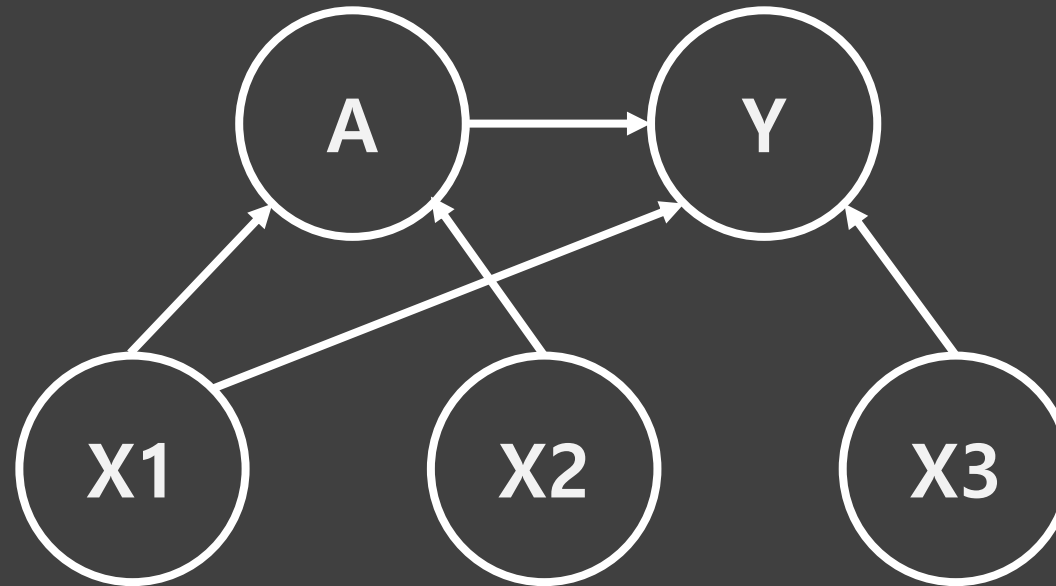


# Estimating Marketing Effects



# 목차

1 . 개 요

2 . 데 이 터 설 명

3 . 연 구 결 과

# 1. 개요

## Marketing Effect Estimating Method

### 기존 추정 방법의 문제점

Attribution

상관에 기반하여 마케팅 효과를 추정하므로  
실제 인과와는 큰 차이를 보일 수 있다.

Incrementality

인과를 보지만, 랜덤화 실험이 필요하다.



관측 자료로부터 인과를 추정하는 방법 필요

## Goal

### 본 연구의 목표

- 마케팅 연구자가 데이터만 format에 맞춰 입력하면, 다양한 인과 추론 작업을 수행해서 결과를 반환해주는 class를 정의.
- 해당 class를 이용하여 Bank marketing data에 적용.
- Simulated data를 생성하여 이에 적용하여 추정 성능을 검증.

## 2. 데이터 설명

## Bank Marketing Data

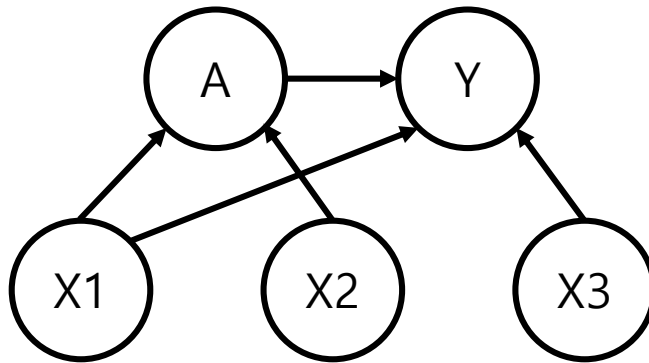
- Y(outcome) : 정기예금 가입여부
- A(treatment) : 1(marketing(phone call 수신) 횟수  $\geq 2$ ) (**Binary**)
- X(covariate) : age, job, marital, education, housing, loan, month, day\_of\_week ... (15개) -> 전처리 후 23개로 변화.
- ATE :  $E(Y^1) - E(Y^0)$
- CATE :  $\tau(x_1) = E(Y^1 - Y^0 | X_1 = x_1)$

