

인공지능 모델 모델의 종류

학습 주제

인공지능 모델은 무엇이 있을까?

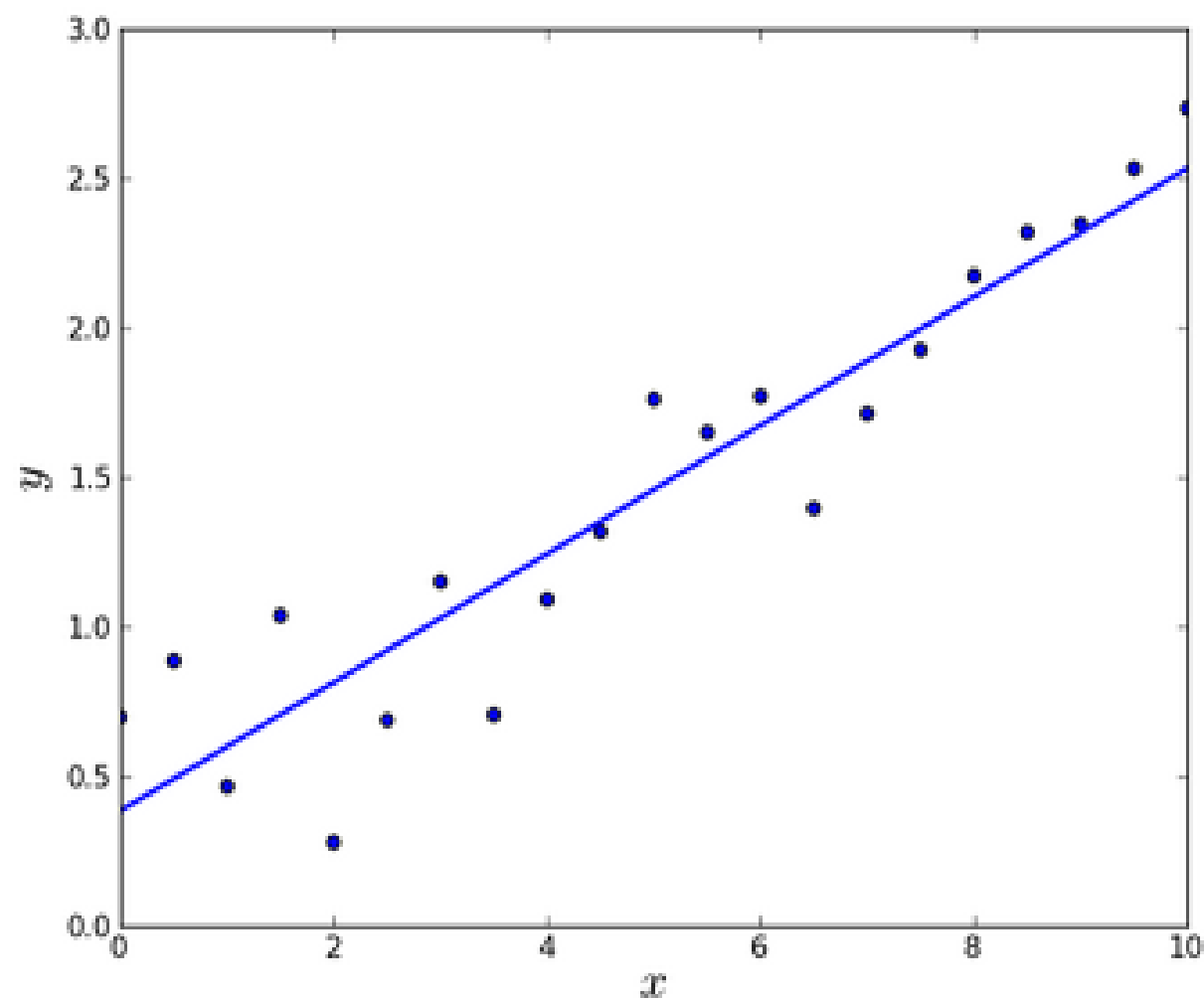
- 01 선형모델이란?
- 02 로지스틱모델이란?
- 03 서포터벡터머신이란?
- 04 의사결정나무란?
- 05 신경망모델이란?

학습 목표

1. 인공지능 모델에 대해 이해할 수 있다.
2. 데이터에 따른 모델을 스스로 결정해 나갈 수 있다.

01 선형모델

선형모델이란?



