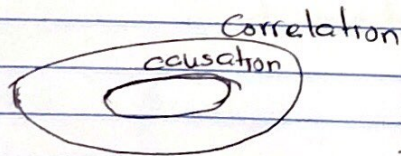
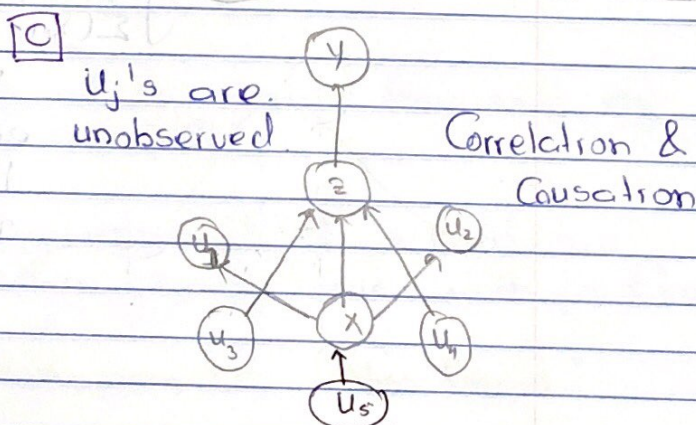
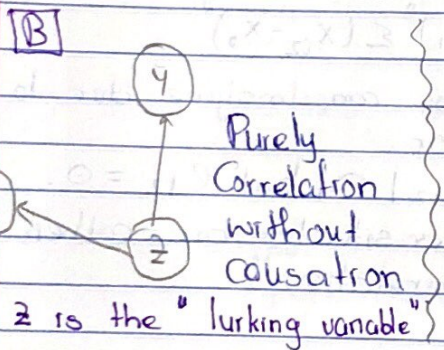
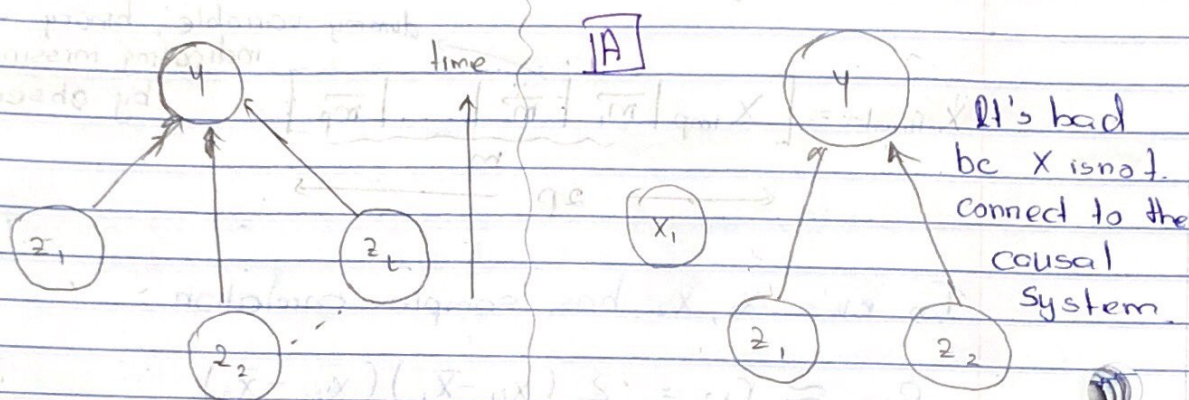


$$y = t(\bar{z}) = f(\bar{x}) + \epsilon$$

\bar{z} : Causal variable of y

\bar{x} : Variable that capture information about \bar{z} .

Causal Diagram



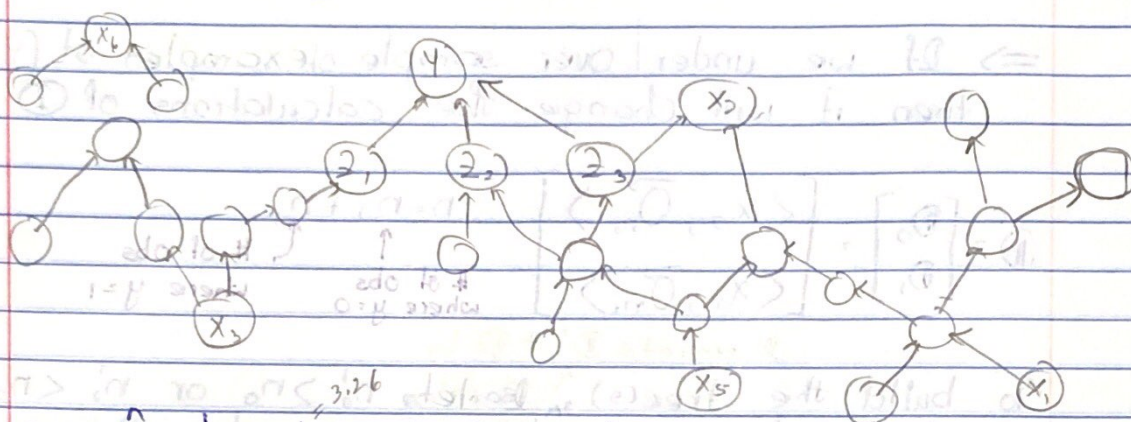
causations \subset correlation.
 \Rightarrow Causation \Rightarrow Correlation, but.
 Correlation \nRightarrow Causation

- z : rainfall
- y : # of car accidents
- x : # of umbrella sold
- u : discount on umbrella

Q1 If x has an indirect connection to y , it is causal
 If x is manipulated, y changes (definition of causality)

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

e.g. phenomenon in the real world may have this causal diagram.



$$\Rightarrow \hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_6 x_6$$

Interpretation of b_1 :-

When comparing two "mutually observed" observations (A) and (B). Sample in the same function as observations in \textcircled{D} , when (A) has an x_1 value one unit larger than (B)'s x_1 measurement, but both (A) & (B) share the same measurements x_2, x_3, \dots, x_p , then (A) is predicted to have a response y that differs by b_1 units or average from the response y of (B) assuming the linear model is true and independent of (A), (B).

COURSE OVER

Asymmetric Cost Classification in trees / bagged trees / RF

Recalled

① Splits in tree construction use weighted Gini.

② leaf assignment uses mode [y's in leaf]

⇒ If we under / over sample observation of ($y=0$) or ($y=1$) then it will change the calculations of ① and ②

$$\mathbb{D} = \begin{bmatrix} \mathbb{D}_0 \\ \mathbb{D}_1 \end{bmatrix} = \begin{bmatrix} \langle x_0, \bar{O}_0 \rangle \\ \langle x_1, \bar{O}_1 \rangle \end{bmatrix}$$

$n = n_0 + n_1$

$\uparrow \quad \uparrow$
of obs where $y=0$ # of obs where $y=1$

we generate $\mathbb{D}' \neq \mathbb{D}$ by

To build the tree(s), \uparrow sampling $n'_0 > n_0$ or $n'_0 < n_0$ and / or $n'_1 > n_1$ or $n'_1 < n_1$ where n'_i is the # sample with replacement from \mathbb{D}_i and n_i is the # of \mathbb{D}_i .

CONFIDENTIAL