

■ 빅데이터분석기사 실기 7 회 - Type3 성별 예측 정답 및 코드

키, 몸무게, 발사이즈로 성별을 분류하는 로지스틱 회귀 모델을 학습하고, Weight의 오즈비와 로그우도, 예측 오차율을 계산합니다.

```
import pandas as pd  
import statsmodels.api as sm  
import numpy as np  
from sklearn.model_selection import train_test_split  
  
# 데이터 로딩 및 전처리  
df = pd.read_csv("07.03.01-gender_prediction_dataset.csv")  
df['Gender'] = df['Gender'].map({'Male': 0, 'Female': 1})  
  
# 변수 설정  
X = df[['Height', 'Weight', 'ShoeSize']]  
y = df['Gender']
```

```
# 훈련/테스트 분할  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, stratify=y,  
random_state=42)
```

```
# 로지스틱 회귀 모델 학습  
X_train = sm.add_constant(X_train)  
model = sm.Logit(y_train, X_train).fit()  
print(model.summary())
```

import statsmodels.api as sm
from statsmodels import Logit
X_train = sm.add_constant(X_train)
model = sm.Logit(y_train, X_train).fit()

```
# (1) Weight 의 오즈비 계산  
odds_ratio = np.exp(model.params['Weight'])  
print("Weight 오즈비:", odds_ratio)
```

```
# (2) 모델의 로짓우도  
log_likelihood = model.llf  
print("로그우도:", log_likelihood)
```

```
# (3) 예측 정확도 및 오차율
```

$$\text{odds} = \exp(\text{Weight}).$$

model.params['Weight']

ML 예측이다. 0 아니면 1

그리고 확률이 0.5 이상인 것 (= 1)

아래인 것 (0) 은 예측하는 것.

이제 예상과 실제를 비교해 예측률을 구함.

```
X_test = sm.add_constant(X_test)  
y_pred = (model.predict(X_test) >= 0.5).astype(int)  
error_rate = round(np.mean(np.abs(y_pred - y_test)), 4)  
print("오차율:", error_rate)
```

*- error_rate =

$(\text{pred} \neq y_{\text{test}}) / \text{mean}()$

Sm. add_constant:

여기서 y_{test} 이 들어온다 |

설명 O

증거에 대한 확률을 구함.

예측