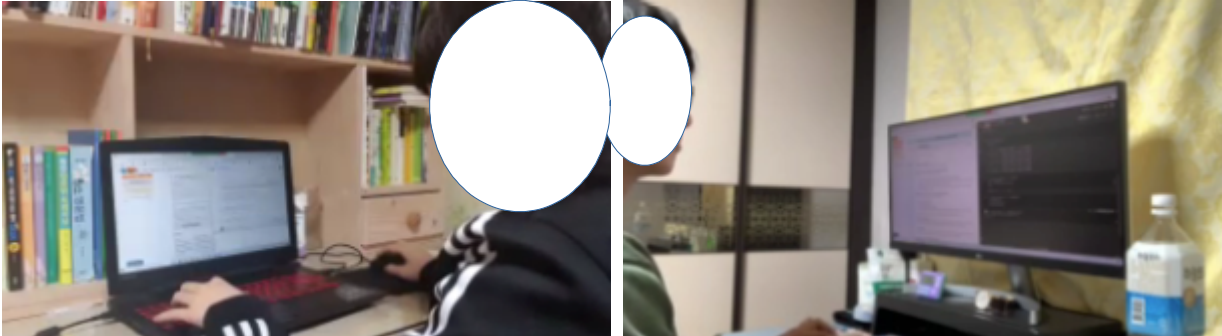


2021.기계학습.기말고사.시험지 - 2021년 6월 14일 (월)

[온라인 시험 바른 자세 예시]

아래와 같이 바른 자세 예시를 따르지 않는 학생은 퇴실 조치 합니다.

자세 검사를 마친 학생에 한하여 시험지의 비밀번호가 제공됩니다.



온라인 시험 주의 사항

- 시험 중 화장실에 다녀올 수 없음
- 화면 녹화 본 제출 필수 (대용량 이메일), 화면 녹화 중단시 녹화된 곳 까지만 점수 인정
- 화면 녹화는 반드시 7시부터 시작, 녹화 시작시 컴퓨터 하단에 그 어떠한 창도 실행되어 있으면 안됨
- 코드 제출 필수(구글드라이브), 반드시 제출 코드로 리더보드 재현이 가능해야 인정
- OJ를 통한 IP 필터링을 통해 모여서 시험을 보는 행위 방지함, 이에 테더링 사용 불가
- 고사 중 부적절한 자세로 인한 경고 3회 이상 누적시 시험 0점 처리
- 듀얼모니터 사용 불가, 시험 시작 후 듀얼 모니터 사용 중이 아님을 입증하는 화면 캡처 제출 필수
- 시험은 오픈 메뉴얼 북 (메뉴얼 안에서 검색 가능)만 허용, 강의안과 책 불가, 인터넷 검색 불가
- 고시 시작 두 시간 이후 퇴실 가능하나 (퇴실 시 반드시 퇴실 하겠다고 알릴 것) 재 입실 불가
- 시험 중 캐글 리더보드 여전히 20회 제출 제한 있음
- 실시간 중계되는 감독 화면은 녹화되어 보관됨
- 고사 중 공지 및 질문을 위해서는 슬랙을 사용할 예정, 반드시 본인 PC에 슬랙 설치 필수
- 고사 중 카카오톡과 같은 메신저는 반드시 종료할 것. 적발시 F 처리
- 기존 코드를 참고하는 행위 절대 불가, 적발시 F
- 대리 시험 절대 불가, 적발시 학사경고

제출

- 화면 녹화 파일 제출: admin@rcv.sejong.ac.kr
 - : 화면 녹화 본 중 듀얼모니터 부분 화면 캡처 제출: 동영상과 함께 이메일로 제출
- 코드 및 기타 파일
 - : 공유된 구글 드라이브

[이론 문제][15점] <https://forms.gle/VtKT3CBaswYkdF259>

총 14문제 중 13문제 객관식, 1문제 주관식

[실습 문제1][3점] 태양광 발전량 예측 <https://www.kaggle.com/t/361727fefccb4194a79cd4e2b45b6b17>

