데이터 전처리

songs 불러와 확인

크롤링하고 전처리가 완료된 가사와 장르들은 csv 파일로 저장되어 있습니다.
csv를 데이터프레임으로 불러오기 위해 pandas를 사용하겠습니다.
판다르를 임포트한 다음 head() 로 출력해보겠습니다.

In [112…

```
# 판다스 임포트
import pandas as pd
# csv 파일 불러오기
songs = pd.read_csv('songs_fin_mod.csv')
# drop genre '포크', '국악'
# songs = songs[songs['genre'] != '포크']
# songs = songs[songs['genre'] != '국악']
# head 출력
songs.head()
```

Out[112…

	genre	song_id	artist_id	song_name	artist_name	lyric
0	포크	31263577	468244	매트리스	10CM	오늘밤 너는 나와 이불 속에 들어가 아무것도 하지 말고 그냥 바라보다가 웃음을 참지...
1	포크	8194007	968452	오늘	오왠 (O.WHEN)	새벽4시 잠들지 않아 돌아갈 수 없는 시간들을 생각하곤 해 습관처럼 마음이 아려...
2	포크	30657311	792022	나의 사춘기에게	볼빨간사춘기	나는 한때 내가 이 세상에 사라지길 바랬어 온 세상이 너무나 캄캄해 매일 밤을 울던...
3	포크	30611680	468244	폰서트	10CM	이건 세상에서 제일 비싼 단독 공연 가수는 나고 관객은 너 하나 화려한 막이 이제 ...
4	포크	9620473	792022	나만 안되는 연애	볼빨간사춘기	왠지 오늘따라 마음이 아픈지 했더니 오늘은 그대가 날 떠나가는 날이래요 왜 항상 나...

In [113…

```
songs.shape
```

Out[113…

```
(2637, 6)
```

장르별 곡 개수 확인

장르를 예측하기 위한 머신러닝을 할 때에는 장르별로 데이터 개수가 균일하게 들어가는 것이 중요합니다.
대략 3:2 비율을 넘어서면(데이터 개수가 가장 많은 것이 가장 적은 것의 1.5배 이상) 균형이 깨졌다고 판단합니다.

In [114…

```
# genre별 갯수 확인 - 6:4 비율 맞는지 체크
songs.genre.value_counts()
```

Out[114…

```
발라드    559
힙합     456
댄스     447
트로트    435
포크     432
```

```
국악      308
Name: genre, dtype: int64
```

데이터가 가장 적은 것은 435개의 트로트, 가장 많은 것은 559개의 발라드입니다.
이정도면 균형이 깨지지 않았다고 판단하고 진행해도 좋습니다.

텍스트(가사) 데이터 정제

이제 2000여개의 데이터들을 확인하고, 가사(lyric) 컬럼을 전처리(정제)하겠습니다.
한글, 영어, 숫자를 제외한 특수문자나 기호는 제거하고, 모든 영문은 소문자로 바꾼 다음, 두 칸 이상 공백이 있다면 한 칸으로 줄여주겠습니다.
그러려면 전처리를 담당한 함수를 만들어야 합니다. 정규표현식을 사용하기 위해 re라는 라이브러리를 임포트하고, 전처리 함수 preprocess()를 만들겠습니다.

In [115…

```
# 정규식 저치 라이브러리 임포트
import re

# 전처리 함수 만들기
def preprocess(text):
  # 한글, 영어, 숫자만 남기기
  text = re.sub(r'[^a-zA-Z0-9가-힣]', ' ', text)
  # 영문을 소문자로
  text = text.lower()
  # 두 칸 이상 공백을 한 칸으로
  text = re.sub(r'\s+', ' ', text)
  return text
```

위에서 만든 함수를 이용해 모든 가사들을 전처리해줍시다.
apply를 이용하면 간편합니다.
전처리된 데이터프레임을 새로운 csv 파일로 저장해주겠습니다.

