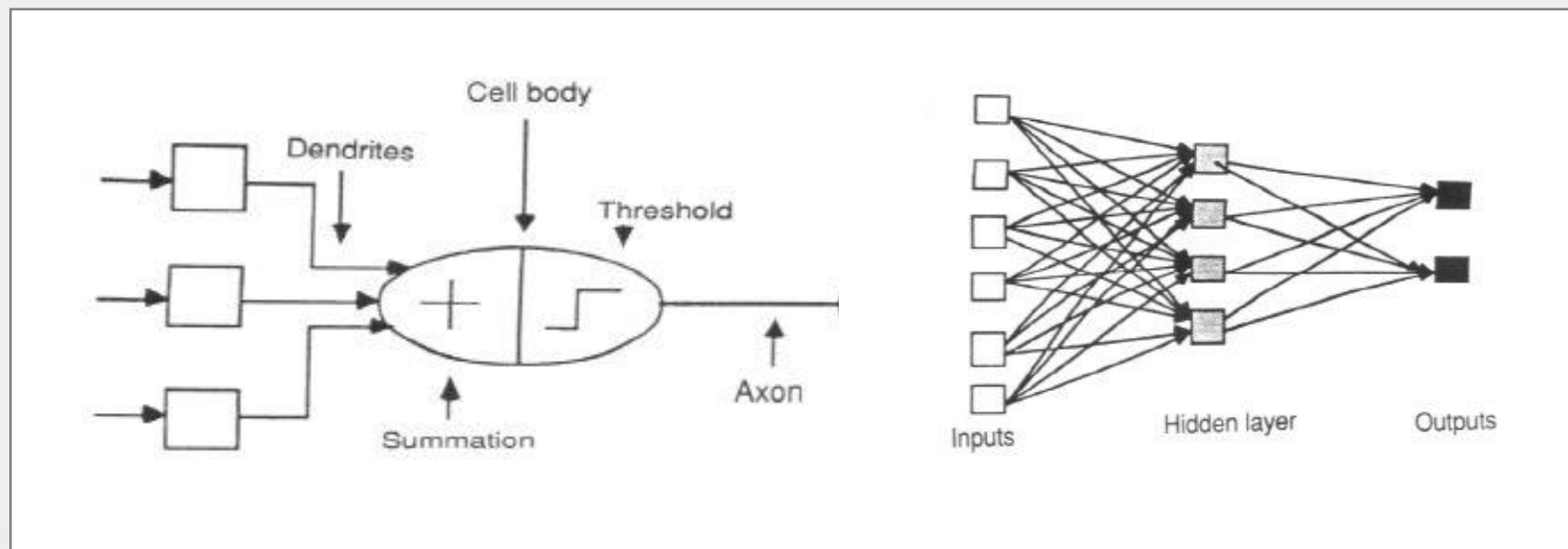
- ㉞ 인간의 뉴런은 시냅스를 통하여 다른 뉴런으로부터 자극을 전달받고 시냅스를 통하여 다른 뉴런에게 자극을 전달하는 과정을 통해서 학습을 진행