최근 태양전지판을 설치하여 전기를 만들어 사용하는 가정집이 늘어나고 있다. 이는 초기 설비만 갓출 경우 반영구적으로 전기를 생산할 수 있고 전기생산과정에서 독성물질이 배출되지 않아 신에너지원으로 각광받고 있기도 하다. 그러나 태양광 에너지는 날씨의 영향을 많이 받아 생산량이 일정하지 않다는 단점이 존재한다. 최근 형준이네 집은 이와 같은 트렌드에 맞춰 태양전지판을 설치하였으나 내일의 전기 생산량을 예상할 수 없다는 어려움을 겪고 있다. 여러분이 지금부터 수업시간에 배운 <머신러닝 기술>을 활용하여, test.csv 파일로 제공된 정보를 통해 태양광 발전량을 예측하는 인공지능 SW를 작성해주길 바란다.

학습 데이터로는 (기상 데이터 정보)와 해당 기상 상태에 따른 (태양광 발전량)을 함께 제공합니다.

Hour	Minute	DHI	DNI	WS	RH	T	TARGET
0	0	0	0	0	1.4	57.76	0
1	15	0	222	18	1.3	26.82	13
2	4	0	0	0	2.6	84.94	-6
3	18	30	0	0	1.5	89.18	2
4	1	30	0	0	2.4	84.35	12
...

테스트 데이터로는 (기상 데이터 정보)만 제공하며, 예측된 (태양광 발전량)은 submit.csv 파일로 저장하여 캐글 리더보드에 제출하셔야 합니다.

Hour	Minute	DHI	DNI	WS	RH	T
0	6	0	54	548	1.2	72.29
1	14	0	297	27	8.0	19.07
2	1	30	0	0	2.9	66.29
3	19	0	0	0	1.8	63.38
4	4	30	0	0	1.7	55.61
...

제공되는 (기상 데이터 정보)는 순서대로 Hour(시간), Minute(분), DHI(수평면 산란일사량), DNI(직달일사량), WS(풍속), RH(상대습도), T(기온), Target(태양광 발전량)입니다.

자, 그럼 테스트 데이터로 주어진 (기상 데이터 정보)를 가지고 (태양광 발전량)을 예측하여 봅시다.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받습니다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. **(미 제출시 10%점 감점)**
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.
(미 제출시 10% 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 상위 랭커 6인에게 추가 점수를 부여합니다. 단, 공동 순위가 6인 이상일 경우 점수를 부여하지 않으며, 또한 풀이방식이 문제가 있다면 상위 랭커 6인 안에 들어가더라도 가점이 부여되지 않습니다.
(20%점 가점)
- ⑤ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑥ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[실습 문제2][2점] 승차인원 예측 <https://www.kaggle.com/t/d47f5d06b5c748129cc87161e2a64dae>

2019년을 기준으로 제주도민과 관광객의 증가로 교통체증 문제가 심각한 사안으로 떠오르고 있다. 제주도는 퇴근시간 대 승차인원을 정확하게 예측하여 제주 버스의 효율적인 운영을 하고자 한다. 제주도 각 지역에서 측정된 데이터를 바탕으로 수업시간에 배운 <머신러닝 기술>을 활용하여 18~20시 퇴근시간 승차인원을 예측하는 인공지능 SW를 작성해주길 바란다.

학습 데이터로는 (버스 운행 정보)와 (18-20시 승차 인원)을 함께 제공합니다.

id	year	month	day	weekday	bus_route_id	in_out	station_code	6~7_ride	7~8_ride	...	9~10_ride	10~11_ride	11~12_ride	6~7_takeoff	7~8_takeoff	8~9_takeoff	9~10_takeoff	10~11_takeoff	11~12_takeoff	18~20_ride
0	2019	9	29	1	3250	0	3281	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1	2019	9	14	1	2871	0	1145	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
2	2019	9	17	0	3074	0	3014	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
3	2019	9	28	1	3082	0	198	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	2019	9	12	0	3247	0	3274	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
...

테스트 데이터로는 (버스 운행 정보)만 제공하며, 예측된 (18~20시 승차인원)은 submit.csv 파일로 저장하여 캐글 리더보드에 제출하여야 합니다.

id	year	month	day	weekday	bus_route_id	in_out	station_code	6~7_ride	7~8_ride	8~9_ride	9~10_ride	10~11_ride	11~12_ride	6~7_takeoff	7~8_takeoff	8~9_takeoff	9~10_takeoff	10~11_takeoff	11~12_takeoff
0	2019	9	10	0	2912	0	2468	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1	2019	9	6	0	3182	0	2972	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	5.0	0.0
2	2019	9	6	0	3133	0	2811	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	2019	9	11	0	3145	0	516	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	2019	9	25	0	2326	0	168	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	2.0	0.0	0.0	0.0

제공되는 (버스 운행 정보)는 고유번호, 년, 월, 일, 평일(0)/주말(1) 여부, 버스 노선 번호, 시내(0)/시외버스 (1) 여부, 정류장번호, 6-7시 승차인원, 7-8시 승차인원, 8-9시 승차인원, 9-10시 승차인원, 10-11시 승차인원, 11-12시 승차인원, 6-7시 하차인원, 7-8시 하차인원, 8-9시 하차인원, 9-10시 하차인원, 10-11시 하차인원, 11-12시 하차인원 입니다.

자, 그럼 테스트 데이터로 주어진 (버스 운행 정보)를 가지고 (18-20시 승차인원)을 예측하여 봅시다.
>> 베이스라인은 1개이며, score 이상으로 받아야 점수를 부여 받습니다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. (미 제출시 10%점 감점)
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.
(미 제출시 10% 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 상위 랭커 6인에게 추가 점수를 부여합니다. 단, 공동 순위가 6인 이상일 경우 점수를 부여하지 않으며, 또한 풀이방식이 문제가 있다면 상위 랭커 6인 안에 들어가더라도 가점이 부여되지 않습니다.
(20%점 가점)
- ⑤ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑥ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[실습 문제3][5점] 구내식당 식수 예측 <https://www.kaggle.com/t/a5962a6d1f684063b616bc7013f9227c>

한국토지주택공사는 빅데이터 분석을 통해 잔반 발생량을 획기적으로 줄이고자 한다. 여러분은 지금부터 수업 시간에 배운 <머신러닝 기술>을 활용하여, 구내식당의 요일별 점심 식사를 먹는 인원을 예측하는 인공지능 SW를 작성해줄기 바란다.

학습 데이터로는 (회사 직원 정보와 식당 메뉴)가 제공되며, (점심 식사 인원 수)가 함께 제공됩니다.

	일자	요일	본사 정원 수	본사휴 가자수	본사출 장자수	본사시간외근무 명령서승인건수	현본사소속재 택근무자수	조식메뉴	중식메뉴	석식메뉴	중식계
0	2016-02-01	월	2601	50	150	238	0.0	모닝콜/편형 우유/두유/주스 계란후라이 호두죽/쌀밥 (쌀:국내산) 완장찌개 취...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 오징어찌개 쇠불고기 (쇠고기:호주산) 계란찜 ...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 육개장 자란고등어구이 두부조림 건파래무침 ...	1039.0
1	2016-02-02	화	2601	50	173	319	0.0	모닝콜/단호박샌드 우유/두유/주스 계란후라이 팔죽/쌀밥 (쌀:국내산) 호박젓국찌...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 김치찌개 가자미튀김 모듬소세지구이 마늘종무...	콩나물밥*양념장 (쌀,현미흑미:국내산) 어묵국 유산술 (쇠고기:호주산) 아삭고추무...	867.0
2	2016-02-03	수	2601	56	180	111	0.0	모닝콜/메이글 우유/두유/주스 계란후라이 표고버섯죽/쌀밥 (쌀:국내산) 콩나물국...	카레알밥 (쌀,현미흑미:국내산) 팔이장국 치킨튀거 (닭고기:국내산) 물면아채무침 ...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 현국장찌개 황태양념구이 (황태:러시아산) 고기...	1017.0
3	2016-02-04	목	2601	104	220	355	0.0	모닝콜/토마토샌드 우유/두유/주스 계란후라이 닭죽/쌀밥 (쌀,닭:국내산) 근대국...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 쇠고기무국 주꾸미볶음 부추전 시금치나물 ...	미니김밥*거자장 (쌀,현미흑미:국내산) 우동 떡시칸샐러드 군고구마 무피를 포...	978.0
4	2016-02-05	금	2601	278	181	34	0.0	모닝콜/와플 우유/두유/주스 계란후라이 쇠고기죽/쌀밥 (쌀:국내산) 재첩국 방...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 떡국 돈육쌈강정 (돼지고기:국내산) 우영찰채...	쌀밥/잡곡밥 (쌀,현미흑미:국내산) 차돌박이찌개 (쇠고기:호주산) 닭갈비 (닭고기:...	925.0
...

테스트 데이터로는 (회사 직원 정보와 식당 메뉴)만 제공되며, 예측된 (점심 식사 인원 수)는 submit.csv 파일로 저장하여 캐글 리더보드에 제출하셔야 합니다.

	일자	요일	본사정 원수	본사휴 가자수	본사출 장자수	본사시간외근무 명령서승인건수	현본사소속재 택근무자수	조식메뉴	중식메뉴	석식메뉴
843	2019-07-17	수	2760	96	259	0	0.0	모닝콜/물호떡 우유/주스 계란후라이 누룽지죽/쌀밥 (쌀:국내산) 연두부탕 장담채...	산채비빔밥 미소장국 아재취침 아삭고추무침 오뎅레포기잡치 (잡치:국내산)	쌀밥/잡곡밥 (쌀:국내산) 두부김치국 돼지고기장조림 아재게란밥 꽃송추무침 열...
844	2019-07-18	목	2760	113	274	310	0.0	모닝콜/민들레샌드 우유/주스 계란후라이 녹두죽/쌀밥 (쌀:국내산) 열무된장국 명...	쌀밥/잡곡밥 (쌀:국내산) 돈육김치찌개 미니채소칼비 모듬목양생장 찜나물생채 ...	쌀밥/잡곡밥 (쌀:국내산) 브로콜리크림스프 옛날왕돈가스 키위드레싱샐러드 편육수...
845	2019-07-19	금	2760	236	291	0	0.0	모닝콜/편형 우유/주스 계란후라이 녹임자죽/쌀밥 (쌀:국내산) 콩나물국 연근조림...	쌀밥/잡곡밥 (쌀:국내산) 소고기무국 갈치구이 비엔나감자볶음 부생채 포기잡치...	쌀밥/잡곡밥 (쌀:국내산) 꽃게탕 한통찜된 해물겨자냉채 간장맛짬뽕 포기잡치 ...
846	2019-07-22	월	2760	223	218	438	0.0	모닝콜/마늘빵 우유/주스 계란후라이 누룽지죽/쌀밥 (쌀:국내산) 호박젓국찌개 김...	쌀밥/잡곡밥 (쌀:국내산) 매운어묵국 소고기숙주볶음 치킨샐러드 양파명아찌 포...	쌀밥/잡곡밥 (쌀:국내산) 달걀국 오리불고기 두부구이 부추생채/상추쌈 포기...
847	2019-07-23	화	2760	206	246	449	0.0	모닝콜/홍우이면 우유/주스 스크램블에그 닭죽/쌀밥 (쌀:국내산) 김자국 동전문어...	쌀밥/잡곡밥 (쌀:국내산) 사금치미장국 등갈비김치찜 김밥 닭은국 해물볶음우동 나초콜샐러드 윤로 포기잡치 (잡치:국내산)	
...

제공되는 (회사 직원 정보와 식당 메뉴)는 순서대로 일자, 요일, 본사정원수, 본사휴가자수, 본사출장자수, 시간외근무명령서승인건수, 현본사소속재택근무자수, 조식메뉴, 중식메뉴, 석식메뉴 입니다.

자, 그럼 테스트 데이터로 주어진 (회사 직원 정보와 식당 메뉴)를 가지고 요일별 (점심 식사 인원 수)를 예측하여 봅시다.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받는다.

>> 메뉴 구성을 1d Vector하기 위해 텀프로젝트에서 사용했던 CounterVectorizer()를 활용 가능합니다

```
from sklearn.feature_extraction.text import CountVectorizer
```

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. **(미 제출시 10%점 감점)**
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요. **(미 제출시 10% 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)**
- ④ 상위 랭커 6인에게 추가 점수를 부여합니다. 단, 공동 순위가 6인 이상일 경우 점수를 부여하지 않으며, 또한 풀이방식이 문제가 있다면 상위 랭커 6인 안에 들어가더라도 가점이 부여되지 않습니다.
(20%점 가점)
- ⑤ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑥ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.

[실습 문제4][5점] 영어음성 국적 예측 <https://www.kaggle.com/t/98ffb0cbe2094035aa72fdb8022c2fc>

정민이는 다국어 음성데이터 세트를 분석하여 English Accent 분류 알고리즘을 개발하려고 한다. 여러분 역시 수업시간에 배운 <머신러닝 기술>을 활용하여, 영어 악센트를 통한 영어 음성 국적을 예측해주는 인공지능 SW를 작성해 주기 바란다.

학습 데이터는 확장자가 wav인 국적별 (음성파일)이 녹음되어 제공되며, 파일의 길이는 서로 다르다.

테스트 데이터는 확장자가 wav인 국적을 확인할 수 없는 (음성파일)만 녹음되어 제공된다.

데이터로 제공되는 (음성 파일)은 비정형데이터로 피쳐 추출 단계를 거쳐 1D 벡터화가 필요하다.

자, 그럼 테스트 데이터로 주어진 (음성 파일)에 맞는 (국적)을 예측하여 보자.

>> 베이스라인은 2개이며, 각 score 이상으로 받아야 점수를 부여 받는다.

>> 음성데이터 처리를 위해 텀프로젝트에서 사용했던 librosa 라이브러리를 사용하세요.

- librosa.load() / librosa.stft() / librosa.feature.melspectrogram() / librosa.feature.mfcc()
- 현재 베이스라인은 sample_rate=16000, n_fft=512로 작성되어 있습니다.

※ 주의사항 ※

- ① 모듈별 코드(데이터전처리, 모델학습, 모델검증 등) 에 주석을 반드시 달아주세요. 특히, 어떤 목적으로 API를 호출했는지 작성되어야 합니다. **(미 제출시 10%점 감점)**
- ② 배운 범위 내의 기계학습 방법론을 사용하여 베이스라인을 넘으면 됩니다.
(인공신경망 방법론 사용의 경우 0점 처리)
- ③ 각 문제 제출 코드 마지막에 텍스트 셀을 추가하여 실험 결과 분석 내용을 서술형으로 작성해주세요.
(미 제출시 10% 감점, 설명이 부족하고 논리적이지 않으면 부분 감점)
- ④ 상위 랭커 6인에게 추가 점수를 부여합니다. 단, 공동 순위가 6인 이상일 경우 점수를 부여하지 않으며, 또한 풀이방식이 문제가 있다면 상위 랭커 6인 안에 들어가더라도 가점이 부여되지 않습니다.
(20%점 가점)
- ⑤ 데이터 전처리는 scikit-learn에 있는 다양한 함수를 모두 사용하셔도 좋습니다.
- ⑥ 오픈라이브러리의 공식 매뉴얼과 stackoverflow의 사용 가능합니다.