

# 머신러닝 개념 입문

202102718 최승윤

# 목차

## 1. 선형회귀(다중선형회귀)

### 1.1 선형회귀

### 1.2 다중선형회귀

## 2. 로지스틱회귀

## 3. 의사결정나무

### 3.1 Data Mining 예측기법

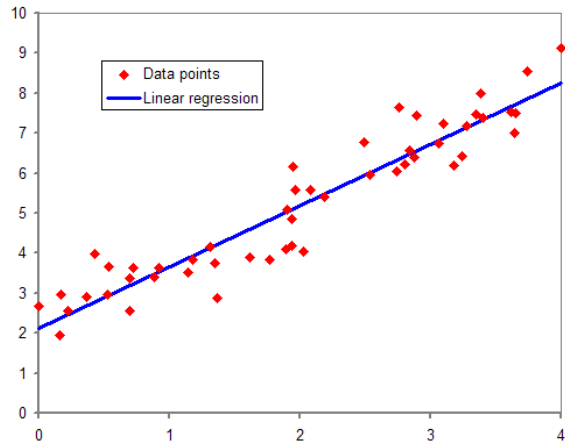
### 3.2 의사결정나무

## 4. 탐색적 데이터 분석

# 1. 선형회귀(다중선형회귀)

## 1.1 선형회귀

**선형 회귀**는 종속변수  $y$ 와 한 개 이상의 독립변수 (또는 설명 변수)  $X$ 와의 선형 상관 관계를 모델링하는 회귀분석 기법이다



선형 회귀의 예

# 1. 선형회귀(다중선형회귀)

## 1.2 다중선형회귀

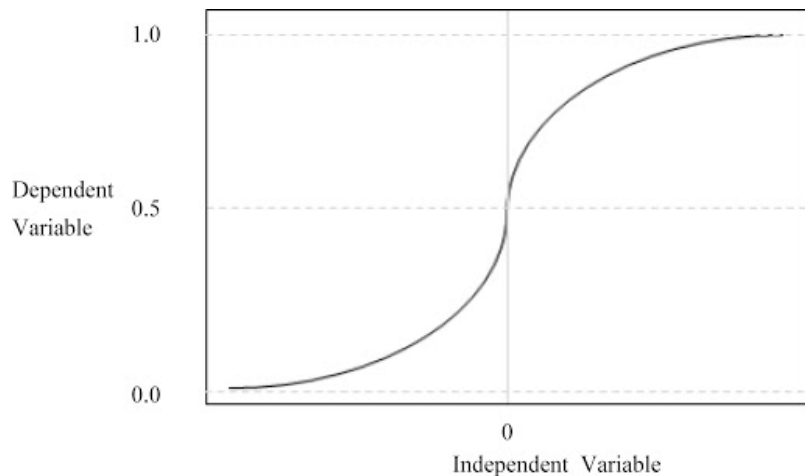
실제 자료들을 보다보면 종속변수  $y$  에 영향을 주는 독립변수  $x$  가 둘 이상인 경우가 대부분이다.  
드에 따라서 다중선형회귀 모델은 다음과 같은 형태를 갖는다.

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n,$$

## 2. 로지스틱회귀

로지스틱 회귀의 목적은 일반적인 종속 변수와 독립 변수간의 관계를 구체적인 함수로 나타내어 향후 예측 모델에 사용하는 것이다.

흔히 로지스틱 회귀는 종속변수가 이항형 문제(즉, 유효한 범주의 개수가 두개인 경우)를 지칭할 때 사용된다. 이외에, 두 개 이상의 범주를 가지는 문제가 대상인 경우엔 다항 로지스틱 회귀 또는 분화 로지스틱 회귀라고 하고 복수의 범주이면서 순서가 존재하면 서수 로지스틱 회귀라고 한다.



