

인공지능2 - 01 인공지능 알고리즘 분류와 회귀분석의 이해

AI Algorithm & Regression analysis

(주)씨에스리 조경미 기술사
(kmicho@cslee.co.kr)

다양한 인공지능 알고리즘 분류해 보자!

| | |
|---------|---|
| Concept | (목표) 인공지능을 구현하기 위한 분석 알고리즘의 활용 목적에 따라, 목표 변수에 따라 분류되는 체계를 이해하고, 가장 많이 활용되는 회귀분석의 유형을 이해한다 (활용 목적 기반) 데이터 분석의 활용 목적에 따라 예측모델과 탐색모델로 분류 (목표 변수 기반) 데이터의 목표 변수의 유무에 따라 지도/비지도 학습 모델로 분류 (회귀 분석 개념) 원인이 되는 값과 결과가 되는 값의 상관관계를 통계적 기법에 의해 조사하는 대표적인 분류 기법 |
| KeyWord | 데이터분석, 데이터마이닝, 학습(지도/비지도/강화), 예측모델, 탐색모델 회귀분석(단순회귀분석, 다중회귀분석, 로지스틱 회귀분석 등) |

1. 진화하다, 인공지능으로

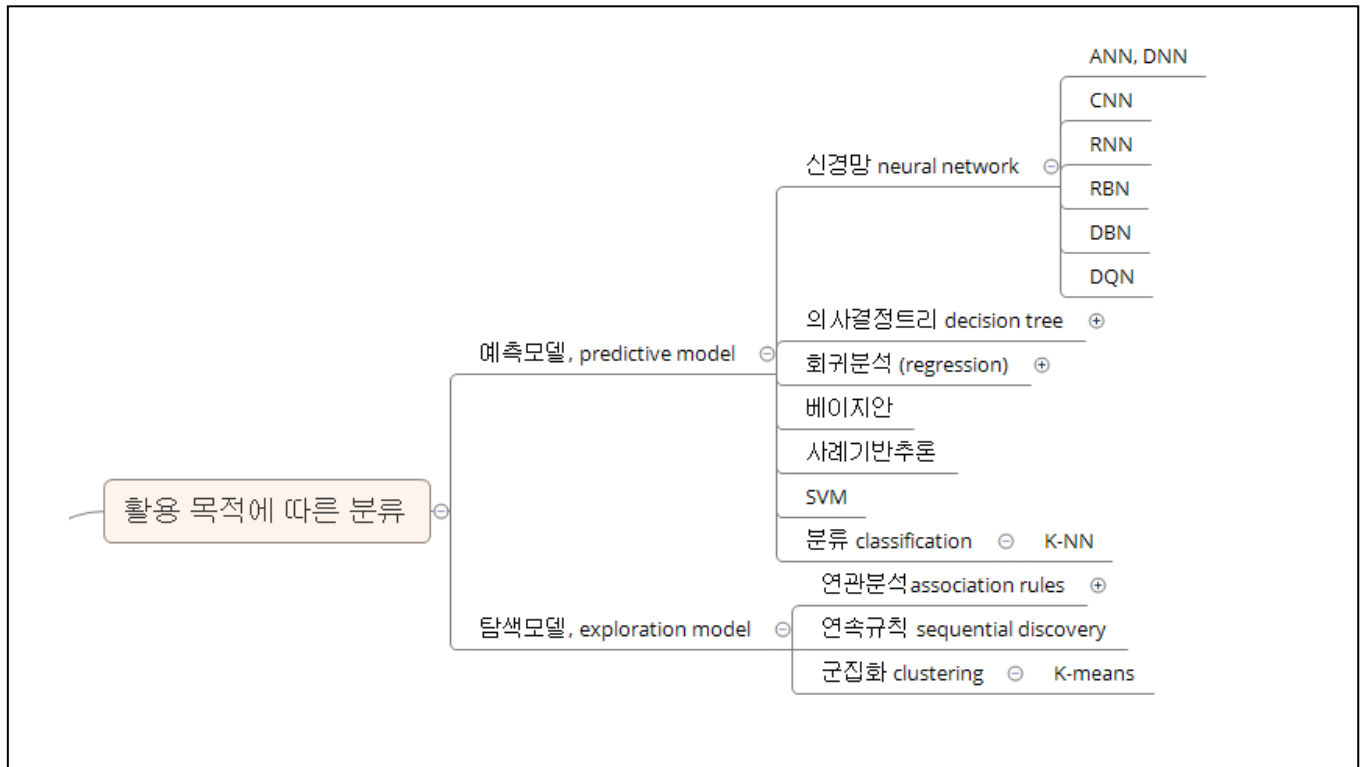
무엇이 인공지능으로 진화했는가? 데이터를 중심으로 어떤 의미를 찾아 내려고 하는 모든 활동이 인공지능으로 수렴하는 것 같다. 데이터를 분석하는 활동이, 데이터에서 의미를 찾는 데이터 마이닝이 신경망알고리즘을 통해 보다 강력한 인공지능으로 진화한다.

