

머신러닝

머신러닝이란?
<ul style="list-style-type: none"> - 학습을 통해 자동으로 개선하는 알고리즘. - 인공지능의 하위 집합 - 학습과 개선을 위해 명시적으로 컴퓨터를 프로그래밍하는 대신, 컴퓨터가 데이터로 학습하고 경험을 통해 개선하도록 훈련하는데 중점을 둠. - 대규모 데이터 세트에서 패턴과 상관관계를 찾고 분석을 토대로 최적의 의사결정과 예측을 수행하도록 훈련됨. - AI > 머신러닝 > 딥러닝 > 신경망 (AI의 하위 집합) - 딥러닝의 머신러닝의 하위 집합으로, 사람의 도움 없이도 데이터를 지속적으로 분석하여 정확한 결정을 내릴 수 있도록 해주는 인공 신경망 알고리즘을 사용함.
Linear Regression_Linear Regression
<ul style="list-style-type: none"> - 지도 학습 알고리즘으로, 주로 수치 예측 문제에 사용함. - 여러 개의 독립 변수와 종속 변수의 관계를 선형식을 이용하여 설명함. - 점들이 퍼져있는 형태에서 패턴을 찾아내고, 이 패턴을 활용해서 예측하는 분석. - 새로운 표본을 뽑았을 때 돌아가려는 특징이 있기 때문에 붙은 이름. - 2차원 좌표에 분포된 데이터를 1차원 방정식을 통해 표현되지 않은 데이터를 예측하기 위한 모델 - 분류 문제를 해결할 수 있지만, 예외적인 데이터에 민감하게 반응하여 결과값이 달라질 수 있어 권장하지 않음.
Linear Regression_Hypothesis(가설함수)
<ul style="list-style-type: none"> - 가설을 세우는 함수로, 1차 방정식을 세워 예측 값과 가장 잘 맞는 선을 찾을 수 있도록 함. - Linear Regression에서 사용하는 1차원 방정식을 가리키는 용어로, 우리말로로는 가설이라고 하며, 수식에서는 $h(x)$ 또는 $H(x)$로 표현함. - $H(x) = Wx + b$에서 x에 대한 1차방정식이며, 기울기(W)와 절편(b)이 반복되는 과정에서 계속 바뀌고, 마지막 루프에서 바뀐 최종 값을 사용해서 데이터 예측에 사용하게 됨.
Linear Regression_Cost
<ul style="list-style-type: none"> - Hypothesis 방정식과의 차이로, 방정식의 결과가 크게 나오면 좋지 않다고 얘기하고, 루프를 돌 때마다 W와 b를 차이가 작은 방향으로 수정하게 됨.
Linear Regression_Cost Function
<ul style="list-style-type: none"> - 원래의 값과 가장 오차가 작은 Hypothesis를 도출하기 위해 사용되는 함수 - Hypothesis의 형태를 결정 짓는 것은 매개변수(parameter)라 부르는 θ임. 이 θ 값을 적절하게 조정하여 실제 값 y에 가장 근접한 Hypothesis를 training set를 이용하여 도출함.
Linear Regression_Cost Function, Objective Function, Loss Function 차이
<ul style="list-style-type: none"> - Objective function: 학습을 통해 최적화 시키려는 함수. 딥러닝에서는 일반적으로 학습을 통해 Cost를 최소화시키는 optimize 작업을 수행하고 이 때 Cost function을 objective function이라 볼 수 있음. - Cost function: 모든 input dataset에 대해서 원래의 값과 가장 오차가 작은 Hypothesis를 도출하기 위해 사용되는 함수 - Loss function: input(x)에 대한 예측값과 실제 label값 사이의 오차를 계산하는 함수임. 예로 linear regression의 경우 최소제곱오차를 Loss function으로 사용함. 하나의 input data에 대해서 오차를 계산하는 함수임.
Gradient Descent
<ul style="list-style-type: none"> - 순간 기울기를 구한 후 오차가 가장 작은 지점을 찾아야 함. 이 때는 미분 시, 기울기가 0이 되는 지점이고, 학습을 만큼 경사를 이동하여 기울기를 변화시키면서 기울기가 0이 되는 지점을 찾아 오차를 가장 작게 함.
왜 Gradient Descent에서는 기울기를 사용하는가?
<ul style="list-style-type: none"> - 각 변수로의 일차 편미분 값으로 구성되는 벡터임. 이 벡터는 f의 값이 가장 가파르게 증가하는 방향을 나타내며, 벡터의 크기는 증가의 가파른 정도를 나타냄. - Gradient의 특성을 이용하여 gradient의 반대 방향으로 조금씩 이동시키면 $f(x)$가 극소가 되는 지점을 찾을 수 있음.
지역 극소점 Local Optima에 빠지는 문제
<ul style="list-style-type: none"> - 에러를 최소화 시키는 최적의 파라미터를 찾는 문제에 있어서 파라미터 공간에 수많은 지역적인 홀(hole)들이 존재하여 최소점이 아닌데도 최소점이라고 인식하는 문제. 하지만 이 문제는 고차원의 공간에서 발생하기 힘든 경

우로 여겨지고, 대부분 경사가 완만한 saddle point로 여겨짐.
Bias란 무엇인가
- 예측값과 실제값이 떨어져 있는 정도 (편차)
Variance란 무엇인가?
- 예측값이 퍼져있는 정도
머신러닝에서 bias와 variance의 관계
- Bias는 데이터의 분포에 비하여 모델이 너무 간단한 경우 과소적합이 발생함.
- Variance 는 모델의 복잡도가 데이터 분포보다 클 때 데이터를 과적합 시키는 문제가 발생함.
ROC Curve
- 다양한 threshold에 대한 이진 분류기의 성능을 한 번에 표시한 것
- FPR, TPR 두가지를 이용해서 표현하며, 좌상단에 붙어 있을수록 좋은 이진 분류기임.
지도학습
- 정답 레이블을 가진 데이터에 대해 학습하는 방법.
- 일반적으로 데이터가 label을 가지기 어려워 구할 수 있는 데이터의 양이 적음.
- 크게 분류와 회귀가 있음.
- 분류: 어떤 데이터에 대해 값을 분류하는 것
- 회귀: 특징을 토대로 값을 예측하는 것. 결과 값은 실수 값을 가질 수 있음.
비지도 학습
- 정답 label이 없는 데이터에 대해 학습하는 방법
- 비슷한 데이터들을 군집화 함.
두 데이터의 유사도를 측정하는 방법
- Euclidean Distance: 두 점 사이의 유클리디안 거리를 구하면 최단 거리가 됨.
- Cosine Similarity: 내적 공간의 두 벡터간 각도와 코사인 값을 이용하여 측정된 벡터간 유사한 정도. 코사인 유사도는 '방향'에 대한 유사도임. 거리는 고려하지 않으며, 2차원보다 높은 차원의 데이터, 벡터의 크기가 중요하지 않은 데이터에 주로 사용됨.
MLE
- Maximum Likelihood Estimation의 약자고, 최대우도법 또는 최대가능도라고 함.
- 관측값이 있을 때, 그 값들을 토대로 θ 를 추정하는 방법.
- Likelihood: 우도 또는 가능도라 불리며, 가설 θ 와 시행 결과 D가 주어졌을 때 θ 가 참이라면, 결과 D가 나올 확률. D라는 결과가 나왔을 때, 이 결과가 나올 수 있는 여러 가능한 가설들(θ)을 평가하는 척도