교수님
이미지 출처
알려주세요

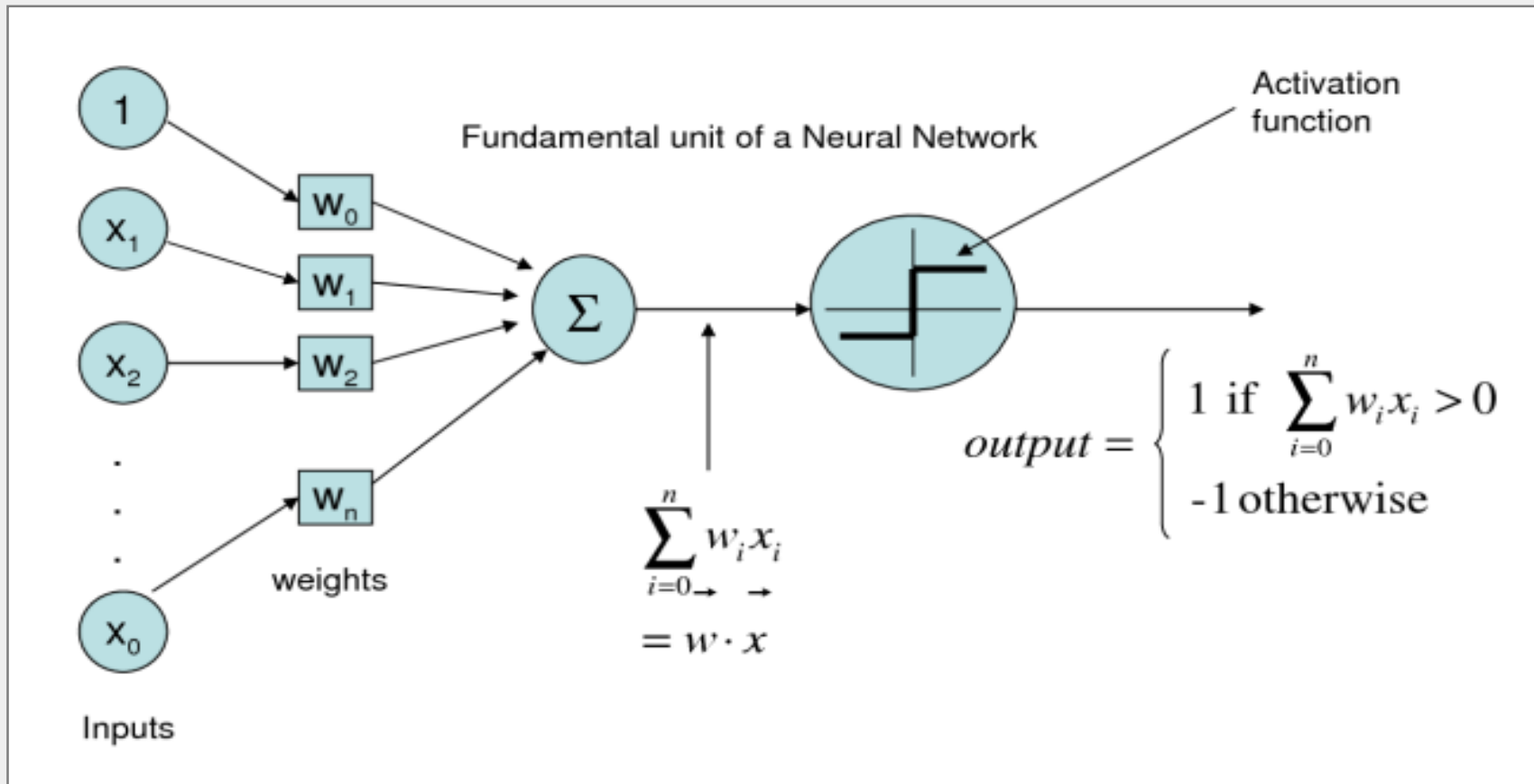
[01] 신경망

- » 자극 전달의 과정을 알고리즘에서 layer와 perceptron으로 뉴런과 시냅스를 구성하여 연결한 것
- » 뉴런은 Neural Network 모형에서 하나의 perceptron과 대응
- » 시냅스의 역할은 여러 layer를 잇는 weight/bias(error)가 됨



