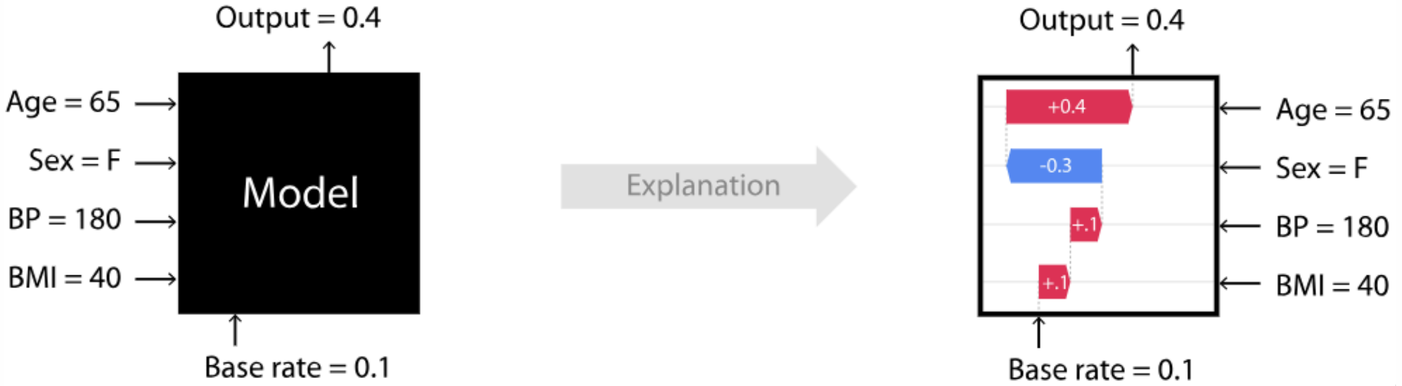
Shap value

SHAP(SHapley Additive exPlanations)

게임이론의 shapley value를 기반으로 한 모델 해석 기법

핵심 아이디어 : 모델 예측을 각 feature의 공정한 기여도로 분해하자!

SHAP



Shap value

Ex) 정훈, 영빈, 소연이 대회에서 500만원을 상금으로 받음.
→ 상금을 어떻게 배분할거냐? 누가 얼마씩 받아야 공평한가?

EDA → 데이터 전처리 → 모델링

순서	기여도 계산	정훈의 기여도
정훈 → 영빈 → 소연	정훈이의 EDA 덕분에 30점을 받음	**30**
영빈 → 정훈 → 소연	영빈이의 EDA는 40점을 받았고 정훈이의 데이터 전처리로 40점을 더 받음	**40**
영빈 → 소연 → 정훈	영빈이의 EDA와 소연이의 데이터 전처리로 90점을 받고 정훈이의 모델링으로 10점을 받음	**10**

Shap value

장점

- 정확한 기여도 계산
- 어느 모델이든 사용 가능
- 직관적으로 해석 가능하고 시각화 도구를 제공

단점

- 계산 비용이 매우 큼
- 딥러닝에서는 정확하지 않을 수 있음
- 대규모 데이터에서는 느림

Shap value

pip install shap xgboost

```
import shap
import xgboost as xgb
import pandas as pd
from sklearn.model_selection import train_test_split

data = load_breast_cancer()
X = pd.DataFrame(data.data, columns=data.feature_names)
y = data.target

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Shap value

