

CUAI 스터디 캐글 데이콘 우수 노트북 분석 - 4팀

2022.05.31

발표자 : 최서희

스터디원 소개 및 만남 인증



05.23 6:00 PM
스터디

스터디원 : 최서희

스터디원 : 박상우

스터디원 : 정재희

5월 16, 23, 30일 스터디 내용

펭귄 몸무게 예측 경진대회

데이콘 베이직 Basic | 정형 | RMSE

₩ 상금 : 참가시 최소 50 XP, 특별상 데이콘 후드

🕒 2021.12.27 ~ 2022.01.07 17:59 [+ Google Calendar](#)

👤 749명 📅 마감

영어 음성 국적 분류 AI 경진대회

음성 분류 | logloss | 월간 데이콘 16

₩ 상금 : 100만원

🕒 2021.05.18 ~ 2021.07.02 17:59 [+ Google Calendar](#)

👤 410명 📅 마감

주식 종료 가격 예측 Pre Competiton(with.데이커)

주식 | 예측 | 시계열 | 종가 | 종료가격

₩ 상금 : DACON RANKING POINT, XP

🕒 2021.08.17 ~ 2021.09.26 23:59 [+ Google Calendar](#)

👤 475명 📅 마감

스터디 내용

펭귄 몸무게 예측 경진대회

데이콘 베이직 Basic | 정형 | RMSE

🕒 상금 : 참가시 최소 50 XP, 특별상 데이콘 후드

🕒 2021.12.27 ~ 2022.01.07 17:59 [+ Google Calendar](#)

👤 750명 📅 마감



참여



```
train.isnull().sum()
```

id	0
Species	0
Island	0
Clutch Completion	0
Culmen Length (mm)	0
Culmen Depth (mm)	0
Flipper Length (mm)	0
Sex	3
Delta 15 N (o/oo)	3
Delta 13 C (o/oo)	3
Body Mass (g)	0
dtype:	int64

• 결측치

- Sex : 펭귄의 성별
- Delta 15 N (o/oo) : 토양에 따라 변화하는 안정 동위원소 15N:14N의 비율
- Delta 13 C (o/oo) : 먹이에 따라 변화하는 안정 동위원소 13C:12C의 비율

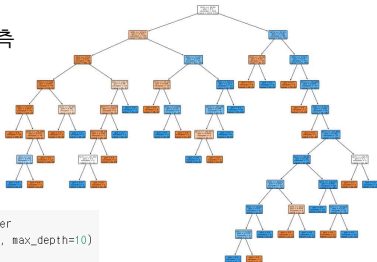
스터디 내용

```
## Delta값의 결측치에는 train 데이터의 평균 입력  
# loc은 행 열 검색  
x_train.loc[:, 'Delta 15 N (o/oo)'] = x_train['Delta 15 N (o/oo)'].fillna(train['Delta 15 N (o/oo)'].mean())  
x_train.loc[:, 'Delta 13 C (o/oo)'] = x_train['Delta 13 C (o/oo)'].fillna(train['Delta 13 C (o/oo)'].mean())  
x_test.loc[:, 'Delta 15 N (o/oo)'] = x_test['Delta 15 N (o/oo)'].fillna(train['Delta 15 N (o/oo)'].mean())  
x_test.loc[:, 'Delta 13 C (o/oo)'] = x_test['Delta 13 C (o/oo)'].fillna(train['Delta 13 C (o/oo)'].mean())
```

Delta 15 N (o/oo) 와 Delta 13 C (o/oo) :
mean() - 평균값으로 대체해줌.

스터디 내용

Sex : DecisionTreeClassifier 예측



```
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier(random_state=1, max_depth=10)
model.fit(x_train, y_train)
print('DTree score: ', model.score(x_train,y_train))
```

예측한 성별 값으로 결측치 대체

```
import numpy as np
```

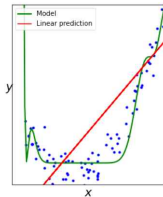
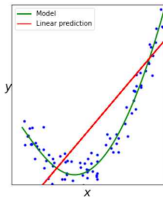
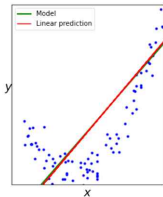
```
test['Sex'].iloc[np.where(test['Sex'].isnull()==True)] = model.predict(x_test)
```

