

Mixtape Sessions: Machine Learning and Heterogeneous Effects

May 15, 2023

A little about me

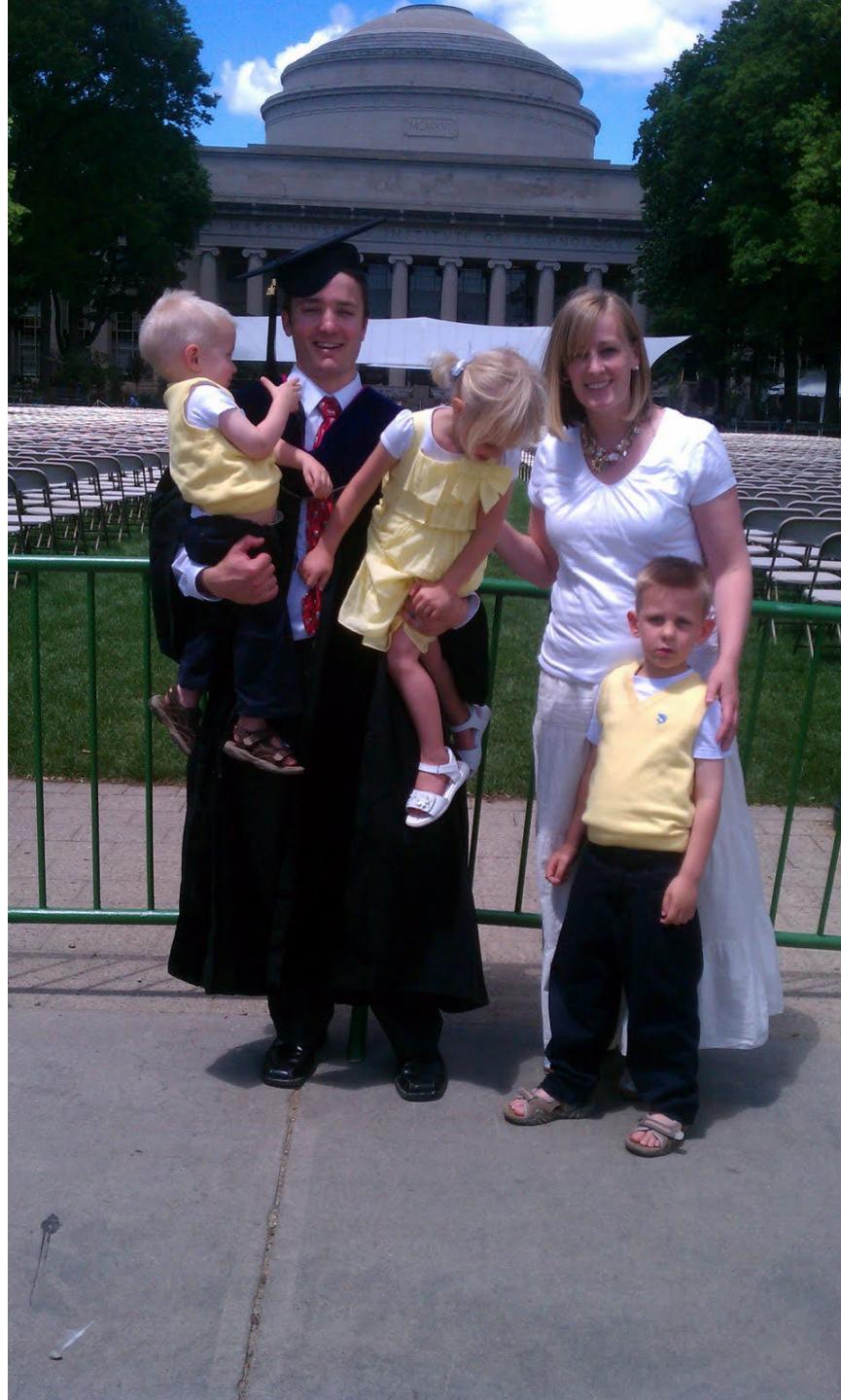














SPLASH MOUNTAIN

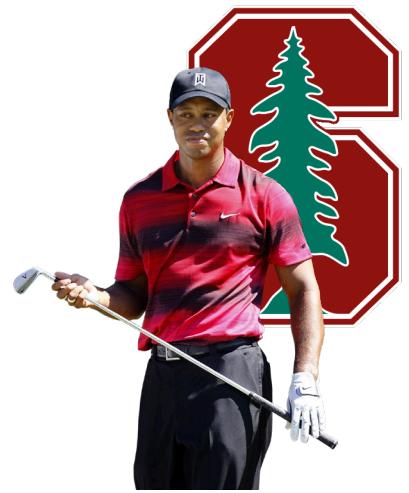
Disneyland

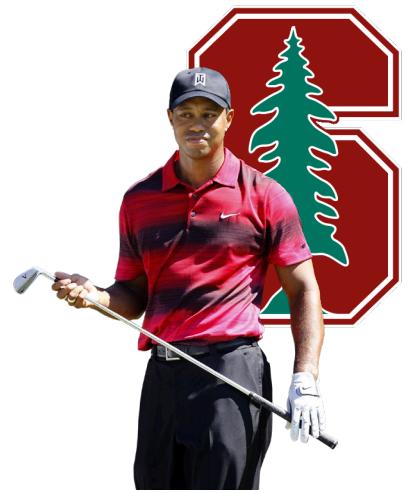






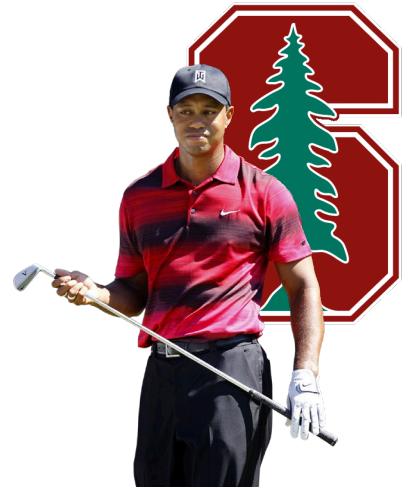






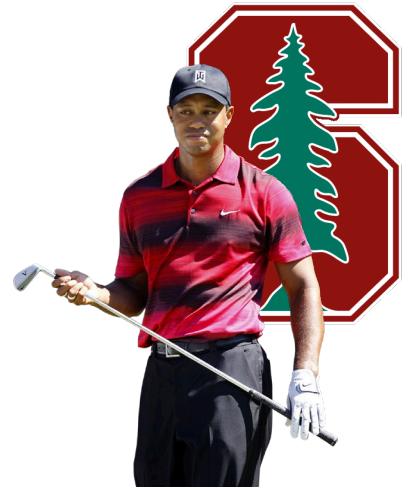
\$\$\$\$

\$43M



\$\$\$\$

\$43M



\$\$\$\$

\$43M



\$

\$700K



\$\$\$\$

\$43M

D = 1

Y(1)

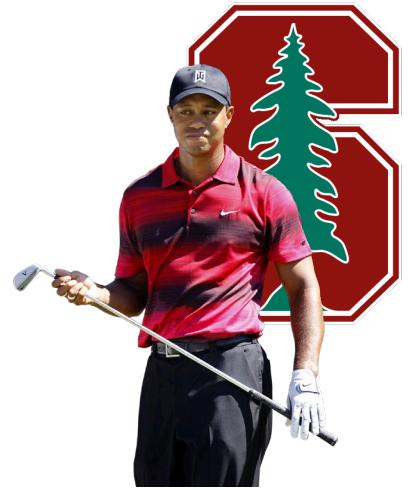
\$

\$700K

D = 0

Y(0)