```
1  model = xgb.XGBClassifier()
2  model.fit(X_train, y_train)
```

✓ 0.0s

XGBClassifier

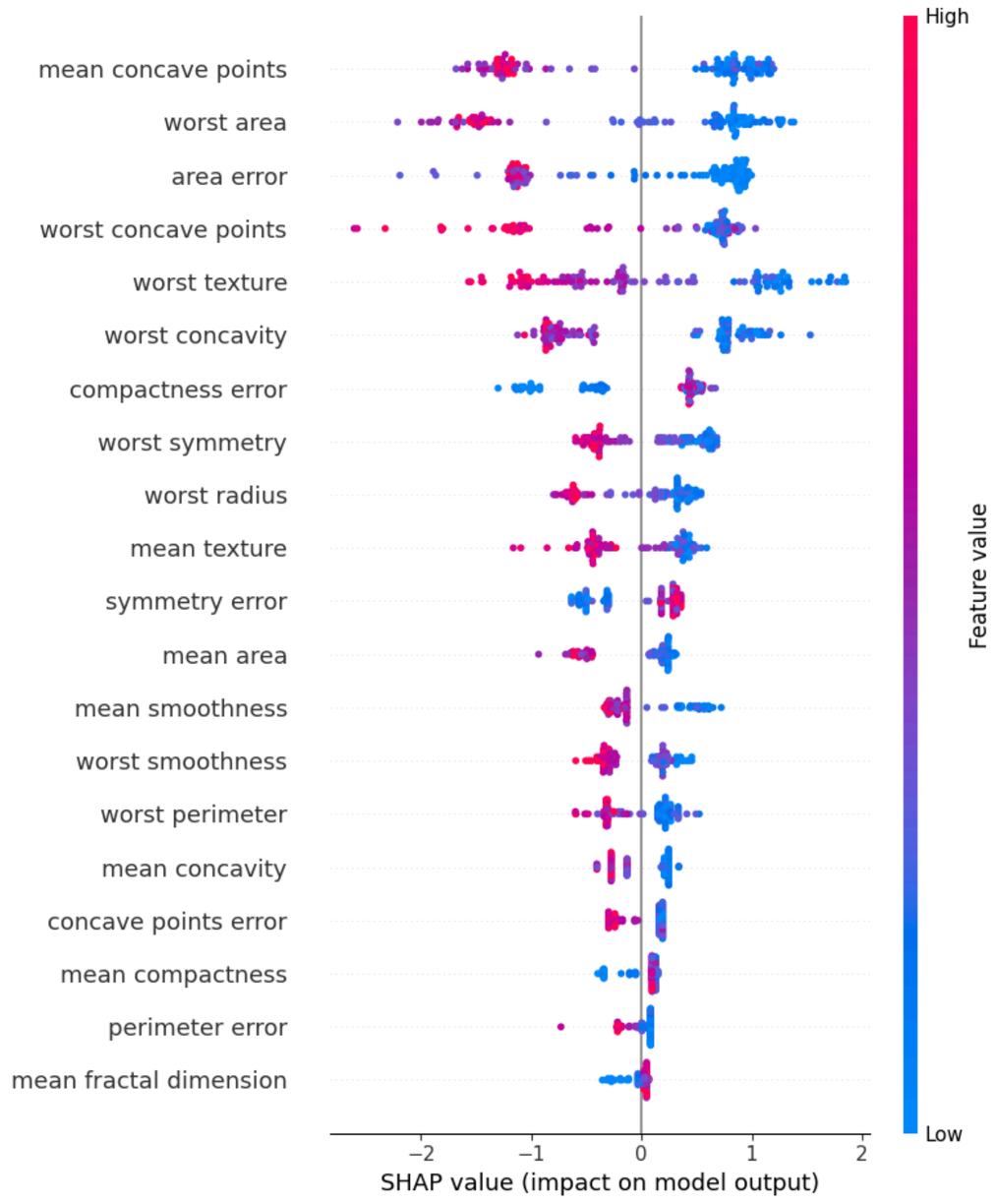
▶ Parameters

Shap value