로지스틱회귀 함수의 예

### 3. 의사결정나무

#### 3.1 Data Mining 예측기법

데이터 마이닝 기법들은 목표변수가 존재여부에 따라 지도예측과 자율예측으로 분류되기도 한다

데이터 마이닝 기법들은 목표변수가 존재여부에 따라 지도예측과 자율예측으로 분류되기도 한다

지도예측(Supervised prediction)	자율예측(Unsupervised prediction)
<ul style="list-style-type: none"><li>■ 판별분석</li><li>■ 회귀분석</li><li>■ 의사결정나무분석</li><li>■ 신경망분석</li><li>■ 시계열분석(forecasting)</li></ul>	<ul style="list-style-type: none"><li>■ 군집분석</li></ul>

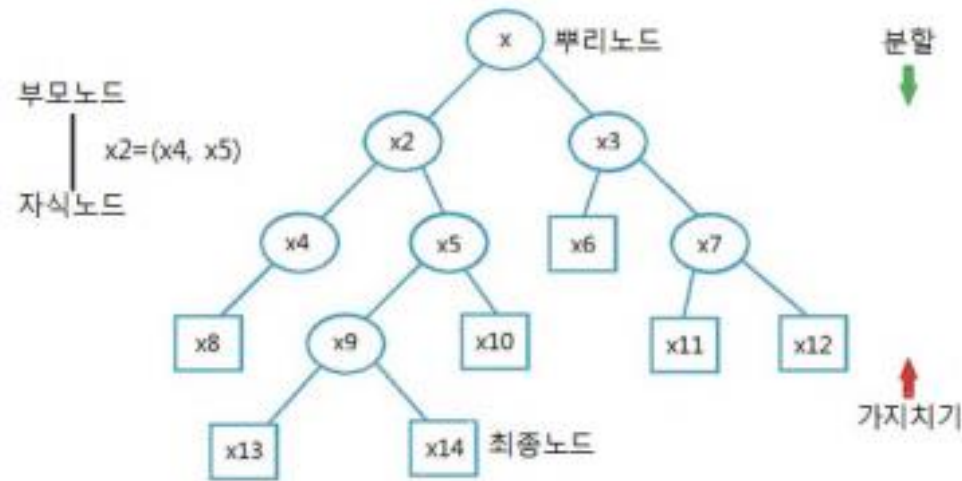
### 3. 의사결정나무

#### 3.2 의사결정나무

의사결정나무의 구조는 다음의 그림과 같다. 이 그림에서 맨 위의 마디를 뿌리노드(root node)라 하며, 이는 분류(또는 예측) 대상이 되는 모든 자료집단을 포함한다

상위 마디를 부모마디(parent node)라 하고, 하위 마디를 자식마디(child node)라 하며, 더 이상 분기되지 않는 마디를 최종노드(terminal node)라고 부른다.

가지분할(split)은 나무의 가지를 생성하는 과정을, 가지치기(pruning)는 생성된 가지를 잘라내어 모형을 단순화하는 과정을 말한다.



-의사결정나무는 목표변수가 이산형인 경우의 분류나무(classification tree)와 목표변수가 연 속형인 경우의 회귀나무(regression tree)로 구분된다

-목표변수가 이산형인 분류나무의 경우 상위노드에서 가지분할을 수행할 때, 분류(기준)변수와 분류 기준값의 선택 방법으로 카이제곱 통계량(Chi-square statistic)의 -값, 지니 지수(Gini index), 엔트로피 지수(entropy index) 등이 사용된다.



## 1.EDA란?

수집한 데이터가 들어왔을 때, 이를 다양한 각도에서 관찰하고 이해하는 과정이다. 한마디로 데이터를 분석하기 전에 그래프나 통계적인 방법으로 자료를 직관적으로 바라보는 과정

## 2. 이상값을 찾아내는 부분

### 1. 개별 데이터 관찰

### 2. 통계 값 활용

### 3. 시각화 활용

### 4. 머신러닝 기법 활용