Potential outcomes



\$\$\$\$

\$43M

-

\$

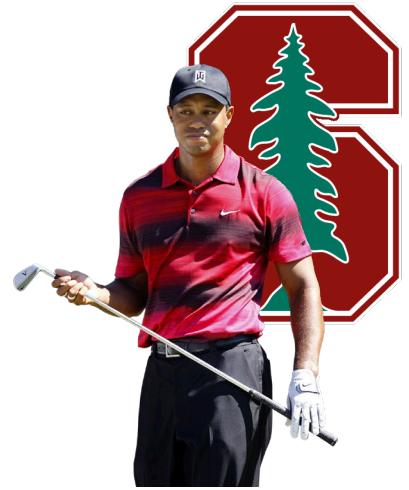
\$700K = \$42.3M

Y(1)

-

Y(0)

= Treatment effect



\$\$\$\$

\$43M

\$

\$700K = \$42.3M

$Y(1)$

-

$Y(0)$

= Treatment effect

counterfactual



\$\$\$\$

\$43M

-

\$

\$700K

=

\$42.3M

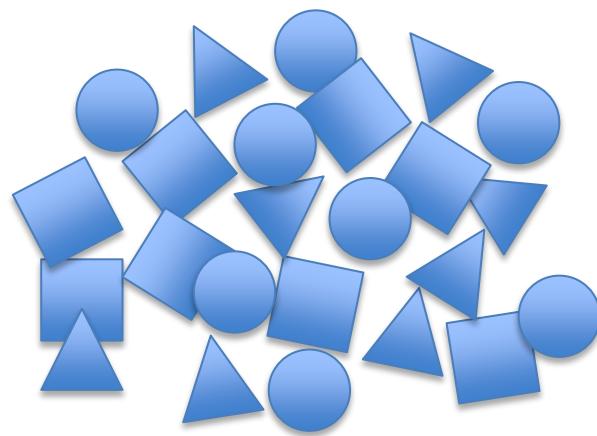
Y(1)

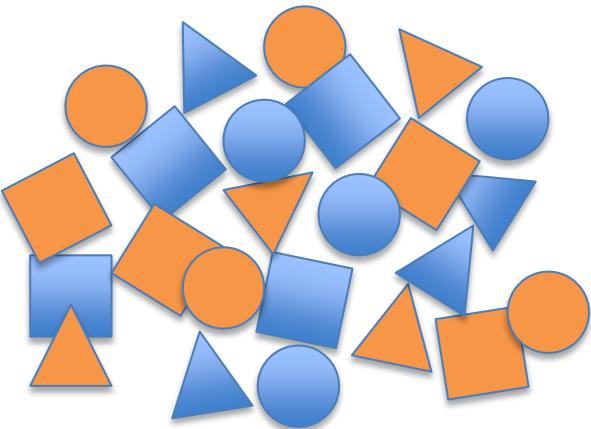
-

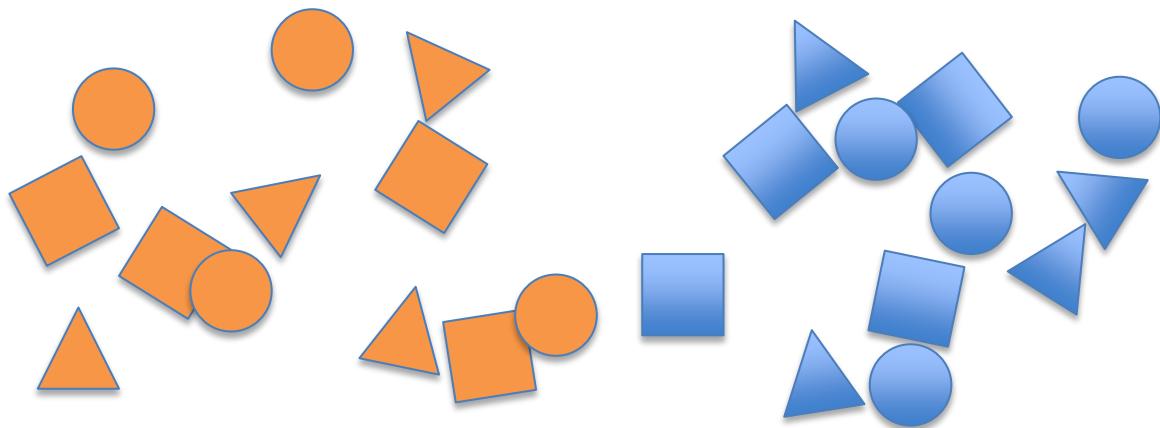
Y(0)