인공지능이 입력변수들을 통해 결과변수를 예측함에 있어, 입력변수와 결과변수가 선형 상관관계를 가지는 모델로 선형회귀라고도 불립니다.

회귀의 사전적 정의는 ‘한 바퀴 돌아 제자리로 돌아가다’로 입력변수들이 어떠한 과정을 통해서 결과변수의 값으로 가는 것입니다.

선형모델

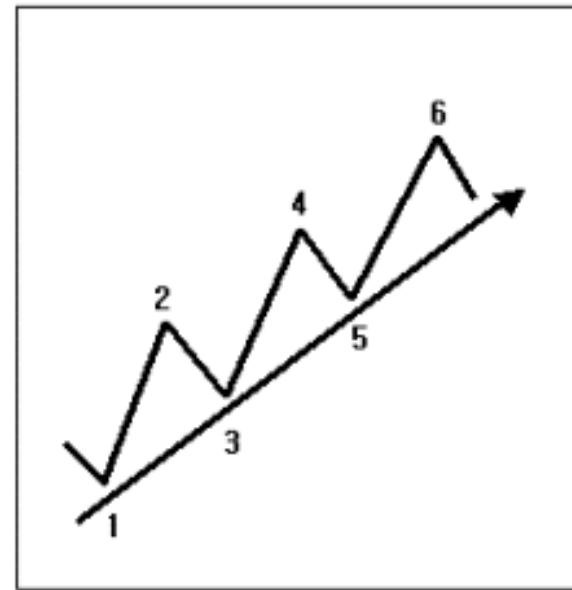
단순선형회귀모델

다중선형회귀모델

일반적으로 입력변수에 가중치를 곱하고 이들을 더하는 것으로 직선으로 표현 될 수도 있지만, 변수자체의 제곱을 통하여 곡선으로 나타날 수도 있습니다. 선형모델의 종류는 하나의 입력변수와 하나의 결과변수만 존재하는 단순선형회귀 모델과 여러 개의 입력변수와 하나의 결과변수를 가지는 다중선형회귀 모델이 있습니다. 선형모델은 생물학, 행동학, 경제학 및 기타 사회과학에서 입력변수들과 결과변수 사이의 관계를 설명하고자 할 때 사용될 수 있습니다. 예를 들어 추세분석과 역학조사에 대해 이야기 해 보겠습니다.

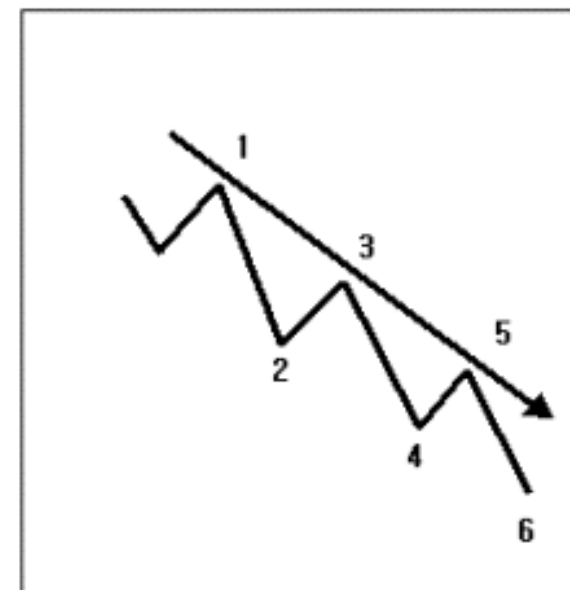
01 선형모델

그림 1)



상승추세에서의 추세선(↗표시)

그림 2)



하락추세에서의 추세선(↘표시)