인공지능으로 진화하고 있는 데이터 분석, 데이터 마이닝, 데이터 학습 등의 의미가 혼선을 일으키는 것 같아, 개념과 차이점을 명확하게 짚어보고, 데이터 활용 목적에 따라 또는 목표변수의 유무에 따라 분석 기법들을 분류하여 정리 해 본다.

| 기법 | 개념 | Focus on |
|---------|--|----------|
| 데이터 분석 | -데이터로부터 의미 있는 패턴을 찾아내기 위해 모델(알고리즘)을 통한 궁극적인 문제해결 기법 -데이터 분석을 위해 데이터 마이닝 알고리즘을 분석 모델로 사용 | 문제 해결 |
| 데이터 마이닝 | -데이터 속에서 유용한 패턴, 새로운 의미 또는 지식을 찾아내는 기법 | 알고리즘 |

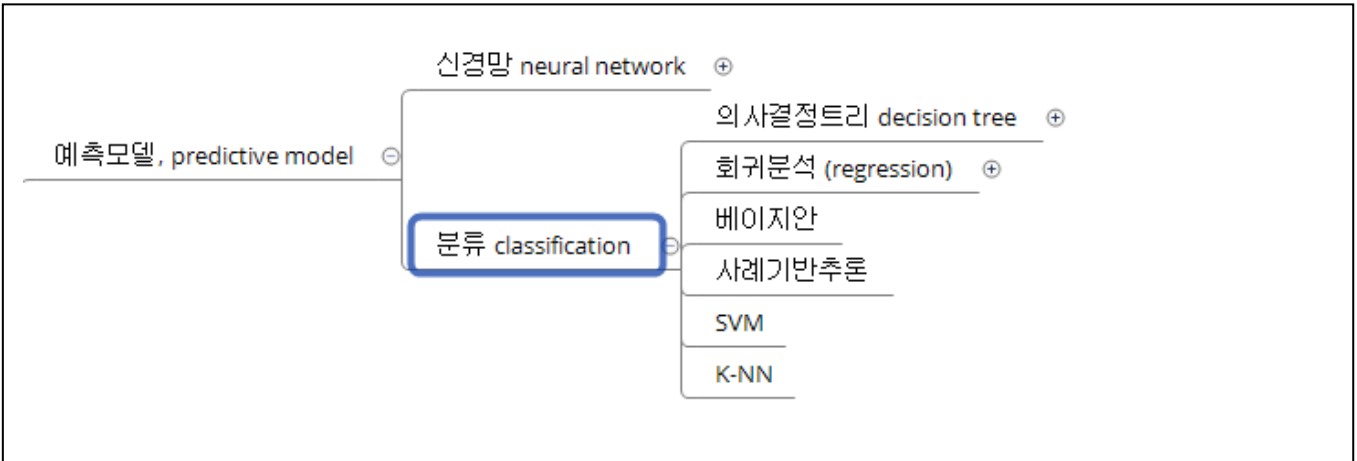
| | | |
|-------------------------------|--|------------------------------|
| <p>Advanced 데이터 분석</p> | <p>-데이터의 숨겨진 패턴을 찾아서 과거와 현재의 비즈니스 현상을 깊이 이해하고, 미래 상황을 예측, 선제적 의사결정을 지원하는 차세대 데이터 분석 기법</p> | <p>미래 예측 선제적 의사결정</p> |
| <p>학습기법</p> | <p>-학습의 결과가 표현되는 유무에 따라 지도학습, 비지도학습으로 구분됨 -지도학습: 원하는 결과가 표현된 학습 데이터 이용한 기계학습 -비지도학습: 원하는 결과가 표현되지 않은 학습 데이터를 이용한 기계학습 -강화학습: 지도학습과 비지도학습의 중간형태로, 입력에 대해 학습자가 행동을 선택하고 그 행동에 대해 보상을 제공하는 기계학습</p> | <p>기계학습</p> |