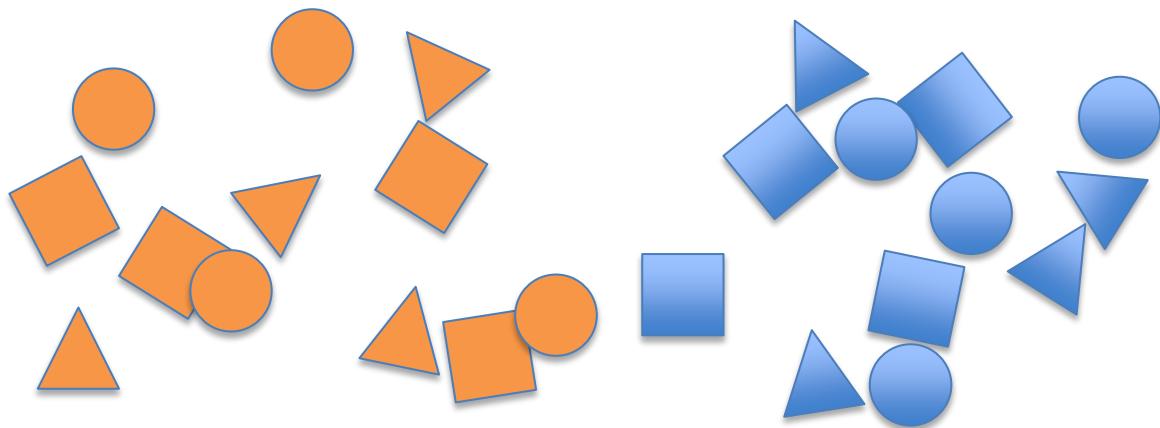
=

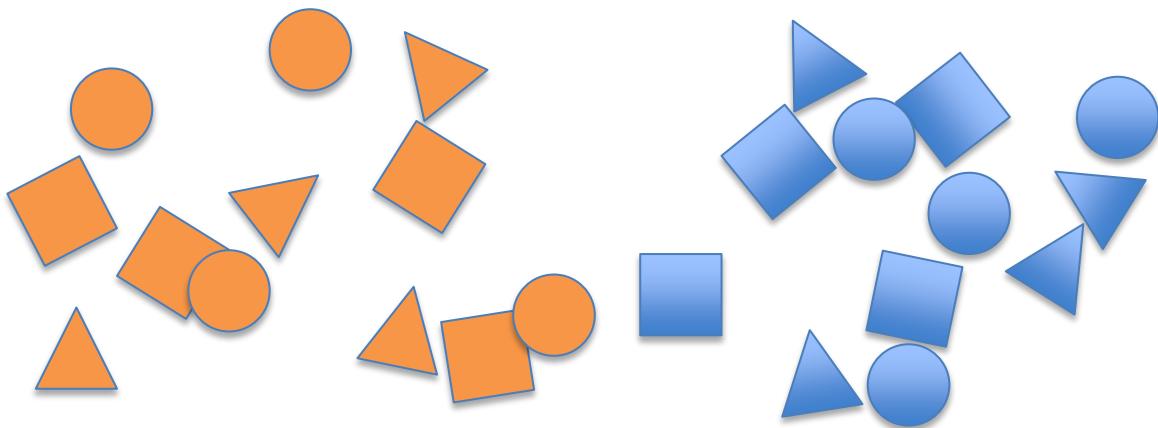
Treatment effect







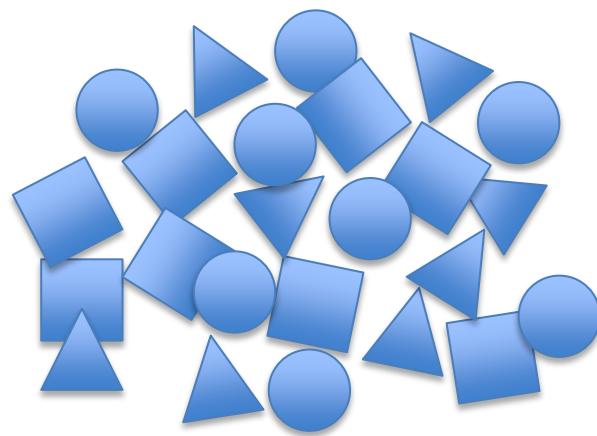

$$E[Y(1)]$$
$$E[Y(0)]$$

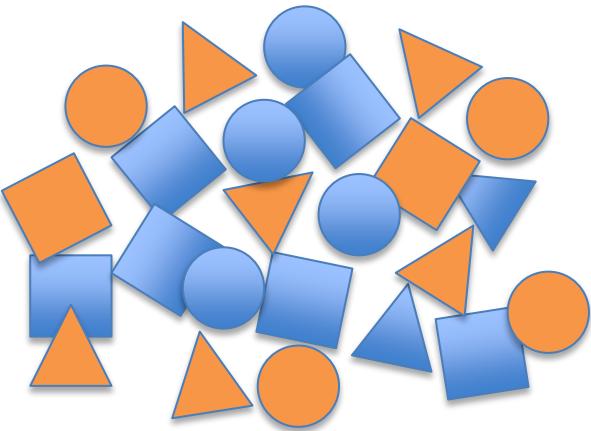


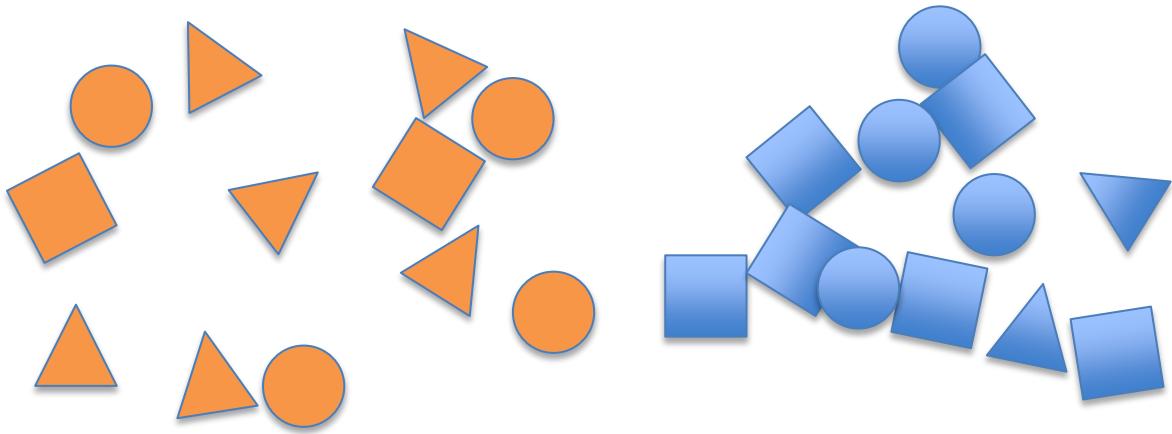
$$E[Y(1)] -$$

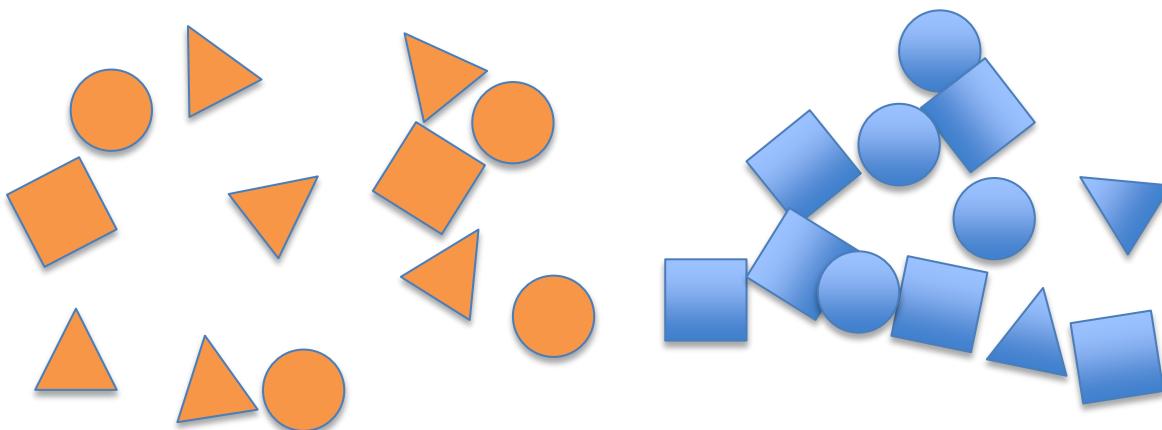
$$E[Y(0)] = E[Y(1) - Y(0)]$$

ATE







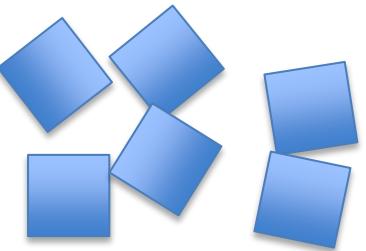
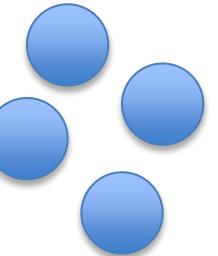
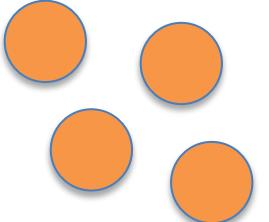
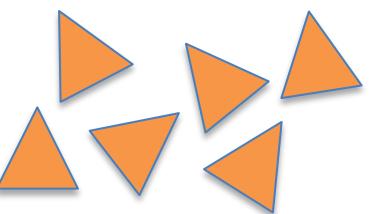


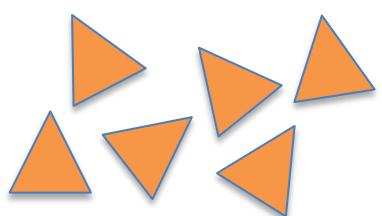
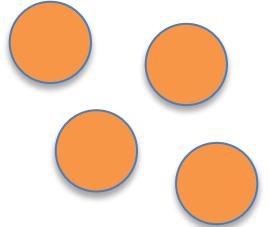
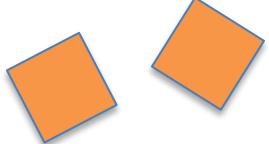
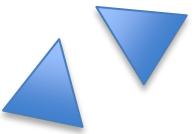
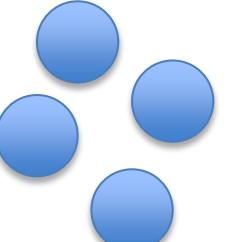
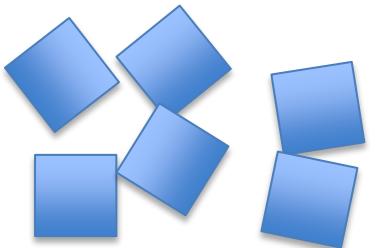
ATT