스터디 내용

1. 다항회귀

예를 들어 degree=3, feature=2라 하면

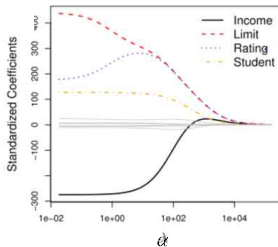
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$



스터디 내용

2. 릿지 회귀

$$J(\theta) = \text{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^n \theta_i^2$$



스터디 내용

```
model = Pipeline([('scl', MinMaxScaler()),
                  ('poly', PolynomialFeatures()),
                  ('fit', Ridge())])

cv = KFold(random_state = 50, n_splits = 10, shuffle = True)

params = {
    'poly__degree' : [2, 3, 4, 5],
    'fit__alpha' : [0, 0.5, 1, 1.5, 2],
    'fit__random_state' : [42]
}

# Pipe, params 설정
start = time.time()
model_ridge = GridSearchCV(model,
                           param_grid = params,
                           cv = cv,
                           n_jobs = -1)
```

```
model_ridge.fit(X_train, y_train)
```

```
GridSearchCV(cv=KFold(n_splits=10, random_state=50, shuffle=True),
             estimator=Pipeline(steps=[('scl', MinMaxScaler()),
                                       ('poly', PolynomialFeatures()),
                                       ('fit', Ridge())]),
             n_jobs=-1,
             param_grid={'fit__alpha': [0, 0.5, 1, 1.5, 2],
                         'fit__random_state': [42],
                         'poly__degree': [2, 3, 4, 5]})
```

```
ridge_best_param = model_ridge.best_params_
print(ridge_best_param)
print('best score : {}'.format(model_ridge.best_score_))
```

```
{'fit__alpha': 2, 'fit__random_state': 42, 'poly__degree': 2}
best score : 0.8263030449314644
```

스터디 내용

```
submission['Body Mass (g)'] = best_model_ridge.predict(test_dt)  
submission
```

	id	Body Mass (g)
0	0	4011.574526
1	1	5671.565630
2	2	3468.228390
3	3	3543.160871
4	4	3445.404592
...
223	223	3843.497097
224	224	4794.423906
225	225	4555.545847
226	226	3839.377977
227	227	4138.180862

스터디 내용

영어 음성 국적 분류 AI 경진대회

음성 분류 | logloss | 월간 데이콘 16

₩ 상금 : 100만원

📅 2021.05.18 ~ 2021.07.02 17:59

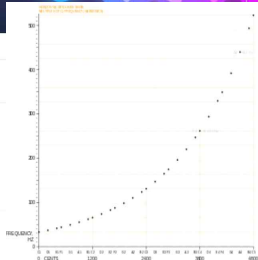
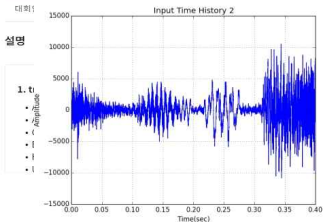
[+ Google Calendar](#)

👤 411명 📅 마감

librosa 패키지



참여중



스터디 내용

주식 종료 가격 예측 Pre Competiton(with.데이커)

주식 | 예측 | 시계열 | 종가 | 종료가격

₩ 상금 : DAICON RANKING POINT, XP

📅 2021.08.17 ~ 2021.09.26 23:59

[+ Google Calendar](#)

👤 475명 🗄 마감



참여

prophet()

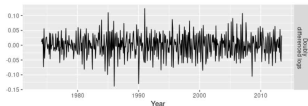
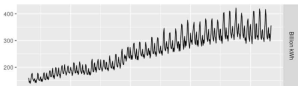
ARIMA

```
globals()['m_{}'.format(i)] = Prophet()
```

```
locals()['m_{}'.format(i)].fit(locals()['df_{}'.format(i)])
```

```
globals()['future_prices_{}'.format(i)] = locals()['m_{}'.format(i)].make_future_dataframe(periods=21)
```

```
globals()['forecast_{}'.format(i)] = locals()['m_{}'.format(i)].predict(locals()['future_prices_{}'.format(i)])
```



감사합니다!