2. 활용 목적에 따른 데이터 마이닝 분류

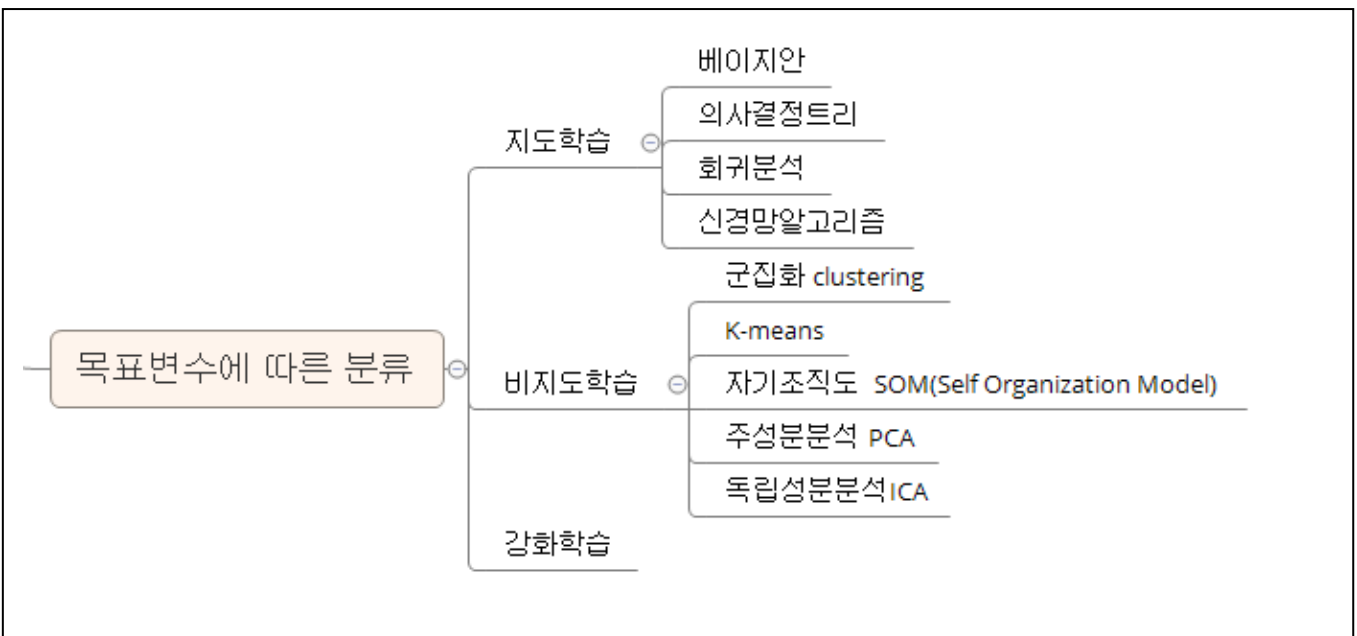


데이터 마이닝은 활용 목적에 따라 예측 모델과 탐색 모델로 구분된다. 예측 모델은 주어진 데이터에 근거한 모형을 이 모형을 이용하여 새로운 입력자료들에 대한 예측을 목적으로 하며, 탐색 모델은 주어진 데이터를 설명하는 패턴을 찾아내는 것이 주 목적이다. 탐색 모델은 서술형 모델(Descriptive Model)이라고 분류하기도 한다.

신경망 분석, 의사결정트리, 회귀분석, 베이지안 알고리즘, SVM 및 분류기법 등을 예측 모델이라고 할 수 있는데, 신경망 분석을 제외한 의사결정트리, 회귀분석, 베이지안 알고리즘 및 SVM 모두는 분류 기법의 한 종류라고 정의할 수도 있다.



3. 목표 변수에 따른 데이터 마이닝 분류



목표 변수가 정해져 있는지 여부에 따라 지도학습(Supervised Data)과 비지도학습(Unsupervised Data)로 분류되기도 한다. 대체로 지도학습은 예측 모델의 알고리즘들로 분류되며, 비지도학습은 탐색 모델의 알고리즘들로 구성이 된다.