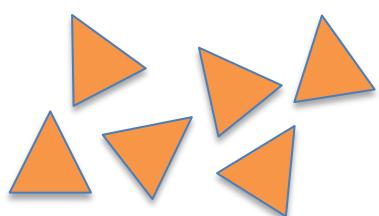
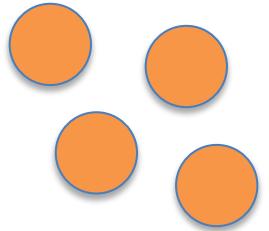
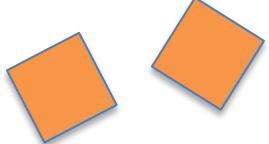
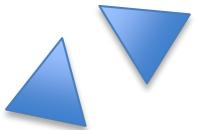
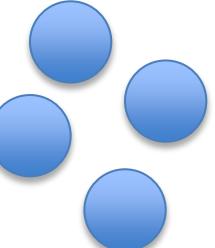
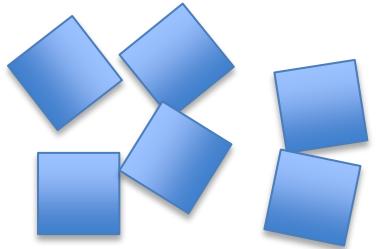
$$E[Y | D=1] - E[Y | D=0] = E[Y(1) - Y(0) | D=1]$$

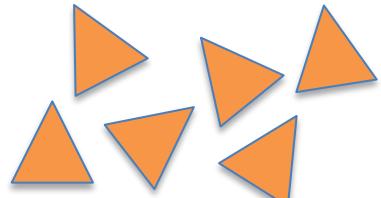
$$+ E[Y(0) | D=1] - E[Y(0) | D=0]$$

selection bias

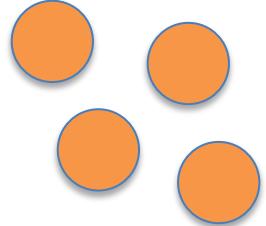


 $E[Y(1) | \triangle]$  $E[Y(1) | \circ]$  $E[Y(1) | \square]$  $E[Y(0) | \triangle]$  $E[Y(0) | \circ]$  $E[Y(0) | \square]$

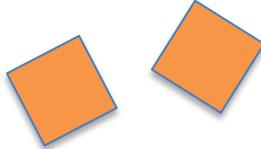
 $E[Y(1) | \blacktriangle]$  $E[Y(1) | \bullet]$  $E[Y(1) | \blacksquare]$  $E[Y(0) | \blacktriangle]$  $E[Y(0) | \bullet]$  $E[Y(0) | \blacksquare]$



$$E[Y(1) | \blacktriangle]$$



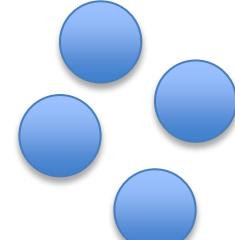
$$E[Y(1) | \bullet]$$



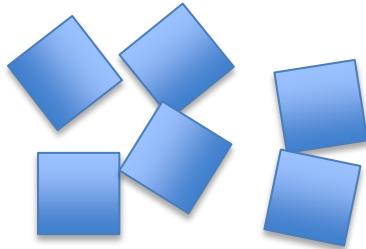
$$E[Y(1) | \blacksquare]$$



$$E[Y(0) | \blacktriangle]$$



$$E[Y(0) | \bullet]$$



$$E[Y(0) | \blacksquare]$$

As long as:

$$E[Y(1) | \triangle] = E[Y(1) | \blacktriangle]$$

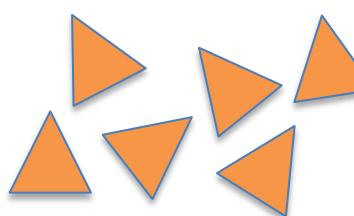
$$E[Y(0) | \triangle] = E[Y(0) | \blacktriangle]$$

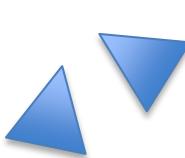
$$E[Y(1) | \circ] = E[Y(1) | \bullet]$$

$$E[Y(0) | \circ] = E[Y(0) | \bullet]$$

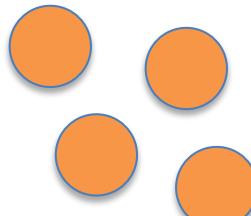
$$E[Y(1) | \square] = E[Y(1) | \blacksquare]$$

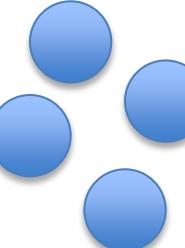
$$E[Y(0) | \square] = E[Y(0) | \blacksquare]$$


$$E[Y(1) | \blacktriangle]$$

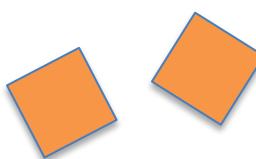


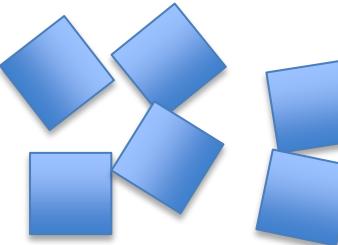
$$E[Y(0) | \blacktriangle] = E[Y(1) - Y(0) | \blacktriangle]$$


$$E[Y(1) | \bullet]$$

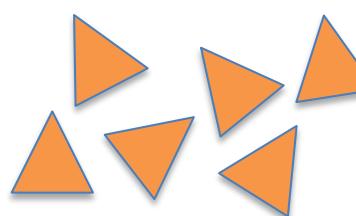


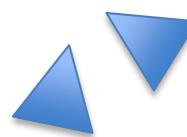
$$E[Y(0) | \bullet] = E[Y(1) - Y(0) | \bullet]$$


$$E[Y(1) | \blacksquare]$$

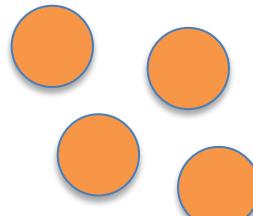


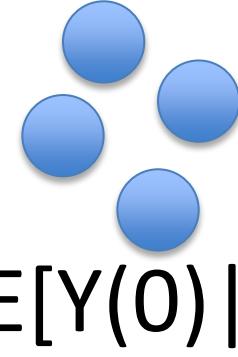
$$E[Y(0) | \blacksquare] = E[Y(1) - Y(0) | \blacksquare]$$


$$E[Y(1) | \blacktriangle] -$$

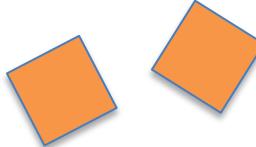


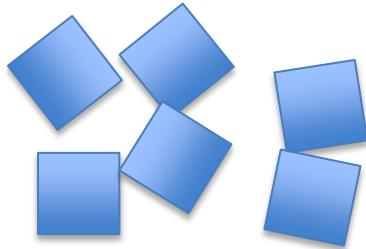
$$E[Y(0) | \blacktriangle] = E[Y(1) - Y(0) | \blacktriangle]$$