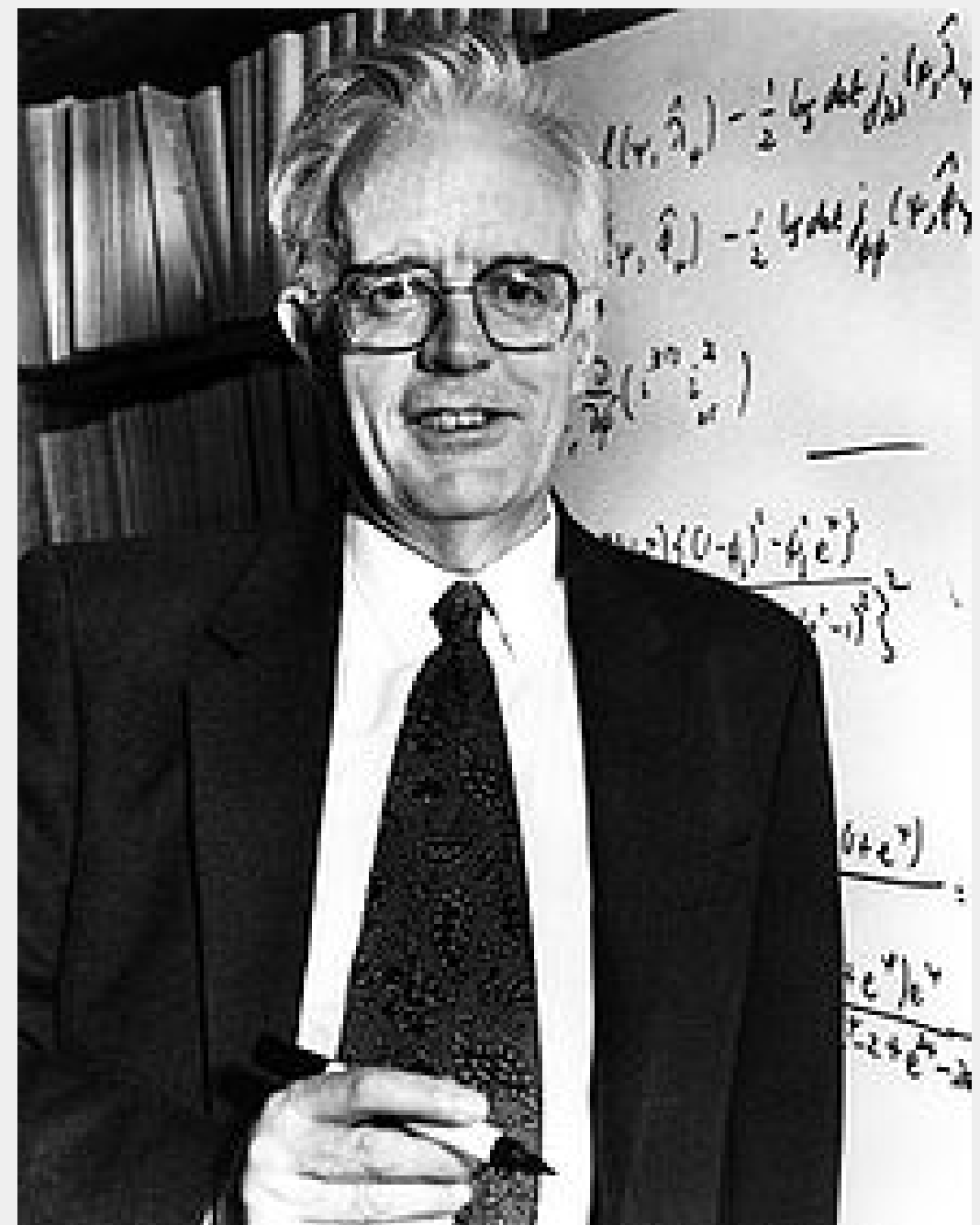
추세분석은 데이터를 시간 축에 놓고, 데이터의 값이 장기적으로 어떻게 변하는지 직선 또는 곡선으로 표현하는 것입니다. 이를 통해 값이 증가세에 있는지 감소세에 있는지 확인 할 수 있습니다. 역학조사는 질병의 수와 사망률사이의 관계성을 알고자 할 때, 질병이 많을수록 사망률이 높게 나타난다면 이는 역학관계가 있는 것입니다.

따라서 선형모델은 회귀모델로 실제값과 같은 연속형 숫자 데이터를 예측하는 회귀모델에 적합하며, 입력변수와 결과변수 사이의 관계가 상관관계를 보일 때 활용할 수 있습니다.