교수님
이미지 출처
알려주세요

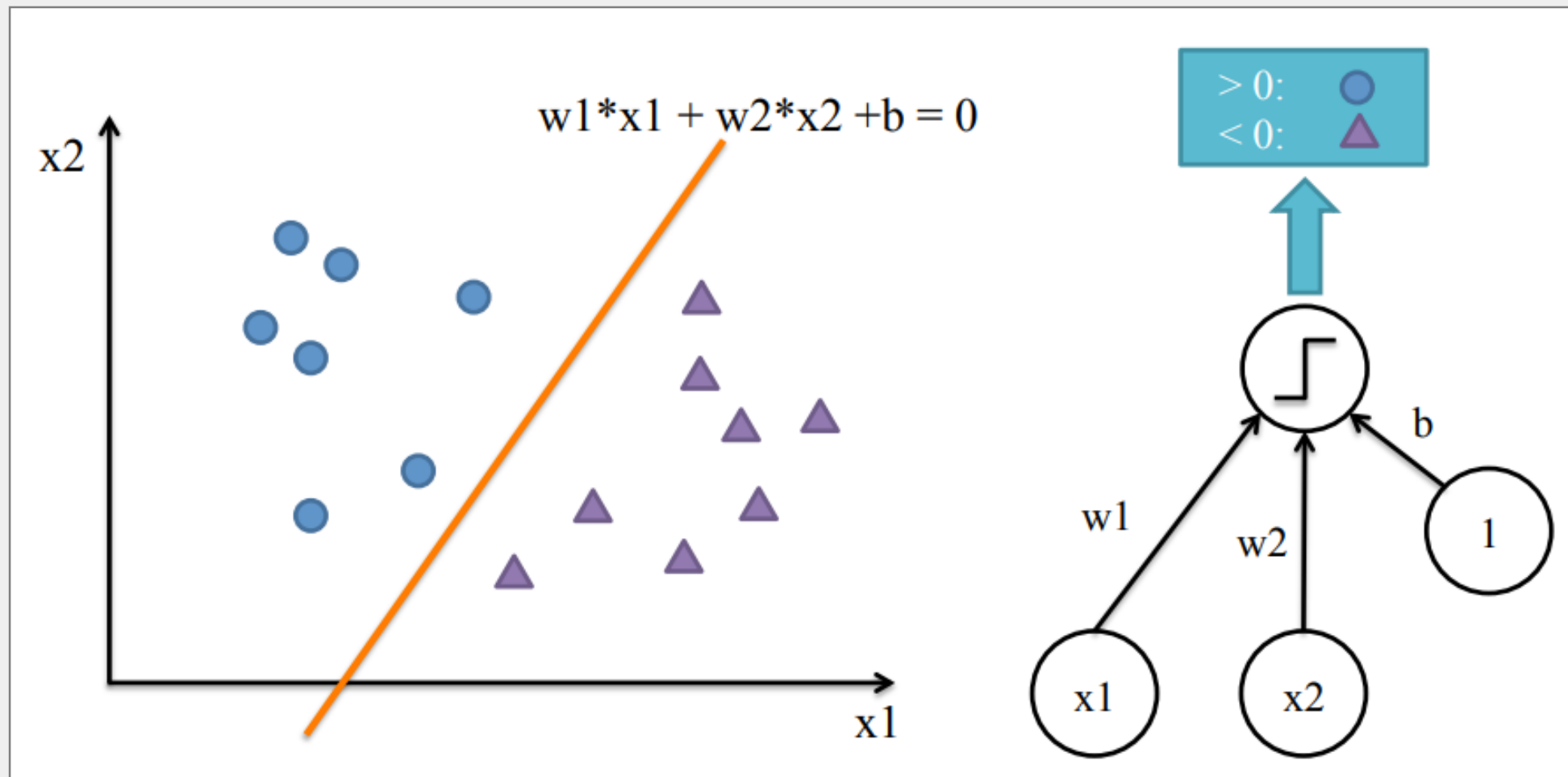
[01] 신경망



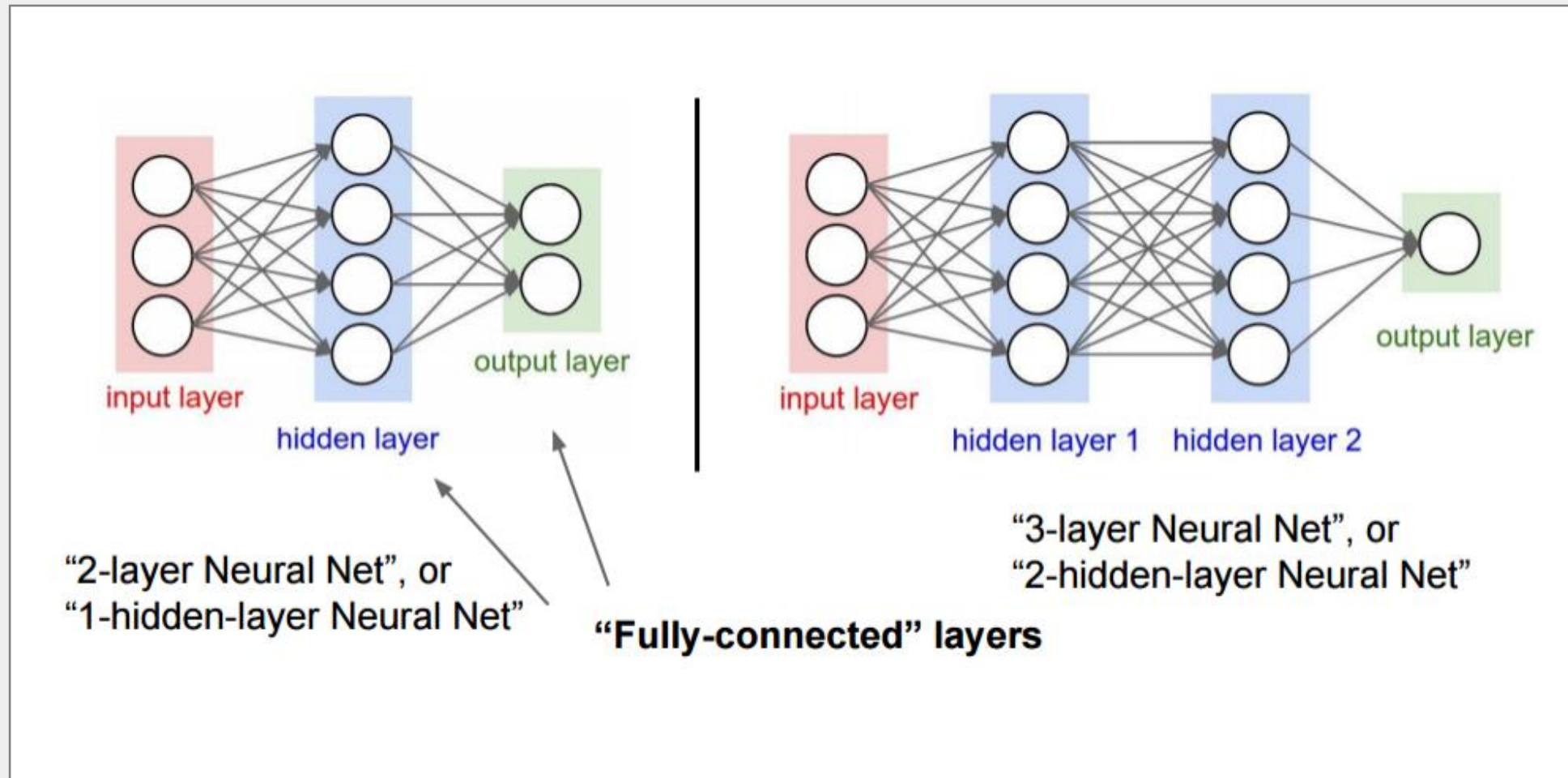
⊕ 퍼셉트론은 inputs / weights / transfer function / activation function으로 이루어짐

⊕ 예) 2개의 입력변수 x_1, x_2 가 존재한다고 할 때 $w_1 x_1 + w_2 x_2 + w_0$

[01] 신경망



[01] 신경망



서포트 벡터 머신(Support Vector Machine)

- 고차원 또는 무한 차원의 공간에서 초평면의 집합을 찾아 이를 이용하여 분류(SVC)와 회귀(SVR) 문제를 수행
 - 비중첩(Non-overlapping) 분할을 제공하며 모든 속성을 활용하는 전역적(Global) 모델
- ② 최대 마진을 가지는 선형 판별에 기반하며 속성들 간의 의존성을 고려하지 않음

결정 경계(Decision Boundary)