$$E[Y(1) | \bullet] -$$

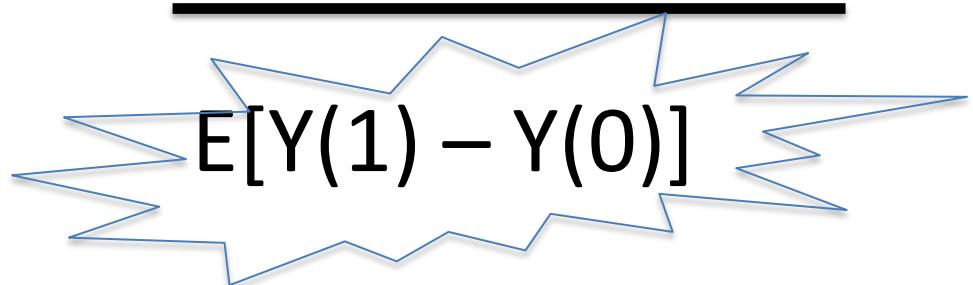


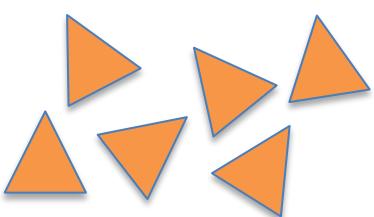
$$E[Y(0) | \bullet] = E[Y(1) - Y(0) | \bullet]$$


$$E[Y(1) | \blacksquare] -$$

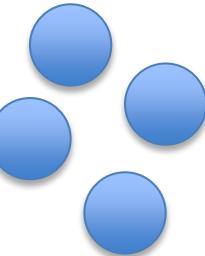
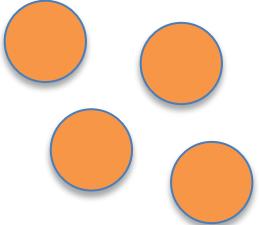


$$E[Y(0) | \blacksquare] = E[Y(1) - Y(0) | \blacksquare]$$

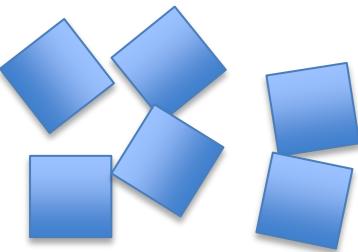
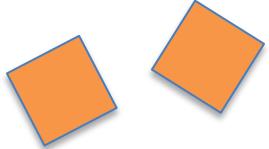

$$E[Y(1) - Y(0)]$$



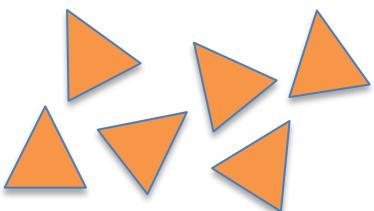
$Y(1), Y(0) \perp\!\!\!\perp D | \blacktriangle$



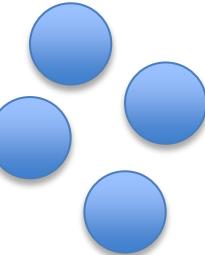
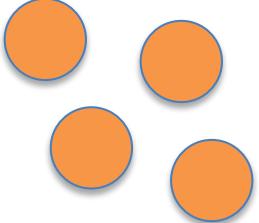
$Y(1), Y(0) \perp\!\!\!\perp D | \bullet$



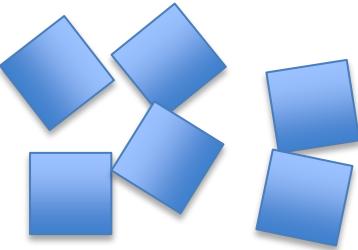
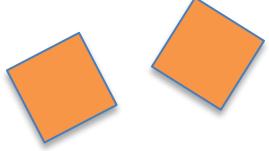
$Y(1), Y(0) \perp\!\!\!\perp D | \blacksquare$



$$Y(1), Y(0) \perp\!\!\!\perp D | \blacktriangle$$



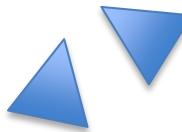
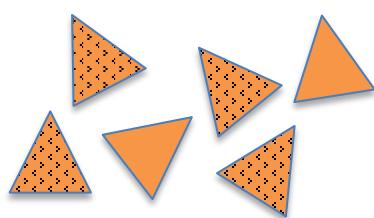
$$Y(1), Y(0) \perp\!\!\!\perp D | \bullet$$



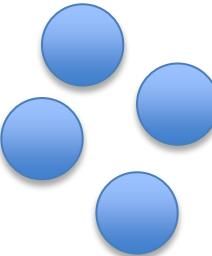
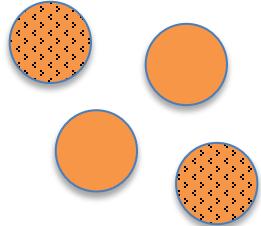
$$Y(1), Y(0) \perp\!\!\!\perp D | \blacksquare$$

Selection on observables,
“unconfoundedness”:

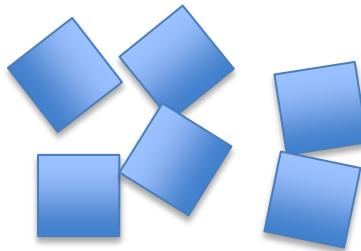
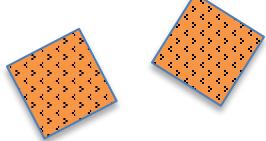
$$Y(1), Y(0) \perp\!\!\!\perp D | X$$



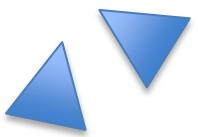
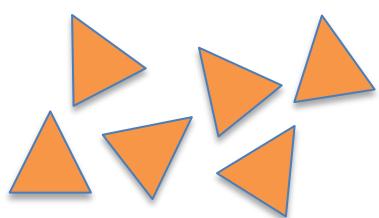
$Y(1), Y(0) \perp\!\!\!\perp D | \blacktriangle$



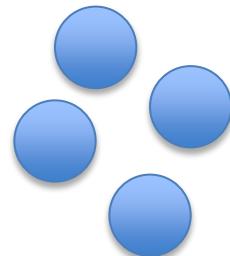
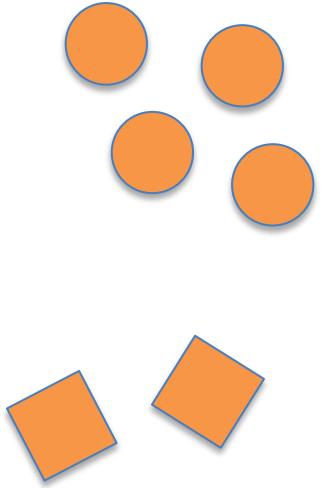
$Y(1), Y(0) \perp\!\!\!\perp D | \bullet$



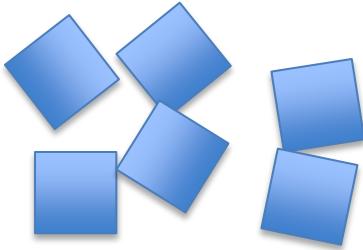
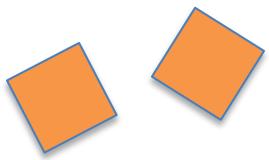
$Y(1), Y(0) \perp\!\!\!\perp D | \blacksquare$



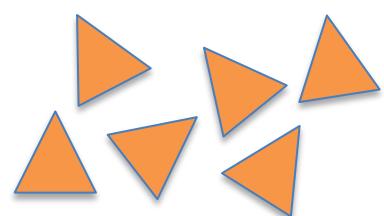
$$\Pr(D=1 | \blacktriangle) = .75$$



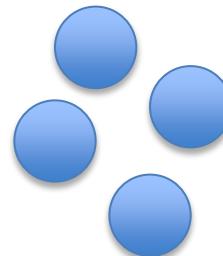
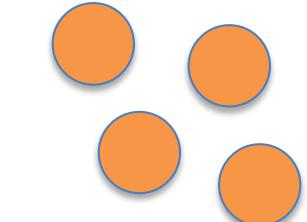
$$\Pr(D=1 | \bullet) = .5$$