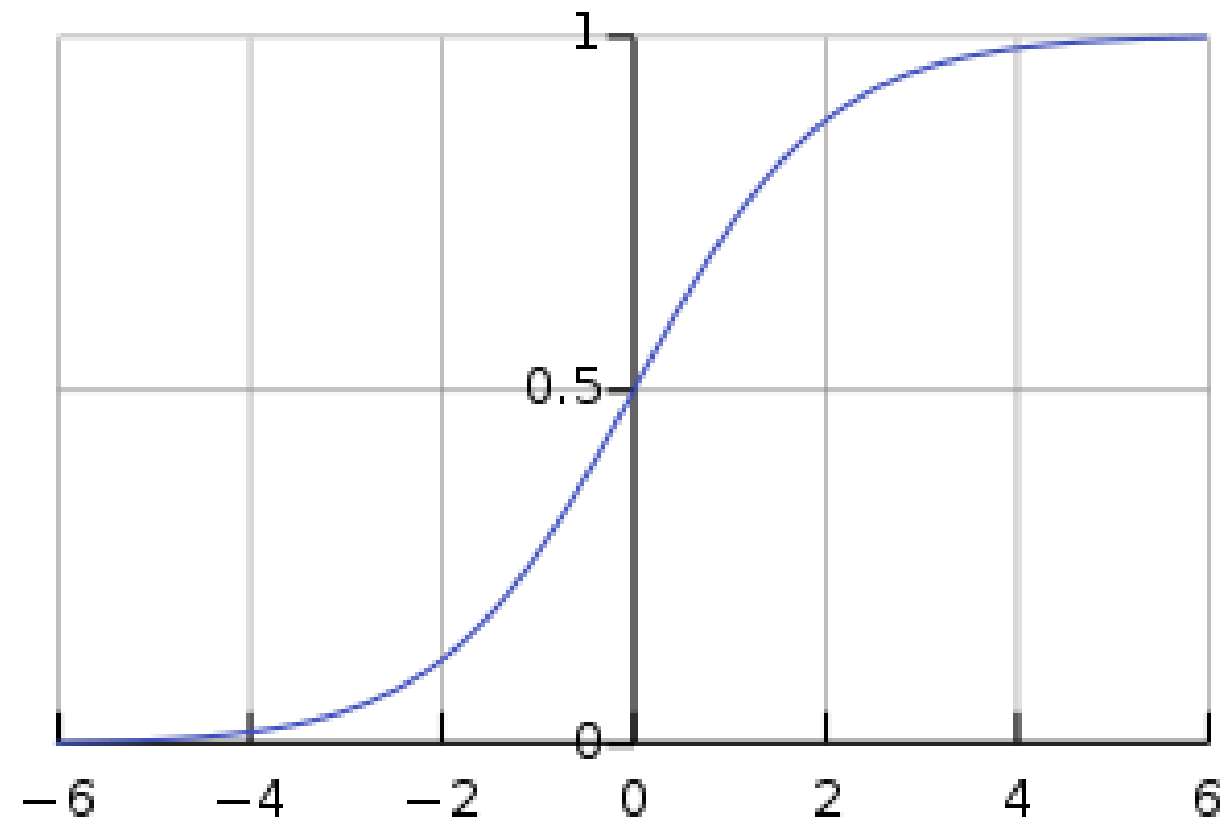
02 로지스틱모델

로지스틱모델이란?

로지스틱모델은 로지스틱회귀라고 불리며 영국 통계학자 콕스가 1958년 제안한 확률 모델로 변수들 간의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계기법입니다. 선형모델과 마찬가지로 입력변수와 결과변수의 사이가 선형적인 관계를 가지며, 입력변수들을 통해 결과변수를 예측하는 모델로 활용합니다.



02 로지스틱모델



하지만 선형모델의 경우 연속형 숫자값을 예측하는데 활용되지만 로지스틱모델은 결과변수가 범주형 데이터를 대상으로 범주 값으로 분류 예측하기 때문에 분류모델에 적합합니다.

특히 참과 거짓으로 결과변수가 라벨을 2가지만 가지는 이항형 문제에 주로 사용되며, 의료, 통신, 데이터마이닝과 같은 다양한 분야에서 분류 및 예측을 위한 모델로 사용되고 있습니다.