In [116…

```
# 전처리 함수로 모든 가사를 전처리하기
songs['lyric'] = songs['lyric'].apply(preprocess)
# 저장하기
songs.to_csv('songs_preprocessed.csv', index=False)
```

가사 토큰화

토큰화는 문장을 구성성분별로 나누는 것을 의미합니다. 예를 들어 '나는 밥을 먹습니다' 라는 문장을 토큰화하면
['나', '는', '밥', '을', '먹습니다'] 가 됩니다.
okt는 명사만 추출하기 기능 등을 포함해 다양한 토큰화 관련 기능을 제공합니다.
토큰화 된 배열을 return하는 tokenize라는 함수를 만들겠습니다.

In [117…

```
# 라이브러리 임포트
from konlpy.tag import Okt, Kkma, Hannanum, Komoran
okt = Okt()
kkma = Kkma()
hannanum = Hannanum()
komoran = Komoran()

# 함수 만들기
def tokenize_okt(text):
  return okt.morphs(text)
```

```
def tokenize_kkma(text):
  return kkma.morphs(text)
def tokenize_hannanum(text):
  return hannanum.morphs(text)
def tokenize_komoran(text):
  return komoran.morphs(text)
```

로지스틱 회귀모델을 활용한 분류

이제 본격적으로 분류 작업을 해보겠습니다. 순서는 아래와 같습니다.

1. train, test 세트 분리
2. cross validation으로 여러 매개변수 조합을 테스트하며 학습하기
3. 학습 결과 시각화하기

1. train, test 분리하기

sklearn 라이브러리에서 제공하는 train_test_split 함수를 사용하면 간편하게 train, test 분리를 수행할 수 있습니다.
data(가사)와 target(장르)을 집어넣고, test set의 사이즈를 지정하고, 재현성을 위해 random_state를 지정하겠습니다.

In [118…

```
# 라이브러리 임포트
from sklearn.model_selection import train_test_split

# train, test 분리
X_train, X_test, y_train, y_test = train_test_split(songs['lyric'], songs['ge
```

2. GridSearchCV 5-fold cross validation을 사용해 매개변수 조합들을 바꿔가며 학습해보기

머신러닝 모델에는 하이퍼파라미터들이 들어갑니다.
어떤 규제를 사용할 것인가, 어떤 토큰화 함수를 사용할 것인가, 규제 강도는 얼마로 할 것인가 등 다양한 하이퍼 파라미터들이 있습니다.

원래대로라면 하이퍼파라미터들을 바꿔가며 모델을 수십 번 학습해야 하지만, 이를 자동화해주는 라이브러리가 있습니다.
GridSearch와 Pipeline을 통해 하이퍼파라미터를 자동으로 바꿔가며 Logistic Regression을 학습해보겠습니다.

학습 파라미터 설정

우리는 param_grid라는 것을 설정해줘야 합니다. 어떤 파라미터들을 바꿔가며 실험해볼지를 지정합니다.
vect**tokenizer(토큰화 함수)로 우리가 위에서 만든 tokenize 함수와 str.split을 바꿔가며 실험해보고,**
clfpenalty(규제)로는 l1과 l2를 바꿔가며 실험해 보겠습니다. 마찬가지로 clf__C(규제값)도 0.1부터 1000까지 바꿔가며 실험합니다.

In [119…

```
# 필요한 라이브러리 임포트
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.feature_extraction.text import TfidfVectorizer

# 바꿔가며 학습 할 파라미터들 목록
param_grid = [{'vect__ngram_range': [(1, 1)],
               'vect__tokenizer': [
                   str.split, tokenize_okt,
                   tokenize_kkma, tokenize_hannanum, tokenize_komoran],
               'regressor__penalty': [('l2')],
               'regressor__C': (1, 10, 100)}]
```

파이프라인 구축

파이프라인은 두 단계로 매우 간단합니다. 문장을 벡터화한다 > 로지스틱 회귀를 수행한다.
이 둘을 파이프라인 안에 배열로 넣어주 두 작업을 자동으로 연달아 수행합니다.