## Synthetic Data

- 실제 데이터는 정답이 되는 ATE와 CATE를 알 수 없음.
- 실제 ATE와 CATE를 알 수 있도록 Synthetic data를 생성한다.

$$X_1, X_2, X_3, \epsilon \sim N(0, 1), \quad A \sim \text{Ber}\left(\text{logit}^{-1}(X_1 + X_2 + 1.5)\right),$$

$$Y = A + X_1 + X_3 + A * (X_1^2 + X_1 + 1) + \epsilon$$





### 3. 연구 결과

## 연구 결과 요약

본 연구의 흐름은 다음과 같이 요약될 수 있다.

- 데이터만 입력하면 유용한 causal inference 관련 결과를 제공하는 **Class의 개발**.
- **Bank marketing data**를 전처리 후, 위 class에 적용하여 분석 결과 도출.
- **Synthetic data**를 generate한 후, 위 class에 적용하여 분석 성능 평가.

## 연구 결과 - Class 개발

- Class CausalForest를 통해, binary treatment data에 대해 random forest를 이용하여 propensity score modeling을 하는 causal inference 방법을 구현하였다.
- propensity score의 추정 성능을 쉽게 체크할 수 있도록 scores / plot\_roc\_curve / importance\_plot 메서드를 제공.
- ATE\_pred / CATE\_pred 메서드를 통해 ATE 추정과 CATE 추정을 수행.

## 연구 결과 - Class 개발

```
[ ] CF = CausalForest(X,A,y)
```

```
[ ] CF.scores()
```

숨겨진 출력 표시

```
[▶] CF.plot_roc_curve()
```

숨겨진 출력 표시

```
[ ] CF.Importance_plot()
```

숨겨진 출력 표시

```
[ ] CF.ATE_pred()
```

숨겨진 출력 표시

```
[ ] CF.CATE_pred('balance')
```

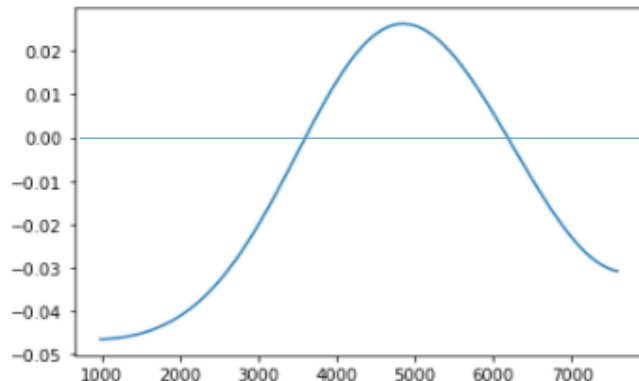
다음과 같이 간편한 입력을 통해  
유용한 결과를 얻을 수 있음.

## 연구 결과 - Bank Marketing Data

```
[ 3] CF.ATE_pred()
```

```
(('ATE, E_y_1, E_y_0'),  
 -0.03694796747816337,  
 0.4494630404397563,  
 0.4864110079179197)
```

```
[144] CF.CATE_pred('balance')
```



위 3개 메서드를 통해서는  
Random forest의 Hyperparameter와 input data 체크.

최종 결과는 ATE\_pred와 CATE\_pred로 확인.

ATE= -0.037이므로 treatment가 예금에 부정적 영향.  
Treatment 받으면 44.9% 예금, 안 받으면 48.6% 예금

CATE 함수를 보면, balance가 약 3500 ~ 6000이면  
Treatment가 긍정적인 것으로 해석됨.