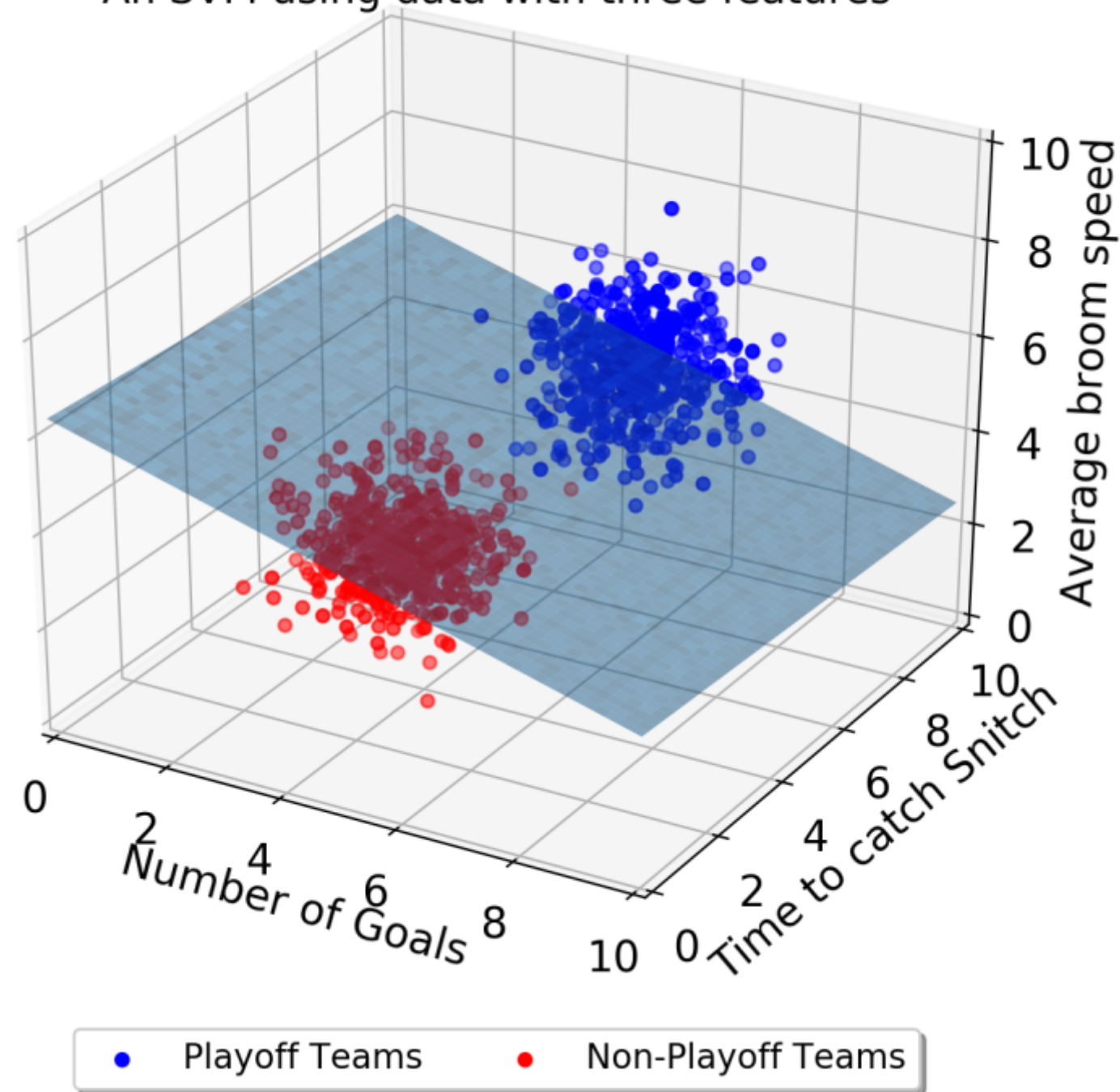
서포터벡터머신이란?

기계학습 분야 중 하나의 방법으로 패턴인식, 자료분석을 위한 지도학습 모델입니다.

주어진 입력변수 데이터들의 집합을 바탕으로 하여 새로운 데이터가 어느 범주에 속할 것인지 판단하는 비확률적인 선형 모델로 특히 참과 거짓으로 결과변수가 라벨을 2가지만 가지는 이항형 문제에 주로 사용됩니다. 기본적인 원리는 결과변수가 2가지 라벨을 가지고 있을 때, 각 라벨 그룹의 데이터들을 고차원의 벡터로 표현하고 그 중심 위치를 정하고 그 가운데에서 최적의 평면을 구함으로써 두 라벨을 나누는 방법으로 학습하게 됩니다. 중심 위치가 직선으로 나타날 경우 선형모델이며 아닐 경우 비선형 모델을 사용하게 됩니다.