4. 대표적인 분류기법 - 회귀분석

회귀분석(Regression)은 원인이 되는 값과 결과가 되는 값의 상관관계를 통계적 기법에 의해 조사하는 대표적인 분류 기법으로 하나 또는 둘 이상의 독립변수들을 기초로 하여 종속변수에 미치는 영향력의 크기를 알아보는 분석기법이다. 단순회귀분석, 다중회귀분석, 로지스틱 회귀분석 등 그 유형이 다양하다.

상관분석(Correlation)은 하나의 변수와 다른 변수와의 어떤 밀접한 관련성을 갖고 변화하는 가를 분석하는 방법으로 변수들간의 상호관계정도를 분석하는 통계적 기법으로 산점도를 통해 시각적으로 그 관계를 쉽게 확인할 수 있다.

| 회귀분석(Regression) | 상관분석(Correlation) |
|-----------------------------------|--------------------------------------|
| -독립변수가 종속변수에 미치는 영향력의 크기를 분석하는 기법 | -하나의 변수와 다른 변수와의 어떤 밀접한 관련성을 분석하는 기법 |

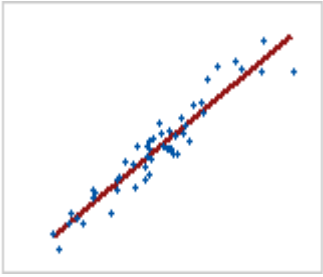
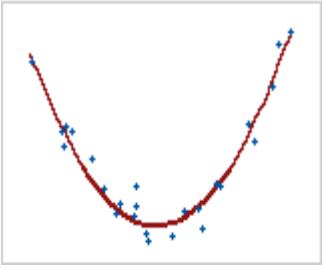
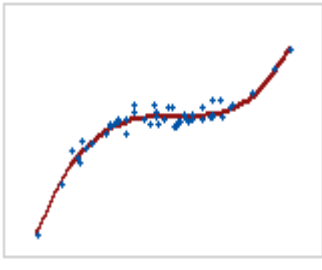
회귀분석의 정의에 사용되는 독립변수, 종속변수 등 변수의 유형들과 다양한 유사명칭들을 확인하고, 회귀분석을 표현하는 상관 관계의 다양한 형태와 그에 따른 회귀분석의 유형을 확인해 보자.

1) 변수의 이해

| 독립변수 | 종속변수 |
|-----------|----------|
| 설명변수 | 반응변수 |
| 예측변수 | 목표변수 |
| 방정식에서 X 값 | 방정식에서 Y값 |
| Input | output |

이제 어떤 이름으로 나오더라도 독립변수와 종속변수를 구분할 수 있다.

2) 관계의 이해

| 선형관계 | 비선형관계 | 단조관계 |
|---|---|---|
|  |  |  |

선형관계는 위와 같이 우측으로 가면서 높아지는 양의 선형관계와 반대로 우측으로 가면서 낮아지는 음의 선형관계가 있다. 단조관계는 독립변수와 종속변수가 동시에 증가하기는 하지만 같은 비율로 증가하지 않는다는 것을 확인할 수 있다.

회귀분석은 이와 같이 독립변수와 종속변수의 개수 또는 직선의 상관관계인지 곡선의 상관관계인지에 따라 다양한 유형으로 분류된다.

3) 회귀분석 유형의 이해

<독립변수 수에 따른 유형>

| 독립변수 수 | 유형 | 회귀식(사례) |
|---------------|--|---|
| 독립변수 1개 | 단순회귀분석 Simple Regression Analysis | $y = \alpha + \beta x + \varepsilon$ |
| 독립변수 2개 이상 | 다중회귀분석 Multiple Regression Analysis | $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$ |

* y=절편, B는 기울기로, 독립변수 1 개를 가지는 회귀식은 기울기가 있는 직선의 상관관계를 가진다.

