

머신러닝

- 데이터에서부터 학습하도록 컴퓨터를 프로그래밍하는 과학
- 데이터를 가지고 학습 => 프로그래밍과 다름
- 사람이 직접 규칙을 만들지 않고 학습을 하는 것.

컴퓨터 프로그램은 작업 T와 성능 측정 P에 대해 경험 E로 학습한 것

<의미>

<4조 의견>

1. 작업(T, Task): 프로그램이 수행하는 작업 또는 일련의 작업

Ex) 이미지 분류, 음성 인식, 자연어 처리 등

2. 성능 측정(P, Performance Measure): 작업의 성과를 측정하는 방법이며, 얼마나 잘 수행되었는지를 나타냄,

성능 측정은 목표를 달성하기 위한 지표

3. 경험(E, Experience): 프로그램이 작업을 수행하면서 얻는 데이터 또는 정보를 의미

이 경험을 바탕으로 프로그램은 작업을 더 잘 수행할 수 있게 됨

이러한 요소들을 통해 컴퓨터 프로그램은 작업을 수행하고, 그 결과를 평가하여 경험을 통해 성능을 향상시킴

,

학습이란?

훈련 데이터로부터 가중치 매개변수의 최적 값을 자동으로 획득하는 것

AND 퍼셉트론: 입력이 모두 참일때만 출력이 참이 되는 논리 연산

※ 퍼셉트론 출력

출력 = w1 * 입력1 + w2 * 입력2 + b (w1, w2는 각 입력에 대한 가중치(weights)이며, b는 편향(bias)을 의미)

<4조 의견>

1. w1 = 0.5, w2 = 0.5, b = -0.7로 설정

2. 입력이 (0, 0) 일 때의 출력,

출력 = 0.5 * 0 + 0.5 * 0 - 0.7 = -0.7 이므로 출력은 0

나머지 입력 조합에 대해서도 출력을 계산해보면 AND 게이트의 동작을 잘 따르는 것 확인 가능

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

베이즈 정리: 증거(사실)가 발생했을 때, 가설이 일어날 확률

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

평균 제곱 오차: 손실함수, 현재 신경망이 훈련 데이터를 얼마나 잘 처리하지 못하느냐를 나타냄

$$\frac{d}{dx} f(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

미분: 특정 순간의 변화량

$$\begin{array}{l} x_1 + x_2 + x_3 = 3 \\ x_1 - x_2 + 2x_3 = 2 \\ 2x_1 + 3x_3 = 5 \end{array} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$$

선형대수학: 행렬, 벡터, 가중치 매개변수

<4조 의견>

Q. 머신러닝 학습이란?

종합의견: 전체 데이터셋에서 훈련데이터의 범위를 지정하고, 가중치 및 파라미터의 최적 값을 자동으로 획득하는 것, 학습 데이터를 일반화하는데 꼭 필요한 요소라 생각

<4조 의견>

모델 훈련과정을 생각하고 코드를 해석하여 서로 의견을 주고받음

먼저, 데이터를 다운로드하여 준비하고 데이터의 특정 컬럼을 각각 x와 y에 설정을 해줌

설정된 x와 y의 값을 간단하게 시각화 해보고, 모델을 학습시키기 위해 LinearRegression을 모델로 선정하여 훈련시킴. 이 훈련된 모델을 가지고 키프로스 1인당 GDP를 예측하였을 때 약 5.9624가 나오는 것을 확인함