COS 534 Problem set 2

#1. Fairness of "optimal" predictions

Counterexample for $R \perp A$ and $R \perp A \mid Y$

We assume $0 < p < 1$.

$P(Y=1 \mid X=0, A=0) = p = r(X=0, A=0)$ $\qquad$ $P(X=0, A=0) = \frac{1}{2}$

$P(Y=1 \mid X=1, A=0) = 1 = r(X=1, A=0)$ $\qquad$ $P(X=1, A=0) = 0$

$P(Y=1 \mid X=0, A=1) = 0 = r(X=0, A=1)$ $\qquad$ $P(X=0, A=1) = 0$

$P(Y=1 \mid X=1, A=1) = 1 = r(X=1, A=1)$ $\qquad$ $P(X=1, A=1) = \frac{1}{2}$

① $R \perp A \rightarrow P(R=1 \mid A=0) = P(R=1 \mid A=1)$

However, we will show $P(R=1 \mid A=0) \neq P(R=1 \mid A=1)$ in the above example.

$$P(R=1 \mid A=0) = \frac{P(R=1, A=0)}{P(A=0)} = \frac{P(R=1, X=0, A=0) + P(R=1, X=1, A=0)}{P(X=0, A=0) + P(X=1, A=0)}$$

$$= \frac{P(X=0, A=0)\, P(R=1 \mid X=0, A=0) + P(X=1, A=0)\, P(R=1 \mid X=1, A=0)}{\frac{1}{2} + 0}$$

$$= \frac{\frac{1}{2} \cdot 0 + 0 \cdot 1}{\frac{1}{2}} = 0$$

Similarly, $P(R=1 \mid A=1) = \dfrac{P(R=1, X=0, A=1) + P(R=1, X=1, A=1)}{P(X=0, A=1) + P(X=1, A=1)} = \dfrac{0 \cdot 0 + \frac{1}{2} \cdot 1}{0 + \frac{1}{2}} = 1$

As $P(R=1 \mid A=0) = 0 \neq P(R=1 \mid A=1) = 1$, $R \perp A$ does not hold.

② $R \perp A \mid Y \rightarrow P(R=1 \mid Y=1, A=0) = P(R=1 \mid Y=1, A=1)$

However, we will show $P(R=1 \mid Y=1, A=0) \neq P(R=1 \mid Y=1, A=1)$ in the above example.

$$P(R=1 \mid Y=1, A=0) = \frac{P(R=1, Y=1, A=0)}{P(Y=1, A=0)} = \frac{P(R=1, Y=1, A=0, X=0) + P(R=1, Y=1, A=0, X=1)}{P(Y=1, A=0, X=0) + P(Y=1, A=0, X=1)}$$

$$= \frac{\frac{1}{2} \cdot 0 + 0 \cdot 1}{p \cdot \frac{1}{2} + 1 \cdot 0} = 0$$

$$P(R=1 \mid Y=1, A=1) = \frac{P(R=1, Y=1, A=1)}{P(Y=1, A=1)} = \frac{P(R=1, Y=1, A=1, X=0) + P(R=1, Y=1, A=1, X=1)}{P(Y=1, A=1, X=0) + P(Y=1, A=1, X=1)}$$

$$= \frac{0 \cdot 0 + 1 \cdot \frac{1}{2}}{0 \cdot 0 + 1 \cdot \frac{1}{2}} = 1$$

As $P(R=1 \mid Y=1, A=0) \neq P(R=1 \mid Y=1, A=1)$, $R \perp A \mid Y$ does not hold.

③ We will show $P(Y=1 \mid R=r, A=a) = r$.

$$P(Y=1 \mid R=r, A=a) = \frac{P(Y=1, R=r, A=a)}{P(R=r, A=a)}$$

$$P(R=r, A=a) = \sum_{x \in X} P(R=r, A=a, X=x) = \sum_{x : r(x,a)=r} P(A=a, X=x)$$

$$P(Y=1, R=r, A=a) = \sum_{x \in X} P(Y=1, R=r, A=a, X=x) = \sum_{x : r(x,a)=r} P(Y=1, A=a, X=x)$$

$$= \sum_{x : r(x,a)=r} \underbrace{P(Y=1 \mid A=a, X=x)}_{= r \text{ because } r(x,a)=r} P(A=a, X=x) = r \cdot \sum_{x : r(x,a)=r} P(A=a, X=x)$$

$$P(Y=1 \mid R=r, A=a) = \frac{P(Y=1, R=r, A=a)}{P(R=r, A=a)} = \frac{r \cdot \sum_{x : r(x,a)=r} P(A=a, X=x)}{\sum_{x : r(x,a)=r} P(A=a, X=x)} = r. \blacksquare$$