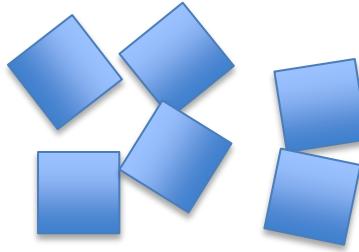
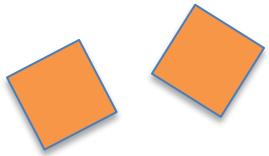
$$\Pr(D=1 | \blacksquare) = .25$$



$$\Pr(D=1 | \blacktriangle) = .75$$



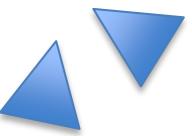
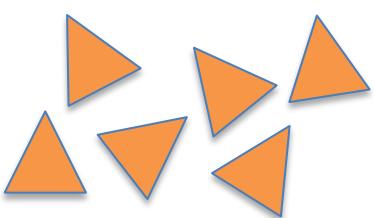
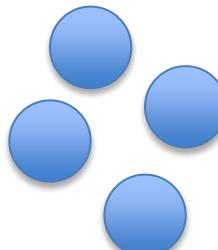
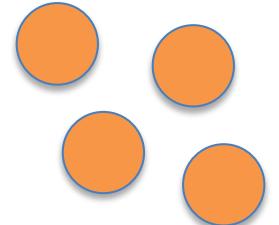
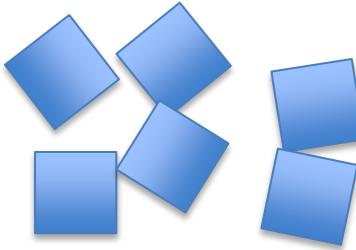
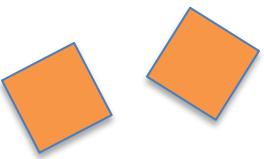
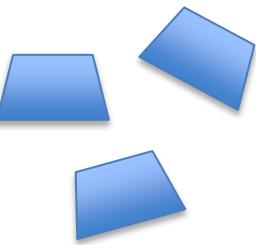
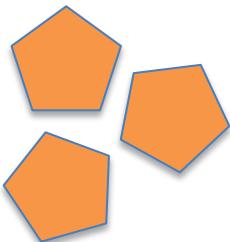
$$\Pr(D=1 | \bullet) = .5$$

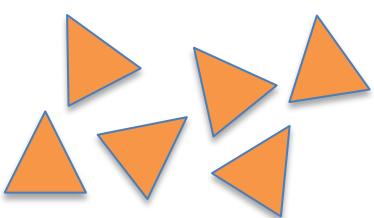


$$\Pr(D=1 | \blacksquare) = .25$$

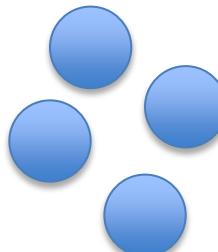
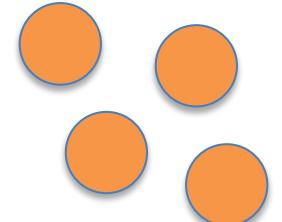
Common support,
“overlap”:

$$0 < \Pr(D=1 | X) < 1$$

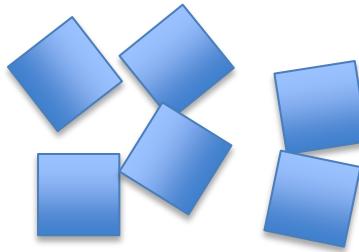
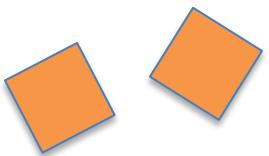

$$\Pr(D=1 | \blacktriangle) = .75$$

$$\Pr(D=1 | \bullet) = .5$$

$$\Pr(D=1 | \blacksquare) = .25$$




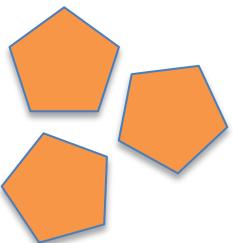
$$\Pr(D=1 | \blacktriangle) = .75$$



$$\Pr(D=1 | \bullet) = .5$$



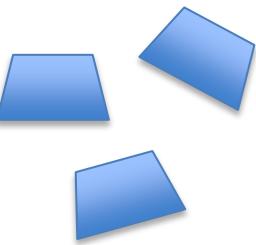
$$\Pr(D=1 | \blacksquare) = .25$$



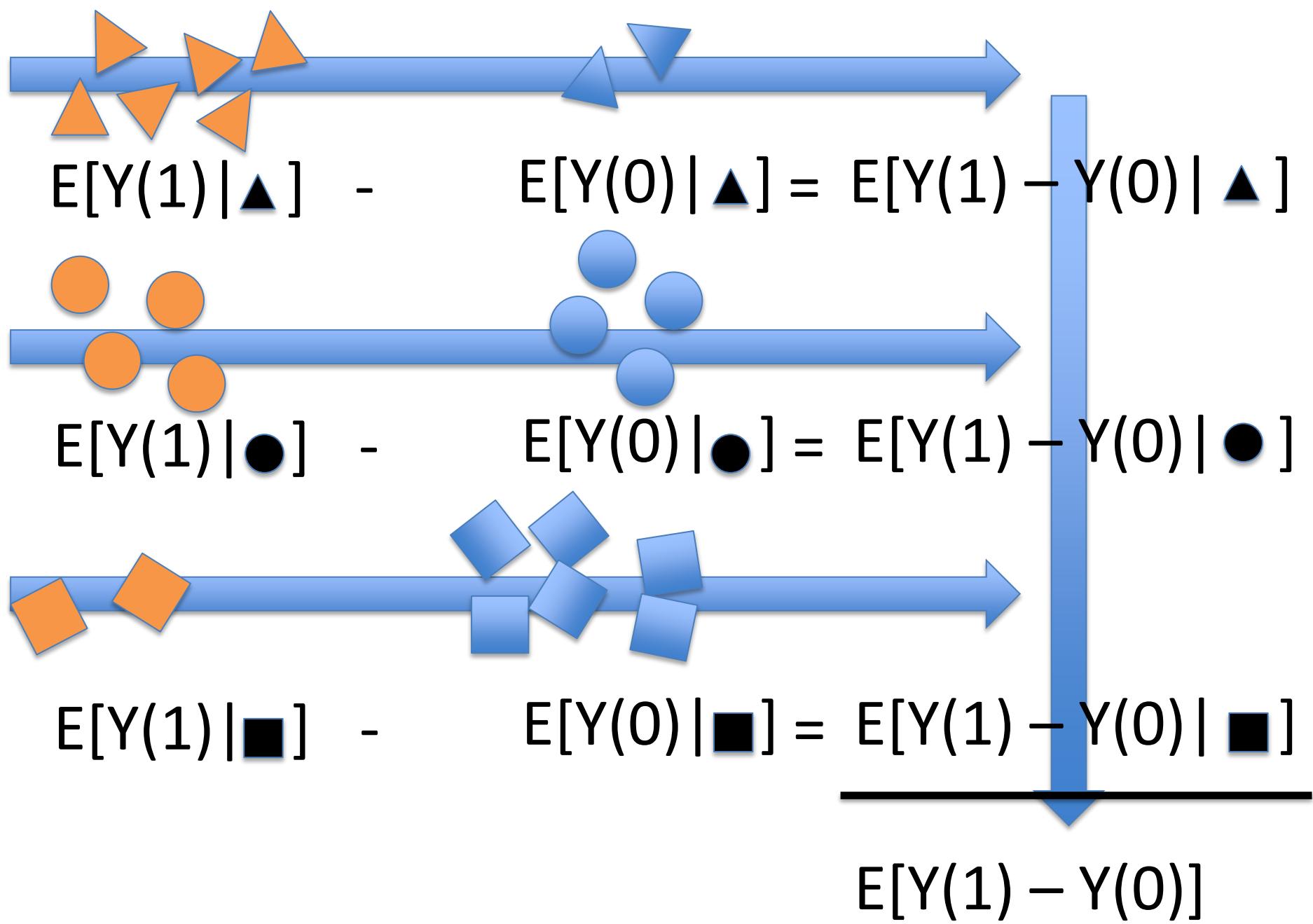
!!

