

1주차

인공지능의 과거와 현재

- 4차 산업혁명: IoT(사물인터넷), 인공지능, 만물초지능 혁명으로 인해 사람, 사물, 공간이 초연결되고 초지능화 되는 사회 전반 시스템의 혁신

2주차

Pandas

- Pandas란?
 - 데이터분석을 위한 Python 라이브러리
 - 대용량의 데이터 처리를 지원함. (자동화된 분석, 대용량의 데이터 처리, 머신러닝, 시각화 등의 데이터 사이언스 관련 라이브러리에서 사용)

Numpy

- numpy란?
 - (Numper / Numerical + Python)
 - 수치 연산, 과학 연산을 위한 파이썬 외부 라이브러리
 - 데이터분석, 시각화, 머신러닝 등의 작업에 필수적
 - 벡터, 행렬 등의 자료구조 및 연산 지원
- 제공하는 기능들
 - 통계 함수들: 최대, 최소, 평균, 중간값, 분산, 표준편차, n분위수
 - 수학 함수들: 삼각함수, 로그함수 등
 - 벡터 및 행렬 연산: 행렬의 곱, 역행렬, 전치행렬(Array라는 이름으로 제공)
 - 공학수학, 선형대수학 등

- numpy의 장점
 - 데이터를 생성할 수 있다. (특정 패턴의 수열, 랜덤 수, 특정 분포에 근거한 데이터, 수학 함수 데이터)
 - 많은 데이터를 쉽고 빠르게 처리할 수 있다. (리스트로 하는 것보다 수행 속도가 훨씬 빠름, 코드도 훨씬 짧음)
 - 복잡한 연산을 수행할 수 있다. (통계, 선형 대수, 푸리에 연산 등)

range()와 arange()

- 공통점
 - (시작값, 종료값, 증감값) 동일
- 차이점
 - range(): 시작값, 종료값, 증감값에 실수 지원 안함
 - arange(): 시작값, 종료값, 증감값에 실수 지원

Matplotlib

- matplotlib란?
 - (mat + plot+ lib)
 - 데이터를 차트나 Plot등 다양한 형태로 그려주는 데이터 시각화 패키지

Seaborn

- seaborn란?
 - matplotlib 위에서 동작되는 시각화 라이브러리
- 사용
 - 두 수치형 칼럼 간의 관계를 파악할 때 사용
 - 수치형 칼럼의 빈도수를 확인하고자 할 때 사용
 - 범주형 칼럼의 빈도수를 확인하고자 할 때 사용
 - 데이터 분포를 백분율로 확인 가능
 - 모든 변수간의 관계를 일괄적으로 확인하고자 할 경우에 사용

4주차

문자열(String) 처리 방식

- 인공지능 현업의 적용 사례
 - 대학교 사례
 - 인공지능 학사 상담 로봇 운영을 시작함
 - 챗팅을 사람이 하는 것이 아니라 인공지능 로봇이 답변
 - 지역 축제 상담
 - 일시적인 상담이 급증하는 상황
 - 고정적인 인력 채용은 불가능한 상황
 - 관광회사
 - 저녁, 야간 시간대 상담 가능 / 여행을 앞두고 긴급한 상담
- 데이터 분석 & 문자열
 - 단순히 단어의 개수를 세는 것으로도 의미 있는 빅데이터 분석을 할 수 있음
 - 아주 간단해 보이지만 처리해야 할 것이 있음
 - 공백을 처리 (띄어쓰기 하는 사람, 안하는 사람)
 - 영어의 경우 대소문자 일치 (어떤 사람은 소문자 어떤 사람은 대문자, 어떤 사람은 앞만 대문자)
 - 시간단위로 검색어의 변화량
 - 두 단어가 얼마나 같이 출현하는가?
 - (갤럭시 노트, 갤럭시 폴더 --- 대박, 실망, 완판, 문제점, 기대)

Word Cloud 데이터 시각화분석

- 워드 클라우드란?
 - 긴 문자열에 대해서 그 안에 단어들을 검색하고 단어들의 빈도 수와 중요성을 큰 단어로 표현하고 그렇지 않은 것은 작은 단어로 표현하는 것