<종속변수 수에 따른 유형>

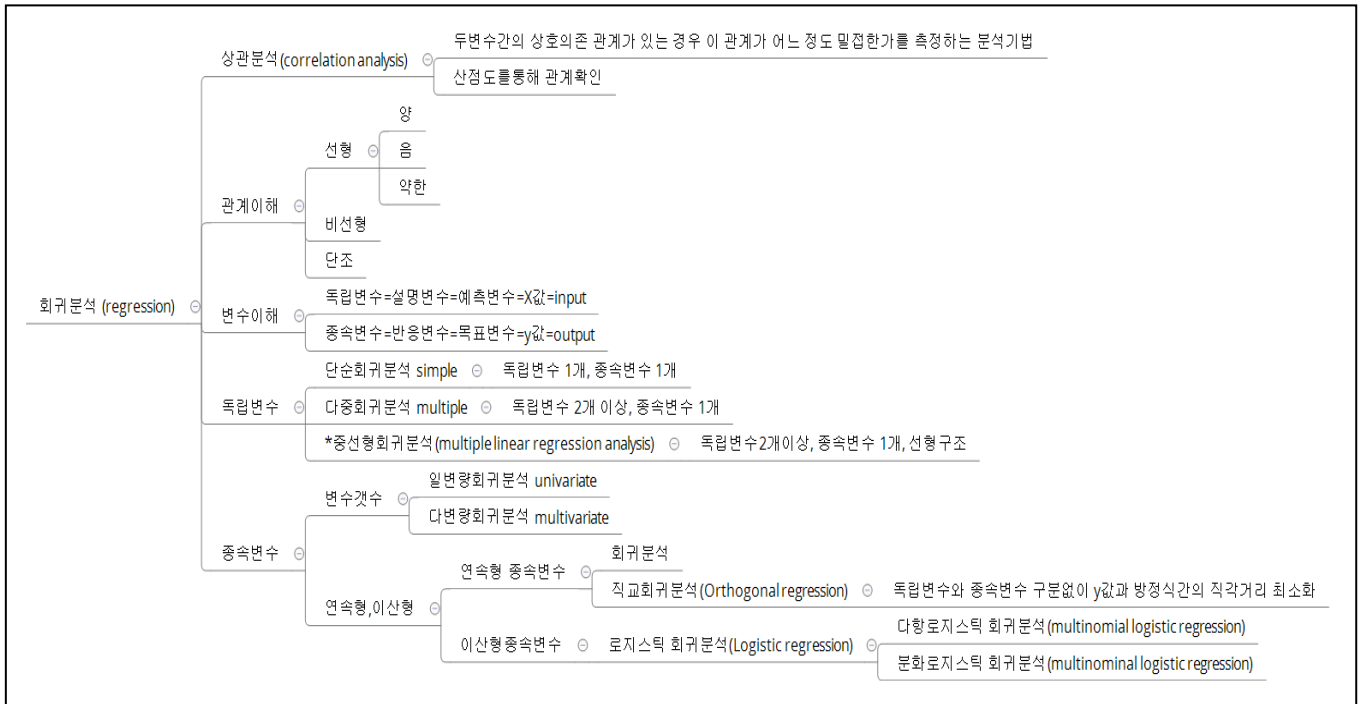
| 종속변수 수 | 유형 |
|---------------|---|
| 종속변수 1개 | 일변량 회귀분석 Univariate Regression Analysis |
| 종속변수 2개 이상 | 다변량 회귀분석 Multivariate Regression Analysis |

<종속변수의 형태에 따른 유형>

| 이산형 | 연속형 |
|--|---|
| 종속변수 Y가 이산형 변수 | 종속변수 Y가 확률값을 갖는 연속형 변수 |
| 회귀분석(Regression Analysis) 직교회귀분석 (Orthogonal Regression Analysis) | 로지스틱 회귀분석 Logistic Regression Analysis |
| | |

- 직교회귀분석은 종속변수 Y 값과 방정식간의 직각거리를 최소화하는 분석기법 이며, 로지스틱 회귀분석은 종속변수가 2 개 이상인 다항 로지스틱 회귀분석과 분화 로지스틱 회귀분석으로 분류된다.

[참고] 회귀분석 관련 용어, 변수 및 그에 따른 유형의 이해를 위한 마인드맵



“끝”

Contents connect communications!!

아이리포에 오시면 더 많은 지식을 가져가실 수 있습니다.

아이리포 온라인 : <http://www.ilifo.co.kr>

아이리포 지덤시리즈 : <http://www.jidum.com>

아이리포 IT지식창고 : <https://www.ilifo.co.kr/boards/knowledge>

아이리포 기술사/감리사 카페 : <http://cafe.naver.com/itlf>

서울시 마포구 상암동 1610번지, DDMC 3층 아이리포 교육센터

TEL: 02-303-9997 | MAIL: edu@ilifo.co.kr