```
explainer = shap.Explainer(model, X_train)
shap_values = explainer(X_test)

shap.summary_plot(shap_values, X_test)
```

Shap value



각 점들 : 각 feature의 각 샘플 값

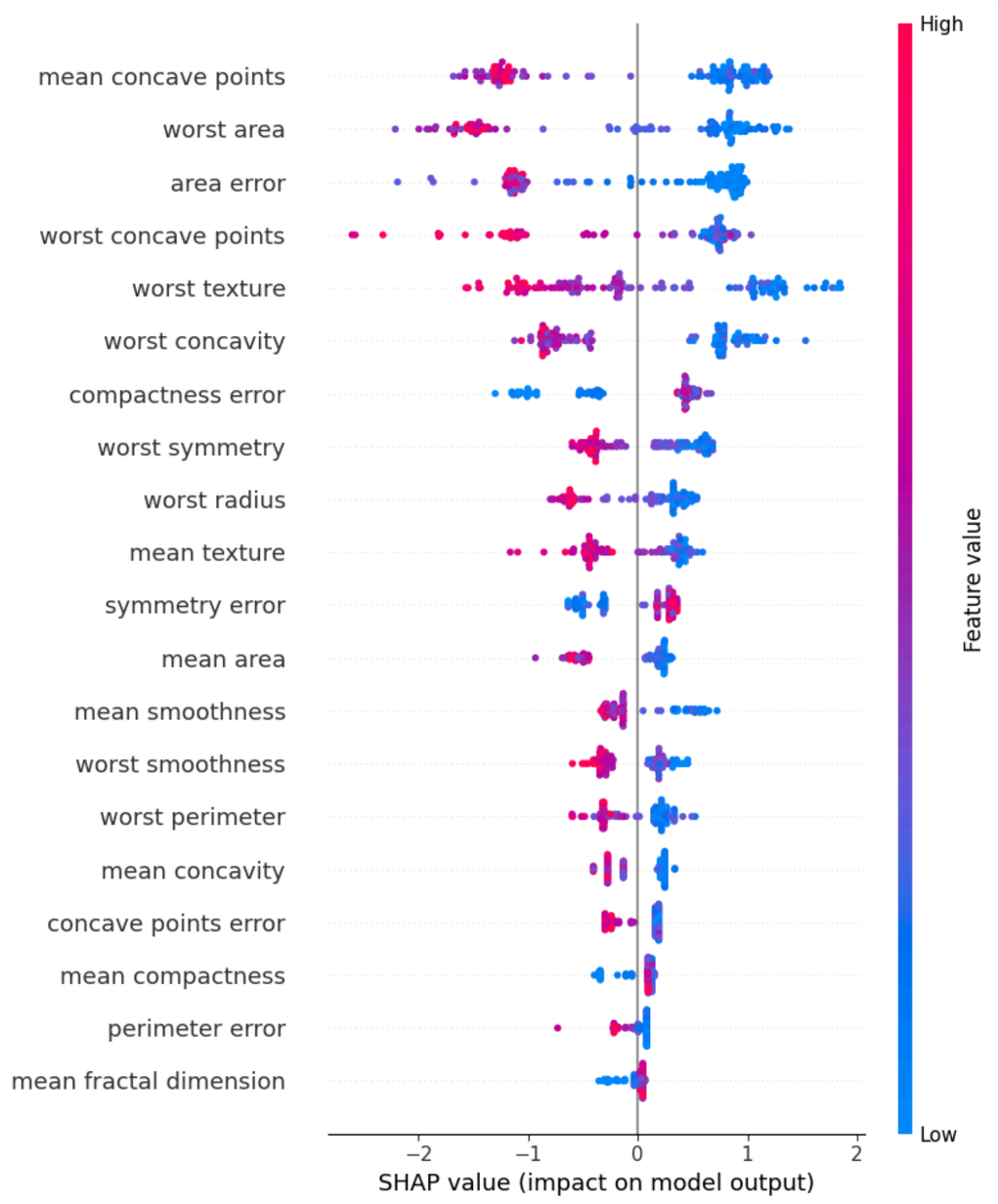
shap value : target값을 맞추는데 얼마나 영향을 줬는지

“mean_concave_points”

값이 작은 데이터들은 예측값이 큰 거에 영향을 많이 줌

값이 큰 데이터들은 예측값이 작은 거에 영향을 많이 줌

Shap value



좋은 feature

값이 골고루 퍼져있는 feature

shap value가 0에 모여있으면 제거 고려