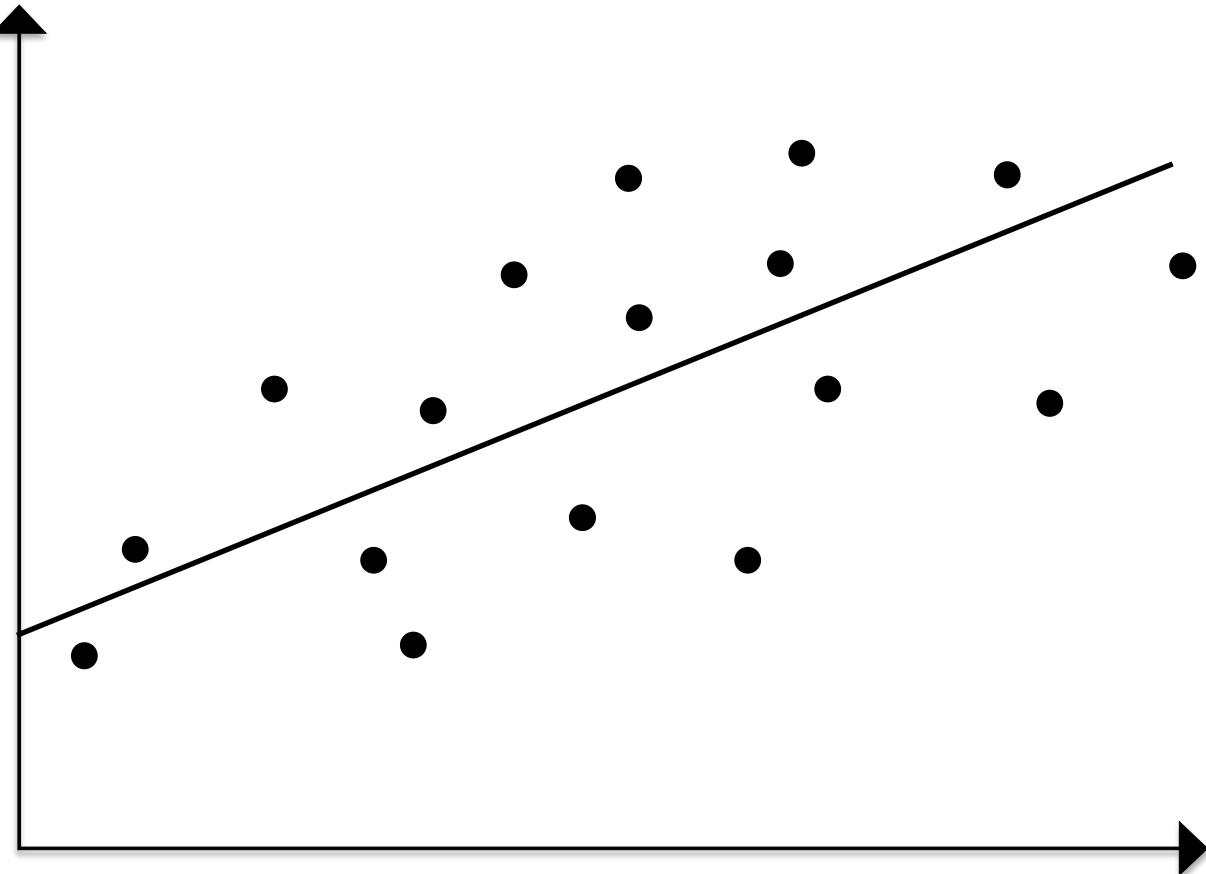
$$\Pr(D=1 | \blacklozenge) = 1$$

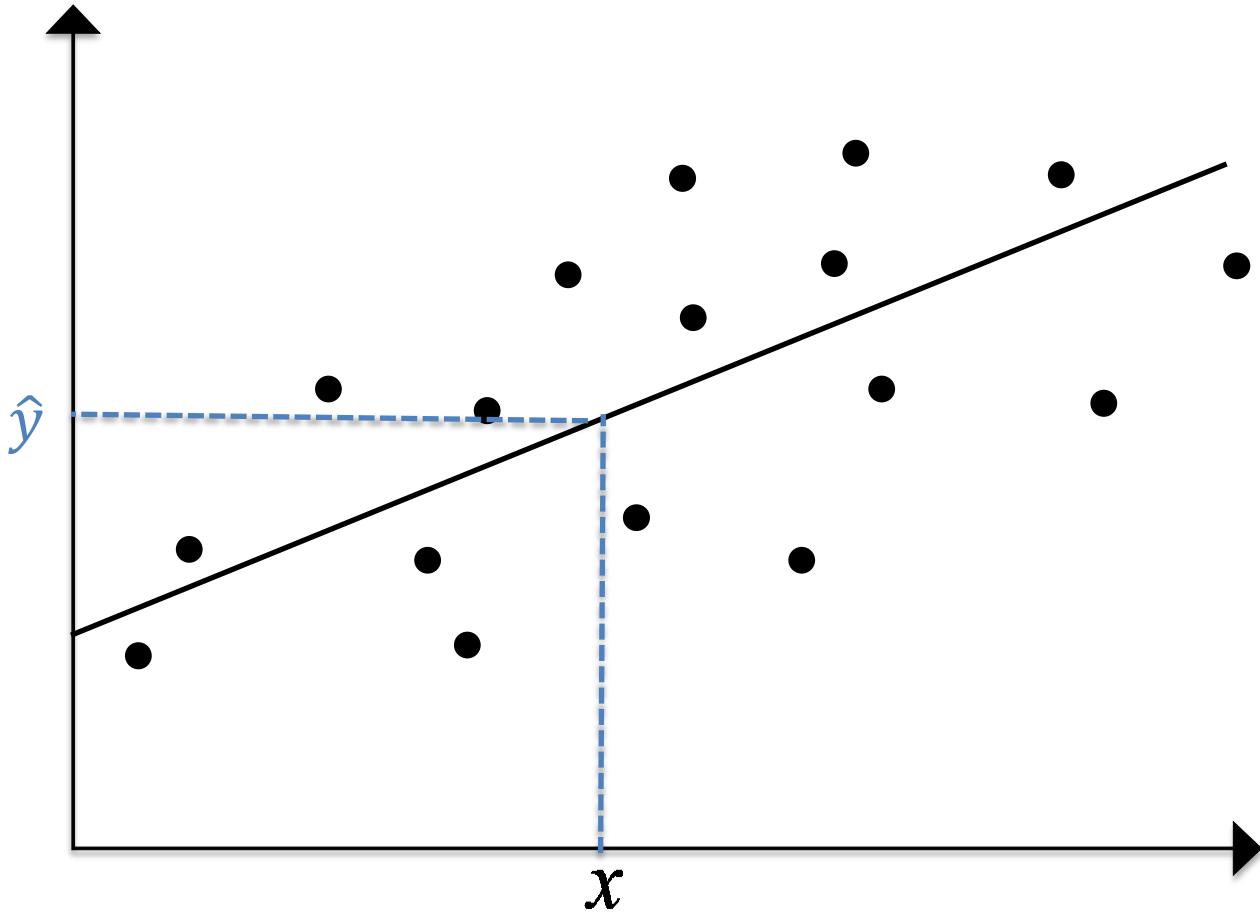
!!