In [120…

```
# TfidfVectorizer: 문장을 토큰화, 벡터화하는 함수

tfidf = TfidfVectorizer(preprocessor=None, lowercase=False, stop_words=None, 

lr_tfidf = Pipeline([('vect', tfidf),
                     ('regressor', LogisticRegression())])
```

그리드서치 수행

이제 우리가 정한 파라미터들을 바꿔가며 파이프라인을 수행합니다.
파라미터를 한 번 바꿀 때마다 5번씩 cross-validation을 수행할 것입니다.
지금 바꿔볼 파라미터는 토큰화함수 2개, 규제방식 2개, 규제값 5개이므로 2x2x5 = 20 종류의 파라미터 조합을 실험해볼 것입니다.
각 조합마다 5-fold cross-valation이 수행되므로, 모델 fit은 총 100번 수행됩니다.

In [121…

```
gs_lr_tfidf = GridSearchCV(lr_tfidf, param_grid, scoring='accuracy', cv=5, ve
gs_lr_tfidf.fit(X_train, y_train)
```

```
Fitting 5 folds for each of 15 candidates, totalling 75 fits
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
```

```
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
```

```
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
```

```
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
```

```
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
/Users/tj/opt/anaconda3/envs/tensorflow/lib/python3.7/site-packages/sklearn/li
near_model/_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (sta
tus=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regress
```

```
ion
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

최적의 매개변수 조합 확인하기

모델은 이미 최적의 파라미터 조합으로 학습을 끝마쳤습니다. 어떤 파라미터들을 선정했는지 확인해봅시다.

In [ ]:
```
# 최적의 매개변수 조합 print
print('최적의 매개변수 조합:', gs_lr_tfidf.best_params_)
```

```
최적의 매개변수 조합: {'regressor__C': 1, 'regressor__penalty': 'l2', 'vect__ngram_
range': (1, 1), 'vect__tokenizer': <function tokenize_kkma at 0x7f8afae54dd0>}
```

3. 학습 결과를 확인하고 시각화하기

정확도 출력

검증 정확도와 테스트 정확도를 확인해보겠습니다.
모델(gs_lr_tfidf)에서 best*score* 를 출력해보면 validation 정확도가 나옵니다.
best_estimator_를 변수에 넣고, 해당 변수를 이용해 score()를 출력하면 test 정확도 역시 확인할 수 있습니다.

In [ ]:
```
print('검증 정확도:', gs_lr_tfidf.best_score_)

clf = gs_lr_tfidf.best_estimator_
print('test 정확도:', clf.score(X_test, y_test))
```

```
검증 정확도: 0.6410554873861602
test 정확도: 0.6799242424242424
```

예측 결과 시각화

confusion_matrix는 예측 결과를 다각도로 분석하도록 도와줍니다.
confusion_matrix를 출력하면 기본적으로 배열을 출력해주지만, 우리는 seaborn을 임포트해서 더 멋진 시각화 그래프를 그려보겠습니다.

In [ ]:
```
from sklearn.metrics import confusion_matrix
# confusion matrix
labels = ['발라드', '힙합', '댄스', '트로트']

y_pred = gs_lr_tfidf.predict(X_test)
cm = confusion_matrix(y_test, y_pred, labels=labels)

print(cm)
import seaborn as sns
import matplotlib.pyplot as plt
# set font Apple Gothic
plt.rc('font', family='AppleGothic')
# visualize confusion matrix
plt.figure(figsize=(9,9))
ax= plt.subplot()

sns.heatmap(cm, annot=True, fmt="d", linewidths=.5)
# label
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
ax.set_xticklabels(labels)
```

```
ax.set_yticklabels(labels)

plt.show()

# 4를 제일 못 맞춤, 0도 그닥
```

```
[[93  0  4  5]
 [ 5 61 14  2]
 [11 10 56  1]
 [ 9  1  3 71]]
```