회귀 모델 평가지표 (Regression Model Evaluation Metrics)

- 회귀 모델은 연속적인 숫자 값을 예측하는 모델이므로, 예측값과 실제값 사이의 오차(Error)를 기반으로 모델의 성능을 평가
- 다양한 평가지표가 있으며, 각 지표는 모델의 특정 측면을 강조

1. **MSE (Mean Squared Error, 평균 제곱 오차)**:
    - **개념**: 예측값과 실제값 차이의 제곱에 대한 평균입니다.
    - **수식**: $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
    - **특징**: 오차가 클수록 제곱되어 더 큰 페널티를 부여합니다. 따라서 이상치(Outlier)에 민감하게 반응합니다. 단위는 종속변수 단위의 제곱입니다.
2. **RMSE (Root Mean Squared Error, 평균 제곱근 오차)**:
    - **개념**: MSE에 제곱근을 취한 값입니다.
    - **수식**: $RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$
    - **특징**: MSE의 단위를 종속변수와 동일하게 만들어 해석을 용이하게 합니다. MSE와 마찬가지로 이상치에 민감합니다.
3. **MAE (Mean Absolute Error, 평균 절대 오차)**:
    - **개념**: 예측값과 실제값 차이의 절대값에 대한 평균입니다.
    - **수식**: $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$
    - **특징**: 오차의 크기에 비례하여 페널티를 부여하므로, MSE/RMSE보다 이상치에 덜 민감합니다. 단위는 종속변수와 동일합니다.
4. **MAPE (Mean Absolute Percentage Error, 평균 절대 백분율 오차)**:
    - **개념**: 예측값과 실제값 차이의 절대값을 실제값으로 나눈 백분율에 대한 평균입니다.
    - **수식**: $MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{y_i - \hat{y}_i}{y_i}| \times 100\%$
    - **특징**: 오차를 백분율로 나타내므로, 서로 다른 스케일의 데이터셋이나 모델 간 성능 비교에 유용합니다. 하지만 실제값이 0에 가까울 경우 무한대로 발산할 수 있는 단점이 있습니다.
5. **R^2 (R-squared, 결정계수)**:
    - **개념**: 모델이 종속변수 분산의 몇 퍼센트를 설명하는지를 나타내는 지표입니다.
    - **수식**: $R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$
        - SS_{res}: 잔차 제곱합 (모델이 설명하지 못하는 분산)
        - SS_{tot}: 총 제곱합 (종속변수의 총 분산)
    - **특징**: 0과 1 사이의 값을 가지며, 1에 가까울수록 모델이 데이터를 잘 설명한다고 해석합니다. 독립변수의 개수가 늘어나면 R^2 값은 항상 증가하는 경향이 있어, 과적합된 모델을 선택할 위험이 있습니다.
6. **Adjusted R^2 (수정된 결정계수)**:
    - **개념**: R^2의 단점을 보완하기 위해 독립변수의 개수와 표본 크기를 고려하여 R^2 값을 조정한 지표입니다.

  - **수식**: $Adjusted R^2 = 1 - (1 - R^2)\frac{N-1}{N-p-1}$
    - N: 표본 크기
    - p: 독립변수의 개수
  - **특징**: 새로운 독립변수가 모델의 설명력을 충분히 높이지 못하면 Adjusted R^2 값은 감소할 수 있습니다. 따라서 모델에 불필요한 변수가 추가되는 것을 방지하는 데 유용합니다.

적용 가능한 상황

- 회귀 모델의 예측 성능을 객관적으로 평가하고, 여러 모델 중 최적의 모델을 선택할 때.
- 모델의 오차 특성(이상치에 민감한지, 백분율 오차가 중요한지 등)에 따라 적절한 평가지표를 선택하여 사용합니다.