자연어 처리와 워드클라우드

- 챗봇
 - 채팅과 로봇이 결합되어 만들어진 용어
 - 인간의 자연언어를 이해하고 대화가 가능한 채팅 로봇
 - 초기 챗봇: 입력한 텍스트를 이해하고 단변하는 텍스트 기반
 - 최근 챗봇: STT(Sound to Text) 및 TTS(Text to Sound) 기술과 접목되어 음성 기반으로 동작
- 자연어 처리(NLP)
 - Natural Language Processing
 - 기계가 자연어를 이해하고 해석하여 처리할 수 있도록 하는 일
 - 활동분야
 - 스팸 메일 분류
 - 뉴스 기사의 내용을 기반으로 연애/정치/사회/문화 중 어떤 카테고리에 속하는지 자동으로 분류
 - SNS의 글을 분석하여 글쓴이의 감정을 분석하는 것, 영화 및 제품의 리뷰를 분석하는 것
 - 구글 번역기나 파파고와 같은 번역기
 - 애플의 시리나 삼성 갤럭시의 빅스비, 챗봇 등
- 한글 자연어 처리: koNLpy 라이브러리
 - 자연어 처리를 위해서는 문장, 형태소, 품사 태깅같은 텍스트 분석이 필요
 - 한글 텍스트 분석은 영어에 비해 조사와 복합 명사 등의 분리가 어려움

5주차

지도학습

- 지도학습
 - 학습할 데이터와 명시적인 정답(레이블)을 이용해 데이터의 특성과 분포를 학습하여 새로운 데이터에 대해 미래 결괏값 예측하는 방법
 - 과거의 데이터로부터 학습하여 결과를 예측하는데 주로 사용
 - 분류(Classification)
 - 학습데이터를 이용해 주어진 입력값이 어떤 종류인지 구별하는 것
 - k-최근접 이웃 알고리즘, 서포트 벡터 머신

- 의사결정트리, 랜덤포레스트, 나이브 베이즈
- 회귀(Regression)
 - 학습 데이터를 이용해 연속적인(숫자) 값을 예측하는 것
 - 선형회귀, 로지스틱 회귀
- 선형회귀의 목표
 - 직선과 데이터 차이가 평균적으로 가장 작아지는 직선 찾기
- 의사결정트리
 - 데이터들 사이에 존재하는 패턴을 예측 가능한 규칙들의 조합의 나무 형태로 나타내어주는 모형
- 서포트 벡터 머신
 - 서로 다른 분류 값을 결정하는 경계선(결정 경계선)을 결정하는 알고리즘

6주차

인공지능과 데이터 윤리

- 인공지능의 분류
 - 좁은 인공지능 (Narrow AI)
 - 스스로 사고해 문제를 해결할 수 있는 능력이 없는 컴퓨터 기반 AI
 - 특정분야에 국한된 인공지능
 - 특정 영역에서 인간보다 나은 성과
 - 일반 인공지능 (General AI)
 - 스스로 사고해 문제를 해결할 수 있는 능력을 가진 컴퓨터 기반 AI
 - 인간 수준의 능력을 가진 인공지능
 - 모든 분야에 적용될 수 있는 인공지능
 - 자각력과 독립성을 갖추
 - 슈퍼 인공지능 (Super AI)
 - 인간보다 100~1000배 (IQ 1만 ~ 10만) 뛰어난 지능을 가진 AI
 - 효율, 자기보존, 자원획득, 창의성 같은 원초적 욕구를 기반으로 끊임없이 자기 발전

웹 크롤링

- 웹스크래핑(Web scraping)
 - 컴퓨터 소프트웨어 기술을 활용해 웹 사이트 내에 있는 특정 정보를 추출하는 일
 - 웹 사이트에 접속해 자료를 추출, 필요한 정보만 가져오도록 구조를 분석하고 정리하는 작업
 - 크롤링(Crawling)
 - 사이트나 네트워크가 제공할 수 있는 것을 끝없이 탐색하면서 스스로 웹 페이지를 조직적, 자동적으로 탐색하는 작업
 - 크롤링하는 프로그램을 '크롤러(Crawler)' 또는 "스파이더(Spider)"라고 함
 - 검색 엔진과 같은 여러 사이트에서 데이터의 최신 강태를 유지하기 위해 사용
 - 크롤링에는 웹 페이지의 특성 수준의 스크래핑이 필요

In []: