

Homework #2

Sept. 2025

<https://linklab.github.io>

[Note] 숙제 시작 전에 link_dl 리포지토리에 대해 git pull 을 수행하여 코드 최신화를 먼저 수행 필요

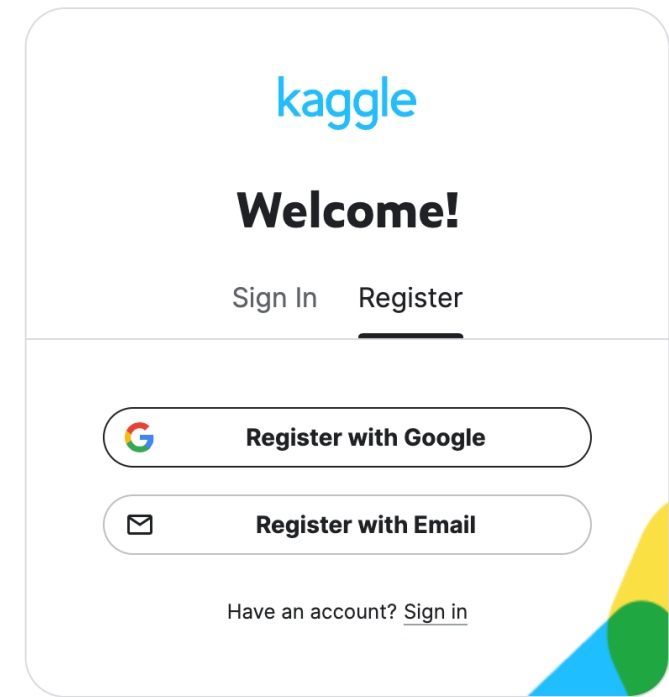
Kaggle Competition

◆ Kaggle 회원 가입

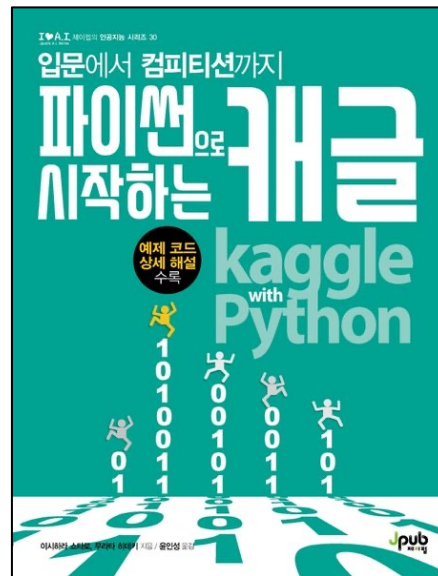
- <https://www.kaggle.com/>
- 가급적 구글 계정으로 회원 가입 추천

◆ Kaggle 둘러보기/알아보기

- <https://musma.github.io/2019/03/04/about-kaggle.html>



◆ 추천서



Kaggle Competition - Titanic

◆ Titanic - Machine Learning from Disaster @ Kaggle

- <https://www.kaggle.com/competitions/titanic>
- This is the legendary Titanic ML competition: the best, first challenge for you to dive into ML competitions and familiarize yourself with how the Kaggle platform works.
- History
 - The sinking of the Titanic is one of the most infamous shipwrecks in history
 - On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" Titanic sank after colliding with an iceberg
 - Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew
 - While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others
 - In this challenge, we ask you to build a predictive model that answers the question:
 - "what sorts of people were more likely to survive?" using passenger data (i.e. age, gender, socio-economic class, etc
- 한글 설명 (반드시 읽어보기)
 - <https://developers.ascentnet.co.jp/2017/11/24/kaggle-process-review/>

Kaggle Competition - Titanic

◆ [데이터 전처리 이전] Titanic - Machine Learning from Disaster @ Kaggle

— Target Feature

- **Survived: 0 = 사망, 1 = 생존**

— Feature

• ~~PassengerId: 승객 번호~~

• ~~Name: 이름~~

- **Pclass: 티켓 클래스**

➢ 1 = 1등석, 2 = 2등석, 3 = 3등석

- **Sex: 성별**

➢ male = 남성, female = 여성

- **Age: 나이**

- **SibSp: 동승한 자매 / 배우자의 수**

- **Parch: 동승한 부모 / 자식의 수**

• ~~Ticket: 티켓 번호~~

- **Fare: 승객 요금**

• ~~Cabin: 방 호수~~

- **Embarked: 탑승지**

➢ C = 세르부르, Q = 퀸즈타운, S = 사우스햄프턴

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhe	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S
12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05		S
14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16		S
17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125		Q
18	1	2	Williams, Mr. Charles Eugene	male		0	0	244373	13		S
19	0	3	Vander Planke, Mrs. Julius (Emelia Mari	female	31	1	0	345763	18		S
20	1	3	Masselmani, Mrs. Fatima	female		0	0	2649	7.225		C
21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26		S

Kaggle Competition - Titanic

◆ [데이터 전처리 이후] Titanic - Machine Learning from Disaster @ Kaggle

– Target Feature

- Survived: 0 = 사망, 1 = 생존

– Feature (총 10개)

- Pclass: 티켓 클래스
 - 1 = 1등석, 2 = 2등석, 3 = 3등석)
- Sex: 성별
 - male = 남성, female = 여성
- Age: 나이
- SibSp: 동승한 자매 / 배우자의 수
- Parch: 동승한 부모 / 자식의 수
- Fare: 승객 요금
- Embarked: 탑승지
 - C = 세르부르, Q = 퀸즈타운, S = 사우스햄프턴)

Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked	title	family_num	alone
0.0	3	1	22.0	1	0	7.2500	2	2	1	0.0
1.0	1	0	38.0	1	0	71.2833	0	3	1	0.0
1.0	3	0	26.0	0	0	7.9250	2	1	0	1.0
1.0	1	0	35.0	1	0	53.1000	2	3	1	0.0
0.0	3	1	35.0	0	0	8.0500	2	2	0	1.0
0.0	3	1	29.0	0	0	8.4583	1	2	0	1.0
0.0	1	1	54.0	0	0	51.8625	2	2	0	1.0
0.0	3	1	2.0	3	1	21.0750	2	0	4	0.0
1.0	3	0	27.0	0	2	11.1333	2	3	2	0.0
1.0	2	0	14.0	1	0	30.0708	0	3	1	0.0

- title:
 - 0 = Mr, 1 = Miss, 2 = Mrs, 3 = Master, 4 = other
- family_num = SibSp + Parch
- alone
 - 0: family_num > 0, 1: family_num가 0

Kaggle Competition - Titanic

◆ 데이터 준비: _03_homeworks/homework_2/titanic_dataset.py (1/5)

```
import os, torch
import pandas as pd
from torch.utils.data import Dataset, DataLoader, random_split

class TitanicDataset(Dataset):
    def __init__(self, X, y):
        self.X = torch.FloatTensor(X)
        self.y = torch.LongTensor(y)

    def __len__(self):
        return len(self.X)

    def __getitem__(self, idx):
        feature = self.X[idx]
        target = self.y[idx]
        return {'input': feature, 'target': target}

    def __str__(self):
        str = "Data Size: {0}, Input Shape: {1}, Target Shape: {2}".format(
            len(self.X), self.X.shape, self.y.shape
        )
        return str
```

Kaggle Competition - Titanic

◆ 데이터 준비: _03_homeworks/homework_2/titanic_dataset.py (2/5)

```
class TitanicTestDataset(Dataset):
    def __init__(self, X):
        self.X = torch.FloatTensor(X)

    def __len__(self):
        return len(self.X)

    def __getitem__(self, idx):
        feature = self.X[idx]
        return {'input': feature}

    def __str__(self):
        str = "Data Size: {0}, Input Shape: {1}".format(
            len(self.X), self.X.shape
        )
        return str
```

Kaggle Competition - Titanic

◆ 데이터 준비: _03_homeworks/homework_2/titanic_dataset.py (3/5)

```
def get_preprocessed_dataset():  
    CURRENT_FILE_PATH = os.path.dirname(os.path.abspath(__file__))  
  
    train_data_path = os.path.join(CURRENT_FILE_PATH, "train.csv")  
    test_data_path = os.path.join(CURRENT_FILE_PATH, "test.csv")  
  
    train_df = pd.read_csv(train_data_path)  
    test_df = pd.read_csv(test_data_path)  
  
    all_df = pd.concat([train_df, test_df], sort=False)  
    all_df = get_preprocessed_dataset_1(all_df)  
    all_df = get_preprocessed_dataset_2(all_df)  
    all_df = get_preprocessed_dataset_3(all_df)  
    all_df = get_preprocessed_dataset_4(all_df)  
    all_df = get_preprocessed_dataset_5(all_df)  
    all_df = get_preprocessed_dataset_6(all_df)  
    ...
```


Kaggle Competition - Titanic

◆ 데이터 준비: _03_homeworks/homework_2/titanic_dataset.py (4/5)

```
if __name__ == "__main__":
    train_dataset, validation_dataset, test_dataset = get_preprocessed_dataset()

    print("train_dataset: {0}, validation_dataset.shape: {1}, test_dataset: {2}".format(
        len(train_dataset), len(validation_dataset), len(test_dataset)
    ))
    print("#" * 50, 1)

    train_data_loader = DataLoader(
        dataset=train_dataset, batch_size=16, shuffle=True
    )

    validation_data_loader = DataLoader(
        dataset=validation_dataset, batch_size=len(validation_dataset)
    )

    test_data_loader = DataLoader(
        dataset=test_dataset, batch_size=len(test_dataset)
    )
```

Kaggle Competition - Titanic

◆ 데이터 준비: _03_homeworks/homework_2/titanic_dataset.py (5/5)

```
if __name__ == "__main__":
    ...
    print("[TRAIN]")
    for idx, batch in enumerate(train_data_loader):
        print("{0} - {1}: {2}".format(idx, batch['input'].shape, batch['target'].shape))

    print("[VALIDATION]")
    for idx, batch in enumerate(validation_data_loader):
        print("{0} - {1}: {2}".format(idx, batch['input'].shape, batch['target'].shape))

    print("[TEST]")
    for idx, batch in enumerate(validation_data_loader):
        print("{0} - {1} ".format(idx, batch['input'].shape))
```

```
[TRAIN]
0 - torch.Size([16, 10]): torch.Size([16])
1 - torch.Size([16, 10]): torch.Size([16])
...
43 - torch.Size([16, 10]): torch.Size([16])
44 - torch.Size([9, 10]): torch.Size([9])
[VALIDATION]
0 - torch.Size([178, 10]): torch.Size([178])
[TEST]
0 - torch.Size([418, 10])
```

Kaggle Competition - Titanic

◆[요구사항 1] titanic 딥러닝 모델 기본 훈련

- `_01_code/_08_learning_and_optimization/c_my_model_training_with_argparse_wandb.py`
코드를 그대로 활용하되 titanic 데이터에 맞게 수정하여 코딩하기
- Wandb로 훈련 과정 데이터 올려 그래프 얻어 내기
 - Training loss
 - Validation loss
 - 위 두 그래프를 보여주는 Wandb URL 얻어내기

Kaggle Competition - Titanic

◆[요구사항 2] Activation Function 과 Batch Size 변경 및 선택하기

- 모델 구성 내에 Activation Function를 변경하여 더 나은 성능을 산출하는 Activation Function 이 무엇인지 조사하기
 - Sigmoid vs. ReLU vs. ELU vs. Leaky ReLU
- 모델 구성 내에 Training Batch Size를 변경하여 더 나은 성능을 산출하는 Batch Size 가 무엇인지 조사하기
 - 16, 32, 64, 128

Kaggle Competition - Titanic

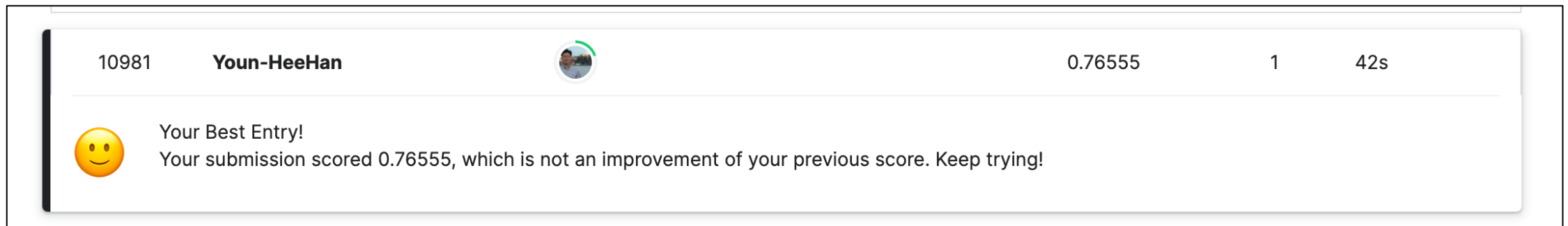
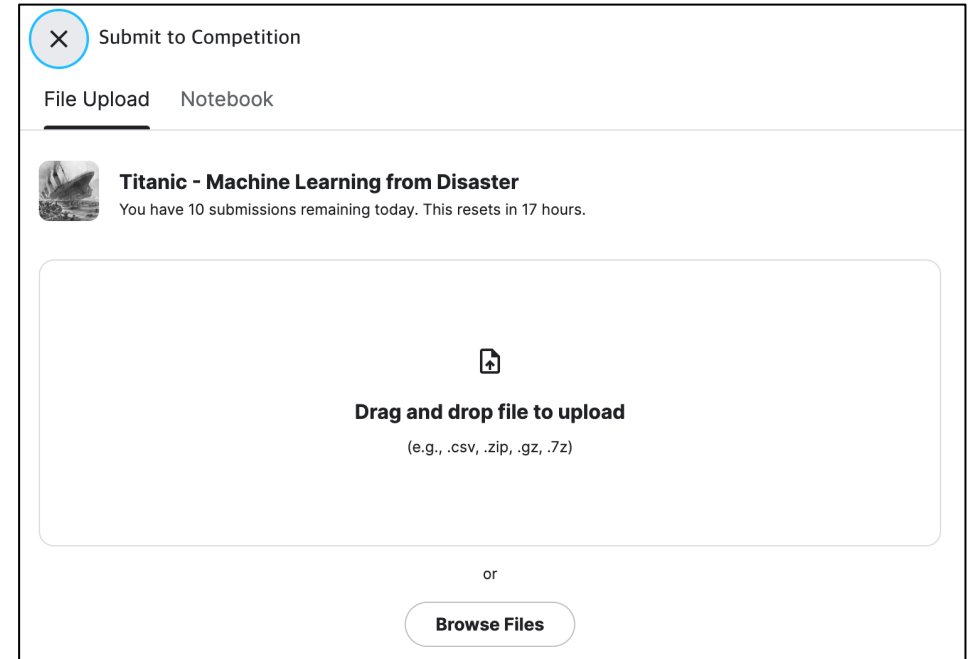
◆[요구사항 3] 테스트 및 submission.csv 생성

- 요구사항 2에서 살펴본 가장 좋은 성능을 보이는 Activation Function 및 Batch Size로 모델 구성하기
- 사전에 테스트 데이터 (즉, `test_data_loader`) 구성하기
- 훈련과정 중 어느 Epoch 시점에 테스트를 수행하여 `submission.csv` 를 구성해야 하는지 고찰하기
- 고찰한 내용에 대한 추가 코딩 수행
- `submission.csv` 생성하기

Kaggle Competition - Titanic

◆[요구사항 4] submission.csv 제출 및 등수확인

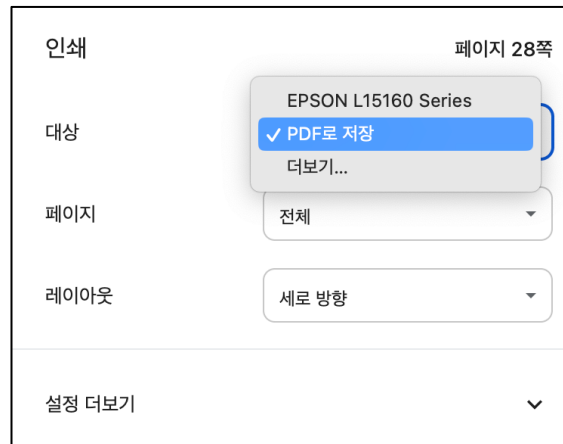
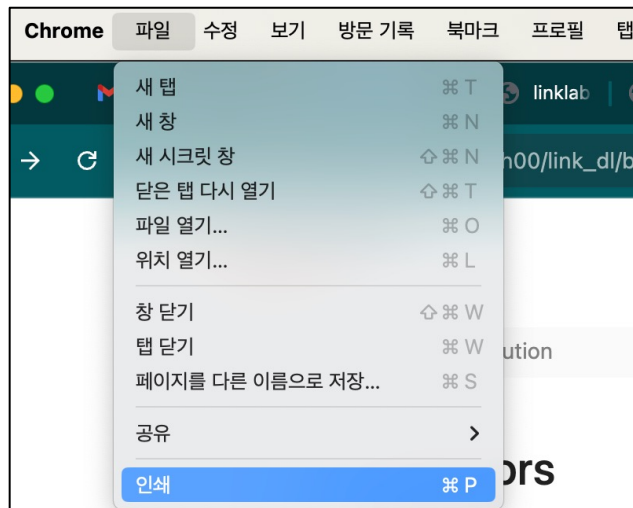
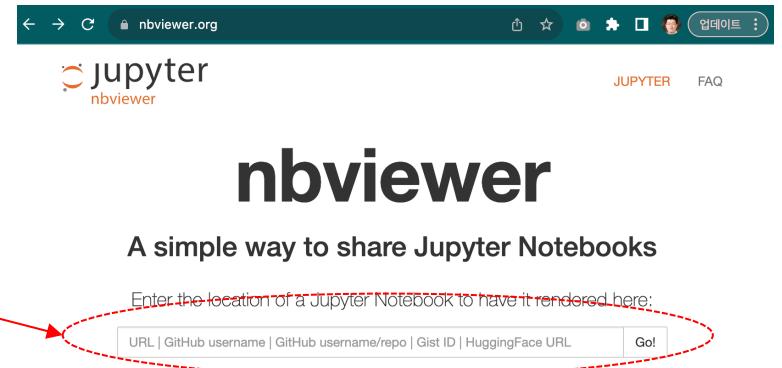
- Kaggle에 로그인 후 "Submit Prediction" 기능을 통한 submission.csv 제출
- LeaderBoard에 등록된 나의 점수 및 위치
스크린 이미지 캡처하여 Jupyter Notebook에 넣기
 - 캡처 이미지를 클라우드에 업로드하여 해당 그림의 URL 생성필요



요구사항

◆ jupyter notebook을 PDF로 인쇄하기

- 1) 로컬에 저장된 hw1_1.ipynb 파일을 클라우드에 업로드 하기
- 2) 업로드한 파일에 대한 URL 얻어오기
 - 업로드한 파일을 Public Access 가능한지 점검 필요
 - 예시: https://github.com/linklab/link_dl/blob/main/_02_homeworks/_01_tensor_handling/hw1_1.ipynb
- 3) 해당 URL을 다음 사이트의 텍스트 필드에 넣기
 - <https://nbviewer.org/>
- 4) nbviewer 사이트에서 보여지는 자신의 숙제 확인하기
- 5) 브라우저에서 PDF 인쇄



요구사항

◆ Wandb 페이지 생성 PDF 인쇄

- x축 Epoch에 대하여, y축에 Training loss 변화를 보여주는 그래프 제시
- x축 Epoch에 대하여, y축에 Validation loss 변화를 보여주는 그래프 제시
- 위 두 그래프를 포함하고 있는 Wandb 제출페이지를 PDF 인쇄

요구사항

◆ 보고서(Report) 내용에 대한 요구사항

- 프로그램 코드가 실행되는 것을 확인할 수 있도록 `jupyter notebook` 내 셀별로 출력 값들을 정확하게 나타낼 것
 - 모든 코드에 대한 출력이 잘 나와 있는지 확인함
- 핵심 코드라고 생각되는 것들에 대해서 주석(한글 또는 영문)을 넣을 것
 - 다다익선
- 코딩을 통하여 자신이 취득한 기술적 사항/고찰 내용을 생각한 바를 제시할 것
 - 다다익선
- [NOTE] 숙제 마지막에는 "숙제 후기" 라는 섹션 제목으로 본 숙제를 한 이후의 느낀점, 하고싶은 말, 또는 불평 등을 반드시 넣을 것
 - "숙제 후기" 섹션이 없으면 감점 처리

요구사항

◆ 제출형태 및 방법

- Jupyter Notebook 파일명: [homework2.ipynb](#) → PDF 생성
 - 요구사항 1~4에 대해 섹션 제목을 정확히 넣고, 섹션이 잘 구분되도록 작성하기
- Wandb 페이지 구성 → PDF 생성
- 숙제 제출 방식
 - <http://el2.koreatech.ac.kr> 의 "딥러닝및실습" 교과목 "과제" 메뉴
 - 게시물 본문에는 앞 페이지에서 설명한 <https://nbviewer.org/> 로 시작하는 페이지의 PDF 업로드
 - 게시물 본문에 Wandb 페이지 PDF 업로드
 - 즉, 총 2개 PDF 업로드
- 숙제에 관한 질의/응답
 - <http://el2.koreatech.ac.kr> 의 "딥러닝및실습" 교과목 "Q&A" 게시판 활용

◆ 제출기한

- 2025년 10월 18일 (토) 23시 59분
- 지각 제출은 받지 않습니다 (0점 처리)