1. 머신러닝 정의

명시적인 프로그래밍 없이 컴퓨터가 학습하는 능력을 갖추게 하는 것

1. 머신러닝 문제유형
2. 레이블된 훈련 세트?

지도학습

1. 가장 널리 사용하는 지도학습 방법 2가지

분류, 예측

1. 보편적인 비지도 학습 작업 4가지

군집, 시각화, 차원축소, 연관규칙학습

1. 사전 정보에 없는 여러 지형에서 로봇을 걸어가게 하면 어떤 종류의 머신러닝 알고리즘?

강화학습

1. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘 사용?

군집

1. 스팸감지의 문제는 지도? 비지도?

분류

1. 온라인 학습 시스템이 무엇인가?

데이터를 순차적으로 한 개씩 또는 미니 배치(작음묶음단위)로 주입하여 시스템을 훈련시킨다.

1. 외부 메모리 학습?

컴퓨터 한 대의 메인 메모리에 들어갈 수 없는 아주 큰 데이터셋을 학습하는 시스템

1. 예측을 하기위해 유사도 측정에 의존하는 학습 알고리즘은?

회귀분석?

1. 모델 파라미터와 학습 알고리즘의 하이퍼파라미터 사이의 차이

모델 파라미터는 데이터에 가장 잘 맞는 파라미터를 찾는다(변함). 하이퍼파라미터란 학습 알고리즘으로부터 영향을 받지 않으며 훈련 전에 미리 지정되고, 훈련하는 동안에는 상수로 남아있다.

1. 모델 기반 알고리즘이 찾는 것은 무엇? 성공을 위해 이 알고리즘이 사용하는 일반적인 전략? 예측은 어떻게 만듬?

샘플들의 모델을 만들어 예측에 사용

1. 머신러닝의 주요 도전 과제는?

나쁜 알고리즘과 나쁜 데이터

1. 모델이 훈련 데이터에서의 성능은 좋지만 새로운 샘플에서의 일반화 성능이 나쁘다면 어떤 문제? 어떻게 해결?

오버피팅(과적합)

1. 테스트 세트 무엇? 왜 사용?

훈련데이터를 훈련세트와 테스트 세트로 나눔, 훈련세트를 사용해 모델을 훈련시킨 후 테스트 세트를 사용해 모델을 테스트 한다.

1. 검증 세트의 목적?

모델이 새로운 데이터에 잘 작동하지 않을 수 있는데 이를 해결하기 위해 사용

1. 테스트 세트를 이용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나?

모델과 하이퍼파라미터가 테스트 세트에 최적화된 모델을 만듬

1. 교차 검증이란? 왜 하나의 검증 세트보다 선호?

훈련 세트를 여러 서브셋으로 나누고 각 모델을 이 서브셋의 조합으로 훈련시키고 나머지 부분으로 검증한다. 검증 세트로 훈련데이터에서 너무 많은 양의 데이터를 뺏김