

<용어체크>

회귀 분석

지도 학습의 하나로 학습 데이터에 부합되는 출력 값이 실수인 함수를 찾는 문제로 가설로 설정한 함수의 파라미터를 추정하는 방법이다. 학습규칙으로는 경사하강법을 사용하며 경사하강법은 반복에 의해 오차함수가 최소가 되는 파라미터를 근사화하는 방법이다. 오차 함수로는 평균제곱오차 함수를 많이 사용한다.

서포트 벡터 머신

기계 학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 두 카테고리 중 어느 하나에 속한 데이터의 집합이 주어졌을 때, SVM 알고리즘은 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 카테고리에 속할지 판단하는 비확률적 이진 선형 분류 모델을 생성한다.

의사 결정 트리

어떤 항목에 대한 관측값과 목표값을 연결시켜주는 예측 모델로 지도학습의 분류 알고리즘으로 사용되는 방법이다. 분류 목적이 범주형일 경우 ID3와 같은 알고리즘이 사용되며 그 외에 수치형일 경우 CHAID나 CART 와 같은 방법을 사용한다. 분류 속성을 결정하기 위해 정보이득, 정보이득비, 지니 계수 등을 사용한다.

<학습내용>

회귀 분석

서포트 벡터머신

의사결정 트리

<학습목표>

회귀 분석을 통한 분류와 회귀에 대해 설명할 수 있다.

서포트 벡터 머신을 이용한 분류에 대해 설명할 수 있다.

의사결정 트리의 종류와 활용방법을 설명할 수 있다.

Q. 지도학습 알고리즘으로 많이 사용되는 회귀 모델은 어떤 특징이 있을까요?

: 회귀 모델은 학습 데이터에 따른 결과값을 예측하기 위해 사용되는 모델이며 사용되는 가설 함수에 따라 선형 회귀 또는 비 선형 회귀 모델이 있습니다. 선형 회귀의 경우 학습 데이터로 사용되는 데이터의 차원이 1차원이나 2차원 형태로 구성되며 경사하강법 등을 이용하면 간단하게 학습 모델을 구성할 수 있는 장점은 있지만 오차 함수를 어떻게 설정하는지에 따라 정확도가 차이가 날 수 있습니다. 경사하강법 또한 반복에 의한 파라미터 근사화 방법이다보니 정확도 부분에 오차를 포함하고 있습니다. 때로는 반복에 의한 추정 방법보다 간단한 가설 함수로 구성되었다면 해석해를 수치해석으로 구하면 조금 더 정확도를 올릴 수 있습니다.

회귀

- ▶ 회귀 분석은 지도 학습의 하나로 학습 데이터에 부합되는 출력 값이 실수인 함수를 찾는 문제이다.
- ▶ 가설 함수에서 시작하여 함수의 파라미터를 학습 데이터를 이용하여 추정하는 지도 학습 방법이다.
- ▶ 경사 하강법을 이용한 학습 규칙을 많이 사용한다.
- ▶ 오차 함수로 평균제곱오차 함수를 많이 사용한다.

서포트 벡터 머신

- ▶ 데이터 분포를 나누는 기준을 결정하는 지도 학습 모델 중 하나이다.
- ▶ 어떤 패턴으로 데이터를 분류한 후 데이터 사이의 거리에 따라 어떤 카테고리에 속할 것인지 판단한다.
- ▶ 서포트 벡터 머신의 손실 함수는 경첩(hinge) 모양을 나타내므로 경첩 함수 또는 경첩 손실 함수라고도 한다.

의사결정트리

- ▶ 정답 데이터를 이용해 결정 트리를 만드는 알고리즘의 하나로 ID3를 사용한다.
- ▶ 의사 결정 트리를 기반으로 모든 데이터를 제대로 분류할 때까지 노드를 추가한다.
- ▶ 정보이득 또는 정보이득비, 지니 계수 등을 이용하여 데이터의 속성을 선택하고 분류한다.
- ▶ 분류 효율성과 의사 결정 트리의 일반성을 고려해 최대한 단순한 형태가 되는 것이 목표이다.