03 서포터벡터머신

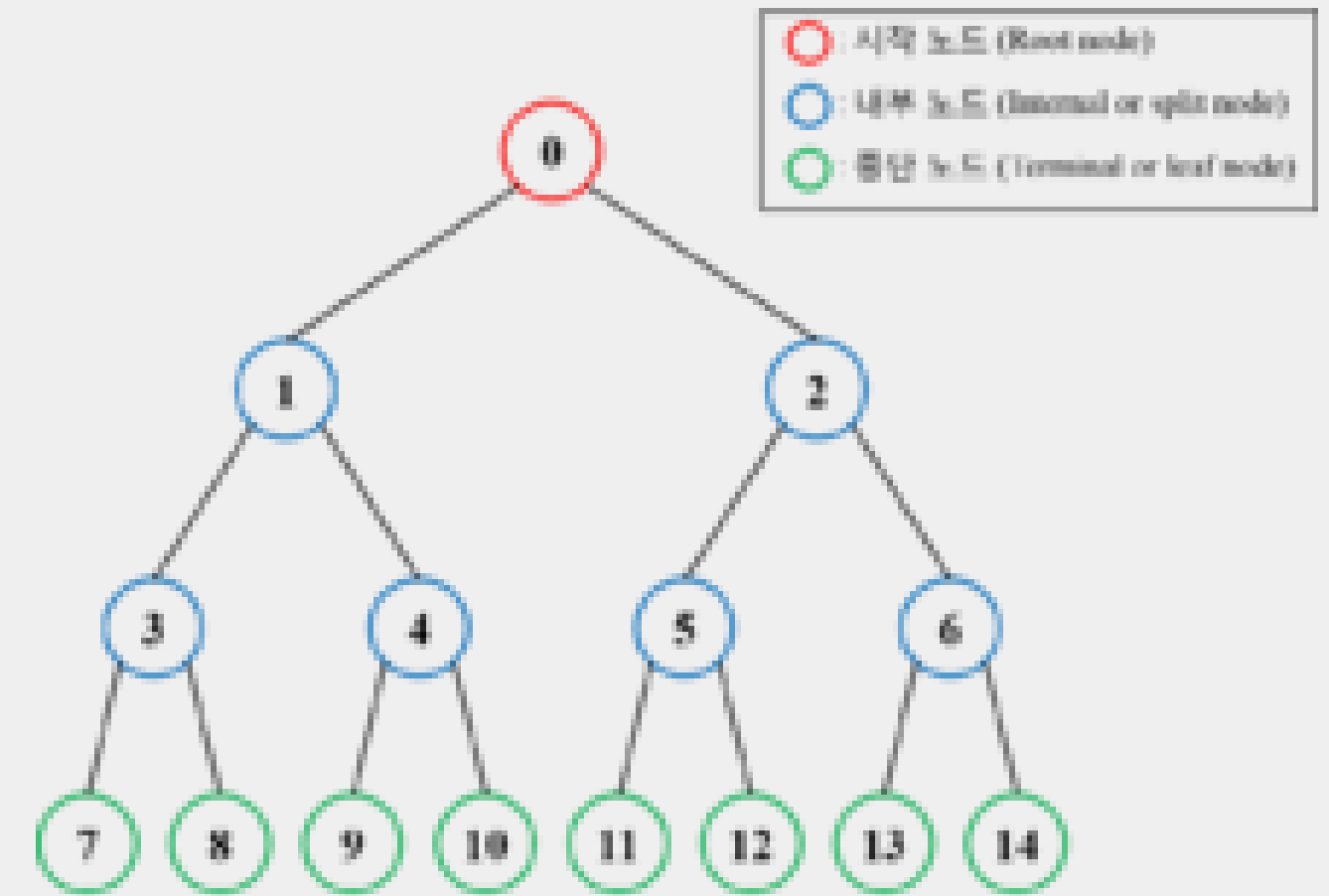
An SVM using data with three features



먼저 선형모델이나 로지스틱모델과는 결과 변수를 회귀 및 분류하는데 있어 고차원에서 분류하기 때문에 선이 아닌 면을 활용하는 것이 특징입니다. 이를 통해 이상값이 존재하는 데이터셋에서도 앞선 모델들에 비해 좋은 성능을 가질 수 있습니다. 서포터벡터머신은 이진 분류모델 이나 회귀모델 두 가지 모두에서 활용가능하며, 회귀모델은 서포터벡터 회귀라고 불립니다. 각 데이터들 간의 계산이 필요하기 때문에 시간이나 비용이 많이 들 수 있습니다.