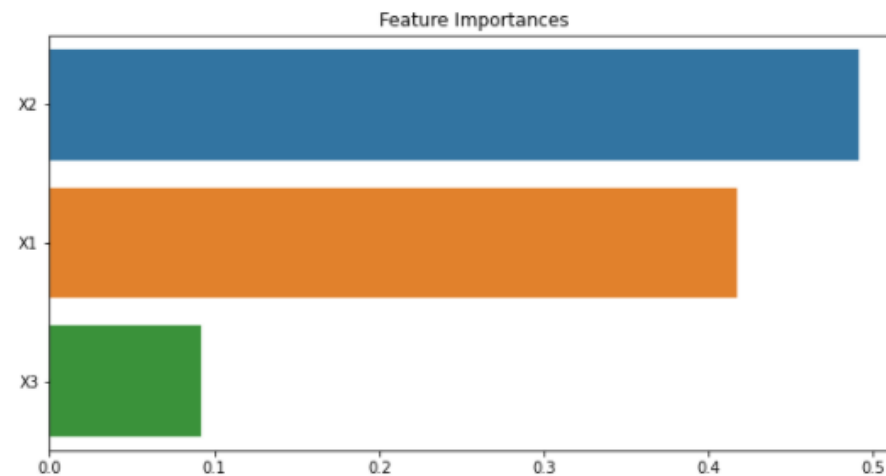
## 연구 결과 - Synthetic Data

$$X_1, X_2, X_3, \epsilon \sim N(0, 1),$$

$$A \sim \text{Ber}\left(\text{logit}^{-1}(X_1 + X_2 + 1.5)\right),$$

$$Y = A + X_1 + X_3 + A * (X_1^2 + X_1 + 1) + \epsilon$$

```
[52] CF.Importance_plot()
```



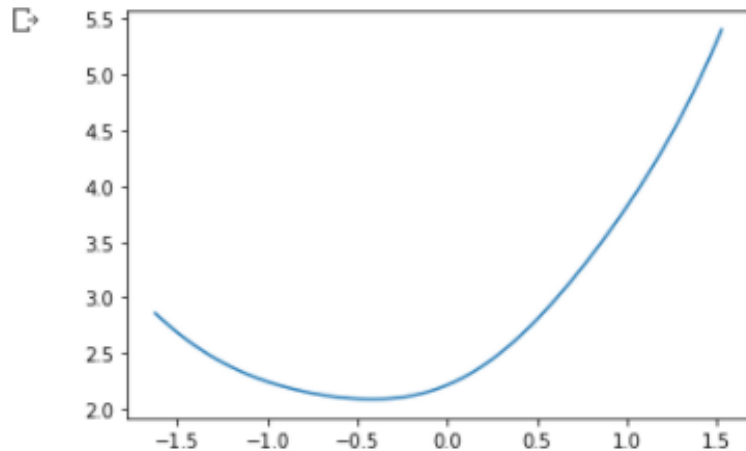
A에 영향을 미치는 변수인 X1, X2가 중요하게 쓰임.

## 연구 결과 - Synthetic Data

```
[127] CF.ATE_pred()
```

```
('(ATE, E_y_1, E_y_0)',  
 3.049198527805677,  
 3.0578071483974094,  
 0.008608620591732598)
```

```
[128] CF.CATE_pred('X1')
```



실제 true ATE와 CATE를 잘 추정함.

$$ATE = E(Y^1) - E(Y^0) = 3 - 0$$

$$CATE(x_1) = \left(x_1 + \frac{1}{2}\right)^2 + \frac{7}{4}$$

## 연구 결과 - Synthetic Data

```
[129] Y[A==1].mean() # 상관을 이용한 추정.
```

```
3.307102393399478
```

```
[130] Y[A==0].mean()
```

```
-0.562034871374579
```

상관을 이용하여 추정하는 경우  
 $ATE=3-0$ 을  $3.31-(-0.56)=3.89$ 로 추정.  
잘 추정하지 못함.



# Thank You

최성식

[btd63@snu.ac.kr](mailto:btd63@snu.ac.kr)