대학원 산업인공지능학과 '어프렌티스 프로젝트'

강의 소감문

이름	방창현	학번	2021254011

강연 제목	AI 기만 공격 기술 (어프렌티스프로젝트)
강사명	박호성 교수님
강연 일자	2021년 12월 09일

1. 강의 요약 및 소감: (문장식으로 작성, 조리 있는 글쓰기 연습의 기회)

1. AI학습
   1) 학습단계 : 다량의 학습 데이터를 이용하여 머신러닝을 돌려 스스로 패턴 분석을 시행한다.
   2) 활용단계 : unseen 데이터를 학습된 머신러닝에 적용하여 결과를 예측한다.
   3) 대표적인 AI 서비스 : 객체인식, 자연어 처리, 상황인식
2. AI와 보안
   1) 보안관점 : AI로 인한 의도치 않은 사고 방지, AI에 대한 공격을 방어, AI를 활용한 보안을 들 수 있다.
   2) 사고방지 : 디자이너의 의도와 다른 유해 행동 가능성 대비가 가능하다. 설계원칙 / 모니터링 / 제약과 감독 등을 들 수 있다.
   3) AI를 활용한 보안 강화 : 사용자 인증, 이상거래 탐지, 악성코드 탐지, 사이버 공격 탐지, 지능형 CCTV, 지능형 보안관제가 있다.
3. AI 보안 위협
   1) Poisoning attack : 잘못된 학습데이터를 주입하여 AI시스템 오동작 유발, 최소의 오염 데이터 추가로 오류를 최대화 시킨다.
   2) Evasion attack : 기만 공격이라고 하며 활용 단계의 분류 데이터를 변조하여 오작동을 유발 시키는 공격이다
   3) Adversarial attack : target classifier 를 속일 수 있도록 변조하는 공격이다.
   4) Model extraction attack : 학습된 모델에 쿼리를 해서 타겟모델 f에 가까운 f'를 만들어 복제하는 공격이다.
   5) inversion attack : 모델에 질의하여 training data를 재현해낸다.
4. AI 관련 연구 소개
   1) security by AI – 목소리 인증, 삼성페이 결재 도용, 모바일 서명 인증, 분장 공격, 블랙박스 공격

1. 강의 요약 및 소감: (문장식으로 작성, 조리 있는 글쓰기 연습의 기회)
2. 개선사항: (강의실 환경, 강의 방법, 강사 음량, 강의 내용 구성, 기타 건의사항)