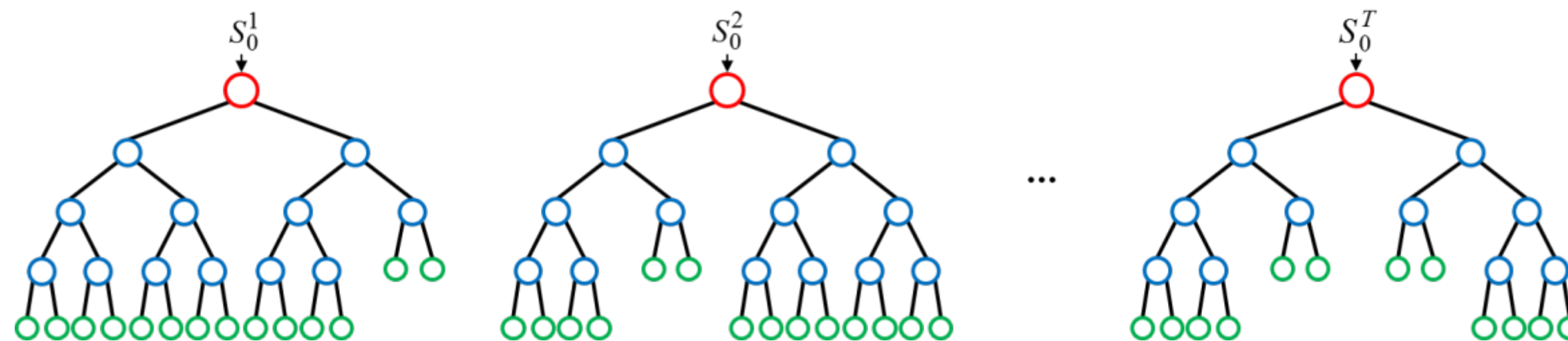
04 의사결정나무

의사결정나무란?



한 번에 하나씩 입력변수를 사용하여 예측 가능한 규칙들의 집합을 생성하는 기계학습 기법입니다. 데이터를 분석하여 이들 사이에 존재 하는 특징을 예측 가능한 규칙들의 조합으로 타내며, 그 모양이 나무가 가지가 나뉘는 것과 유사한 모양을 가지고 있기 때문에 의사결정나무라고 불립니다. 이는 질문을 던져서 대상을 좁혀나가는 스무고개 놀이와 비슷한 개념을 가지고 있습니다.

04 의사결정나무



의사결정나무는 서포터벡터머신과 마찬가지로 분류모델 이나 회귀모델 두 가지 모두에서 활용이 가능하며, 결과변수가 3개 이상의 라벨을 가질 수 있습니다.