주의사항

- **지표 선택**: 어떤 평가지표가 가장 중요한지는 비즈니스 목표나 문제의 특성에 따라 달라집니다. 예를 들어, 이상치 예측이 중요하다면 MSE/RMSE가, 이상치에 덜 민감한 평가가 필요하면 MAE가 적합할 수 있습니다.
- **단위**: MSE, RMSE, MAE는 종속변수와 동일하거나 제곱된 단위를 가지므로, 스케일이 다른 모델 간 비교 시 주의해야 합니다. MAPE는 백분율이므로 스케일에 독립적입니다.
- **과적합**: R^2는 독립변수 개수에 따라 항상 증가하므로, 모델의 복잡도를 고려한 Adjusted R^2를 함께 보는 것이 좋습니다.

```python
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import statsmodels.api as sm
import pandas as pd

# 1. 실제값과 예측값 데이터 준비
y_true = np.array([10, 12, 15, 13, 18, 20, 22, 25, 23, 28])
y_pred = np.array([11, 11, 14, 14, 17, 21, 21, 26, 24, 27])

# 2. MSE, RMSE, MAE 계산
mse = mean_squared_error(y_true, y_pred)
rmse = np.sqrt(mse)
mae = mean_absolute_error(y_true, y_pred)

print(f"MSE: {mse:.3f}")    # 1.000
print(f"RMSE: {rmse:.3f}")  # 1.000
print(f"MAE: {mae:.3f}")    # 1.000

# 3. MAPE 계산 (수동 구현)
def calculate_mape(y_true, y_pred):
    # 실제값이 0인 경우를 처리하기 위해 작은 값(epsilon)을 더해줌
    return np.mean(np.abs((y_true - y_pred) / (y_true + np.finfo(float).eps))) *
100

mape = calculate_mape(y_true, y_pred)
print(f"MAPE: {mape:.3f}%") # 5.971%

# 4. R-squared (R²) 계산
r2 = r2_score(y_true, y_pred)
```

```
print(f"R-squared (R²): {r2:.3f}") # 0.969

# 5. Adjusted R² 계산 (statsmodels OLS summary에서 확인 가능)
# 예시를 위해 간단한 OLS 모델을 가정
# 실제 모델링에서는 X_train, y_train으로 학습 후 X_test, y_test로 예측하여 평가
# 여기서는 y_true를 종속변수, y_pred를 예측값으로 사용하여 R2 계산
# Adjusted R2는 모델의 독립변수 개수(p)와 표본 크기(N)가 필요
N = len(y_true) # 표본 크기
p = 1 # 독립변수 개수 (예시를 위해 1로 가정)

adjusted_r2 = 1 - (1 - r2) * (N - 1) / (N - p - 1)
print(f"Adjusted R²: {adjusted_r2:.3f}") # 0.965

# statsmodels OLS summary 예시 (실제 모델링 시 사용)
X_example = sm.add_constant(np.arange(N)) # 가상의 독립변수
model_sm = sm.OLS(y_true, X_example)
results_sm = model_sm.fit()
print(results_sm.summary()) # R-squared (Adj.) 항목 확인
'''
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.945
Model:                            OLS   Adj. R-squared:                  0.938
Method:                 Least Squares   F-statistic:                     136.5
Date:                Mon, 13 Oct 2025   Prob (F-statistic):           2.63e-06
Time:                        21:10:55   Log-Likelihood:                -17.118
No. Observations:                  10   AIC:                             38.24
Df Residuals:                       8   BIC:                             38.84
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          9.9273      0.881     11.272      0.000       7.896      11.958
x1             1.9273      0.165     11.682      0.000       1.547       2.308
==============================================================================
Omnibus:                        4.080   Durbin-Watson:                   2.897
Prob(Omnibus):                  0.130   Jarque-Bera (JB):                2.005
Skew:                          -1.095   Prob(JB):                        0.367
Kurtosis:                       2.882   Cond. No.                         10.2
==============================================================================
'''
```

결과 해석 방법

- **MSE, RMSE, MAE**: 값이 작을수록 모델의 예측 오차가 작다는 의미이므로 좋은 모델입니다.
- **MAPE**: 백분율 오차이므로, 10% 미만이면 매우 좋음, 10~20%는 좋음, 20~50%는 보통, 50% 이상은 나쁨으로 해석하는 경우가 많습니다.
- **R^2**: 1에 가까울수록 모델이 종속변수의 변동성을 잘 설명한다고 해석합니다.
- **Adjusted R^2**: R^2와 함께 보며, 모델에 추가된 변수가 모델의 설명력을 유의미하게 높였는지 판단하는 데 사용합니다. Adjusted R^2가 R^2보다 현저히 낮거나, 새로운 변수 추가 시 감소한다면 해당 변수는 불필요

할 수 있습니다.

장단점 및 대안

- **장점**:
  - 각 지표는 모델의 예측 정확도, 오차의 크기, 설명력 등 다양한 측면을 정량적으로 평가할 수 있습니다.
  - 모델 간 성능 비교 및 최적 모델 선택에 객관적인 기준을 제공합니다.
- **단점**:
  - 단일 지표만으로는 모델의 모든 측면을 평가하기 어렵습니다. (e.g., MSE는 이상치에 민감, MAPE는 0에 가까운 값에 취약)
  - 비즈니스 목표에 따라 어떤 지표가 더 중요한지 판단해야 합니다.
- **대안**:
  - **시각적 분석**: 잔차 플롯(Residual Plot), 실제값 vs 예측값 플롯 등을 통해 모델의 오차 패턴을 시각적으로 확인하는 것이 중요합니다.
  - **교차 검증**: 단일 테스트 세트에서의 평가가 아닌, 교차 검증을 통해 여러 번의 평가를 수행하여 모델 성능의 안정성을 확보합니다.