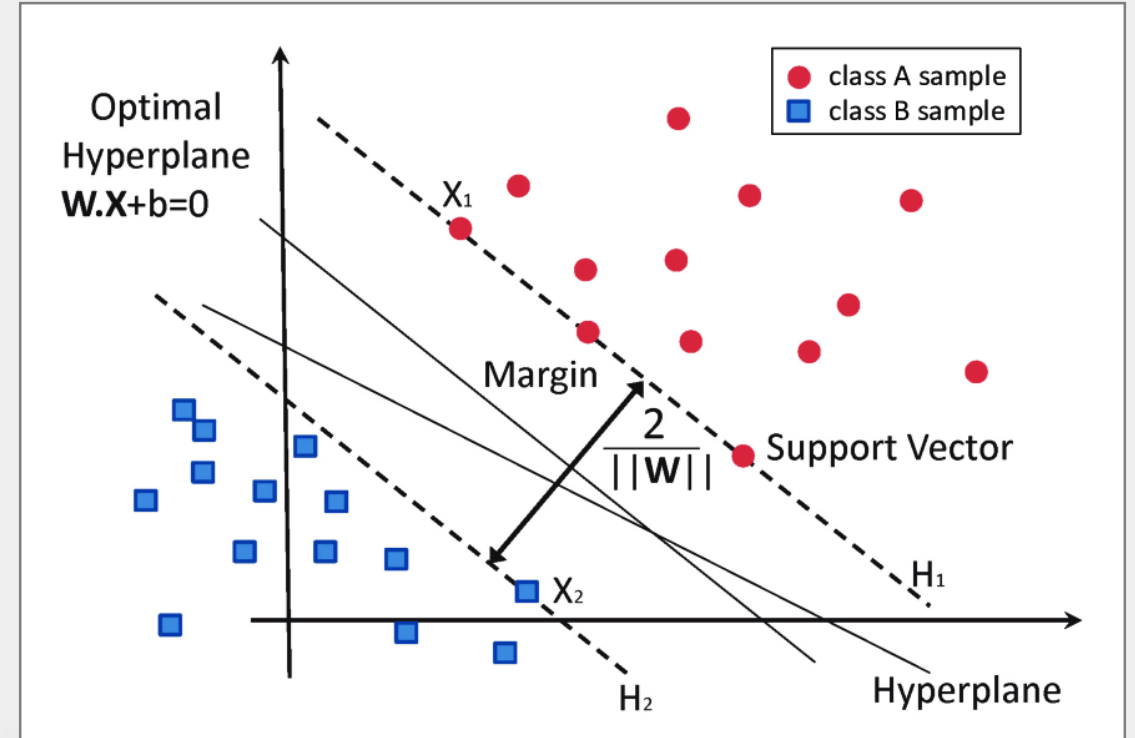
- 두 개의 클래스를 가장 잘 분류할 수 있는 초평면을 의미

서포트 벡터(Support Vector)

- 결정 경계와 가까이 있는 데이터 포인트

마진(Margin)

- 결정 경계와 서포트 벡터 사이의 거리
- 최적의 결정 경계는 마진을 최대화 함





서포트 벡터 머신 알고리즘의 매개변수

② C(Cost)

SVM의 오차를 어느 정도 허용할 것인지를 설정

② Kernel

데이터가 선형으로 분리하기 어려울 때 비선형으로 차원을 변화 시킴

② Gamma

데이터를 구분 짓는 경계선을 얼마나 유연하게 그을 것인지를 정함



서포트 벡터 머신 알고리즘의 매개변수

장점

- 과대적합/과소적합의 정도를 고려한다면 매우 우수한 예측 성능 도출
- 서포트 벡터만 잘 선정할 수 있다면 다른 데이터를 사용하지 않더라도 좋은 성능을 보여줄 수 있음

단점

- 초매개변수의 조정으로 인해 최적화된 모델을 만들기 위한 모델 구성 시간이 오래 걸림
- 입력 데이터 셋이 많을 경우 학습속도가 현저히 느림
- 결과에 대한 설명력이 떨어짐

2차시

PRACTICE

「 실습 」

|

SVM 실습



SVM 모델 구현



학습 정리



신경망

- 신경망 모형은 인간의 뉴런의 자극전달 과정에 아이디어를 착안하여 발생한 머신러닝 알고리즘



SVM

- 고차원 또는 무한 차원의 공간에서 초평면의 집합을 찾아 이를 이용하여 분류(SVC)와 회귀(SVR) 문제를 수행함

참고
문헌



강전희외1명, 혼자 공부하는 R 데이터 분석, 한빛미디어, 2022



양윤석외2명, R로 배우는 데이터 과학, 한빛아카데미, 2019



김영우, Do it! 쉽게 배우는 R 데이터 분석, 이지스퍼블리싱, 2017