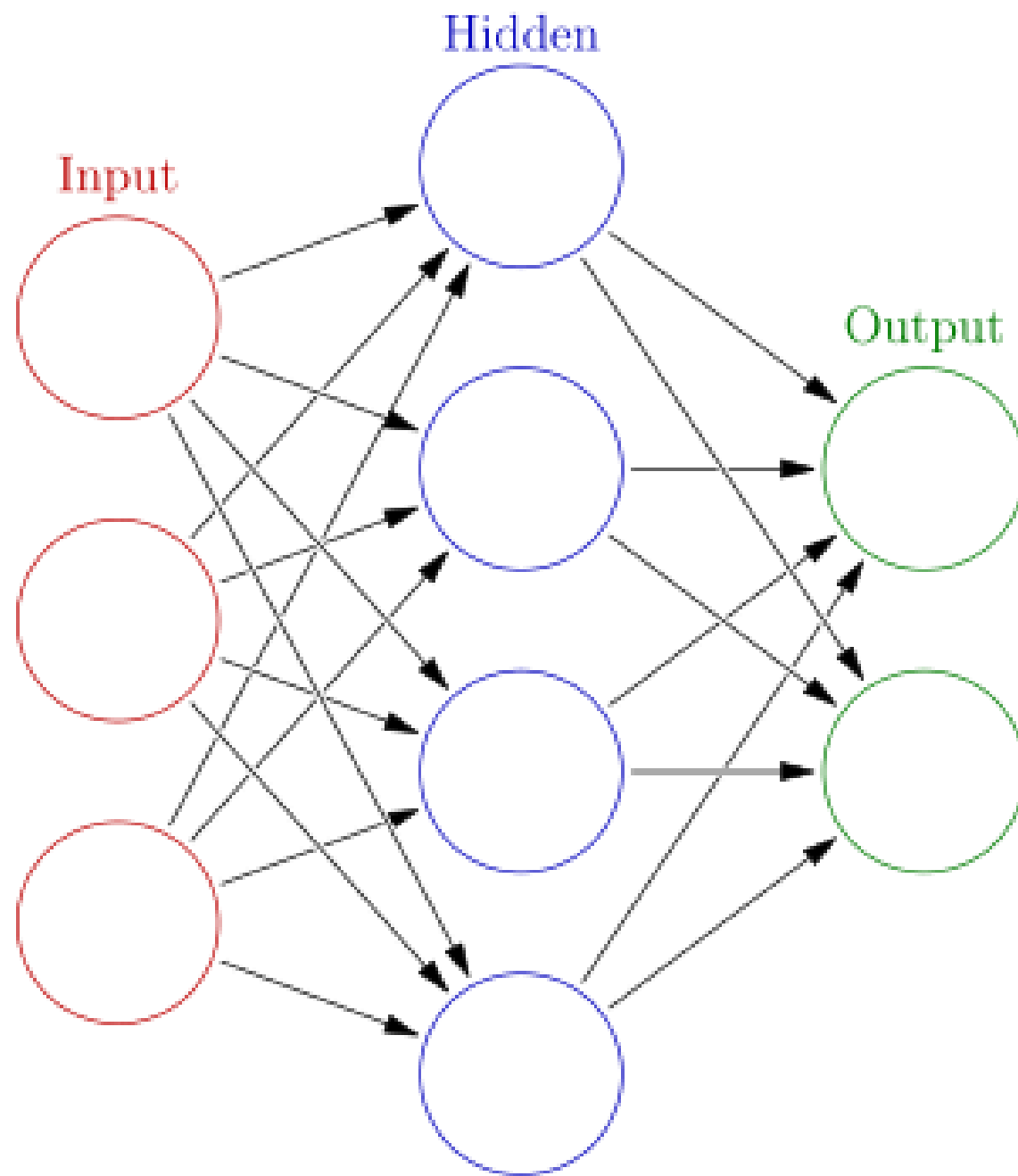
이러한 의사결정나무를 기반으로 한 랜덤포레스트는 각기 조금씩 다른 의사결정나무들을 여러 개를 가지며, 새로운 데이터가 들어오면 각 의사결정나무들의 결과들을 종합하여 투표를 통해 분류를 하거나, 평균을 통회 회귀하는 방식으로 모델이 구성됩니다.



신경망모델이란?

동물의 신경세포인 뉴런을 모방하여 만들어진 모델입니다. 뉴런의 동작원리와 같이 각 뉴런이 입력 신호를 받으면, 특정 자극을 통해 다음 뉴런으로 신호를 보내는 것과 같이 입력변수데이터에 가중치를 주어 다음 층으로 값을 전달하는 것입니다. 이러한 뉴런들이 쌓여서 최종적으로 결과변수를 예측하는 뉴런 단계에서 확률을 기반으로 입력변수들의 데이터를 통해 결과값을 예측하게 되는데 뇌가 신호를 판단하는 것과 유사하게 결정을 내리는 역할을 하게 됩니다.

05 신경망모델



뉴런에서 얼마의 가중치를 가지는 알 수 없으며, 학습단계에서 실제값과 예측값의 오차를 각 뉴런에 전달하여 가중치를 스스로 수정하여, 실제값과 예측값의 오차를 줄이도록 학습하게 됩니다. 일반적으로 뉴런이 많고 층이 높을수록 각 특징들을 분류하거나 회귀하는데 도움을 줄 수 있지만 너무 많은 경우에는 학습한 데이터와 거의 유사해야지만 정답으로 인정할 수도 있기 때문에, 모델을 설계할 때 뉴런의 수를 정하는 것이 중요합니다.

신경망 모델은 앞선 서포트벡터머신, 의사결정나무 기법들과 마찬가지로 분류 모델 이나 회귀모델 두 가지 모두에서 활용가능하며, 이미지 데이터, 음성 데이터 등과 같은 매우 복잡한 입력데이터를 가지고 있어 계산이 어려운 문제에서 유용하게 활용가능합니다.