4주차

기계학습

1. 지도학습

* 분류
* 회귀

1. 비지도학습
2. 강화학습

선형회귀분석

회귀분석: 관계성을 나타내는 함수를 학습하는 기법

최소제곱법

* 오차를 최소화하는 직선을 찾아내는 방법
* y=ax+b 식을 이용
* a와 b를 구하는 방법

(1) 오차의 총합을 a와 b에 대해서 각각 미분

\*오차의 총합식은 음수를 제거하기 위해 제곱을 이용한 이차식으로 그래프의 꼭짓점에 해당하는 a와 b의 값을 구해야 함. 이를 위해 미분을 활용함.

(2) 기울기 y가 0이 되는 2개의 식을 찾기

(3) a와 b를 구함

다중회귀분석

* 여러 개의 값을 이용하여 하나의 수치를 예측
* z=ax+by+c 식을 이용
* a, b, c 값을 구하는 방법

🡪 반복법: 적당한 값에서 출발하여 올바른 값에 가까워지도록 조금씩 값을 갱신해가는 방법

* 경사하강법: 현재의 기울기를 이용하여 값을 갱신해 나가는 반복법

\*합성함수의 미분법

* y=(x+1)^2과 같은 식의 미분에 z=x+1과 같은 식을 대입(치환)
* 식을 간단히 풀 수 있도록 함
* x로 y를 구할 수 있을 때 y의 x에 의한 미분은 y 방향의 변화량을 x 방향의 변화량으로 나눈 것

🡺 δy/δx = δy/δz \* δz/δx

분류

* 최소제곱법: z=ax+by+c 식의 결과 값에 기대값을 더한 오차의 제곱의 총합을 각 요소로 미분하는 방법
* 경사하강법: z=ax+by+c 식의 값이 어떠한 기대값에 더 가까운지 확인하여 예측하는 방법

결정 트리 학습

* 혼잡도(엔트로피)를 줄여 나가는 과정
* 순서

1. 현 상태의 혼잡도를 계산
2. 데이터를 어떠한 조건으로 분류했을 때 전체 혼잡도가 많이 개선되는지 계산

\*전체 엔트로피 계산 🡪 각 특징 별 엔트로피를 구하여 엔트로피 기대치를 계산 🡪 전체 엔트로피에서 각 특징의 엔트로피를 빼서 엔트로피 개선 값을 계산

1. 해당 조건을 찾아 정리 및 반복 적용

* 엔트로피 계산법

엔트로피 = -p(현상 1) \* log2(p(현상 1)

-p(현상 2) \* log2(p(현상 2)

…

-p(현상 N) \* log2(p(현상 N)

* 관점에 따른 분류: 특징 별 엔트로피 개선 값이 큰 정도에 따라 순차적으로 분류
* 분류 방식이 규칙 형식으로 학습되어 학습 결과로 인간도 알기 쉬운 판정 기준을 얻을 수 있음