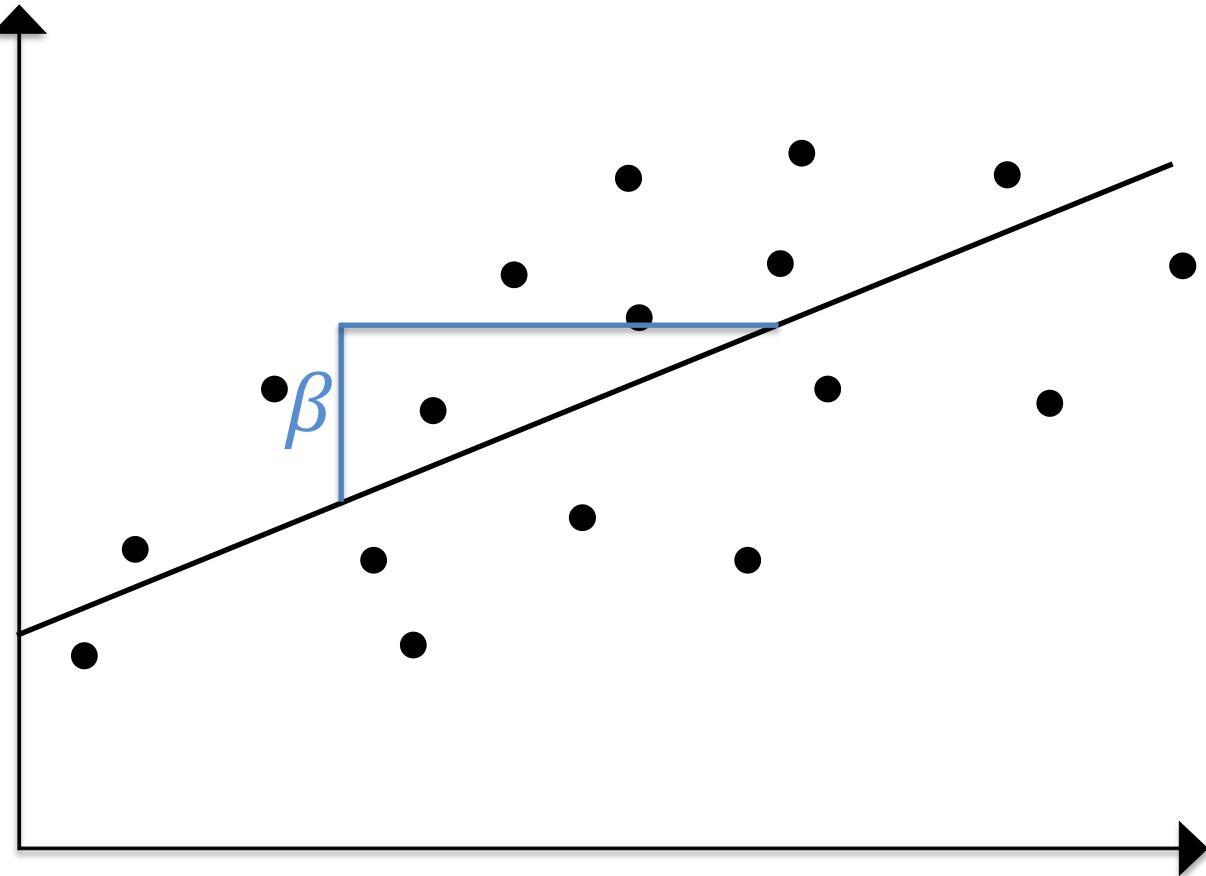


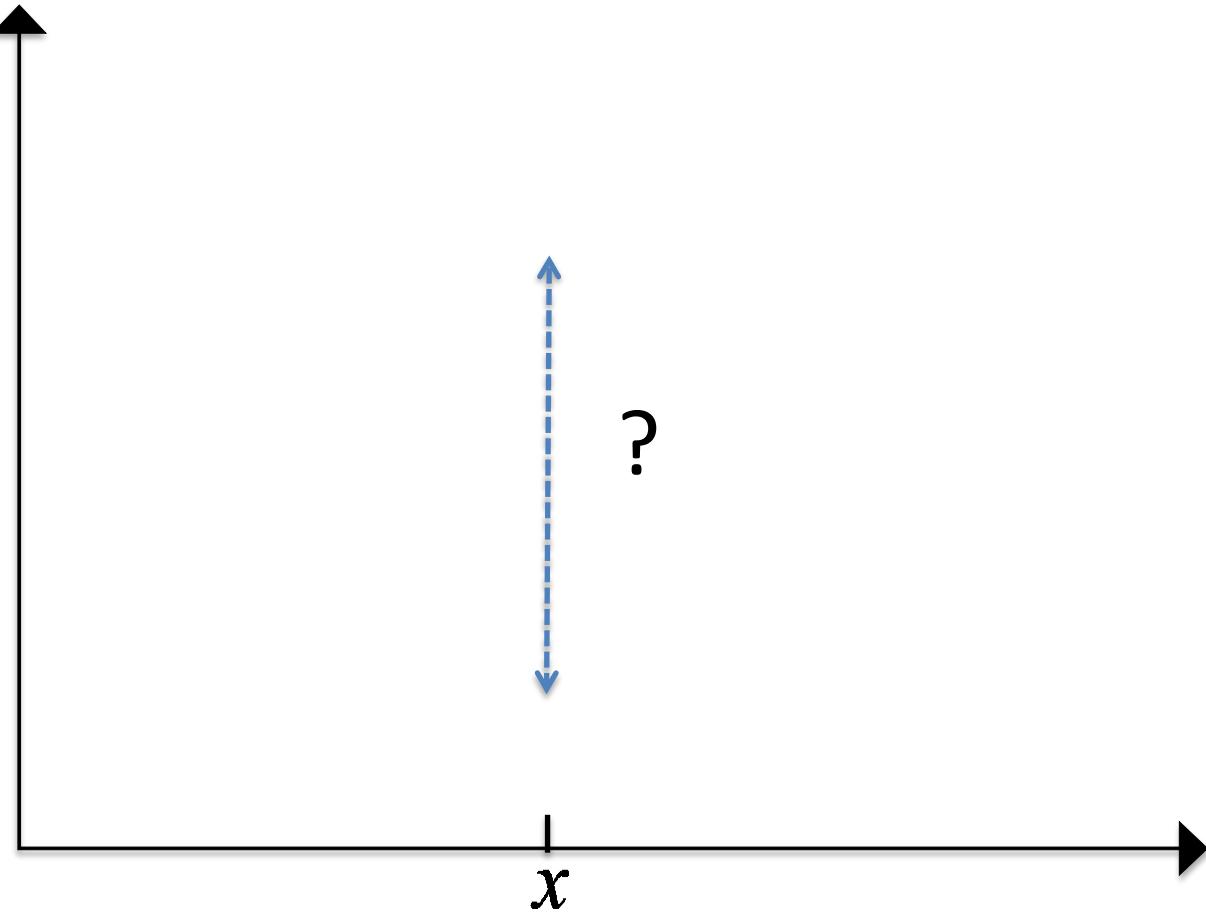
$$\Pr(D=1 | \blacksquare) = 0$$

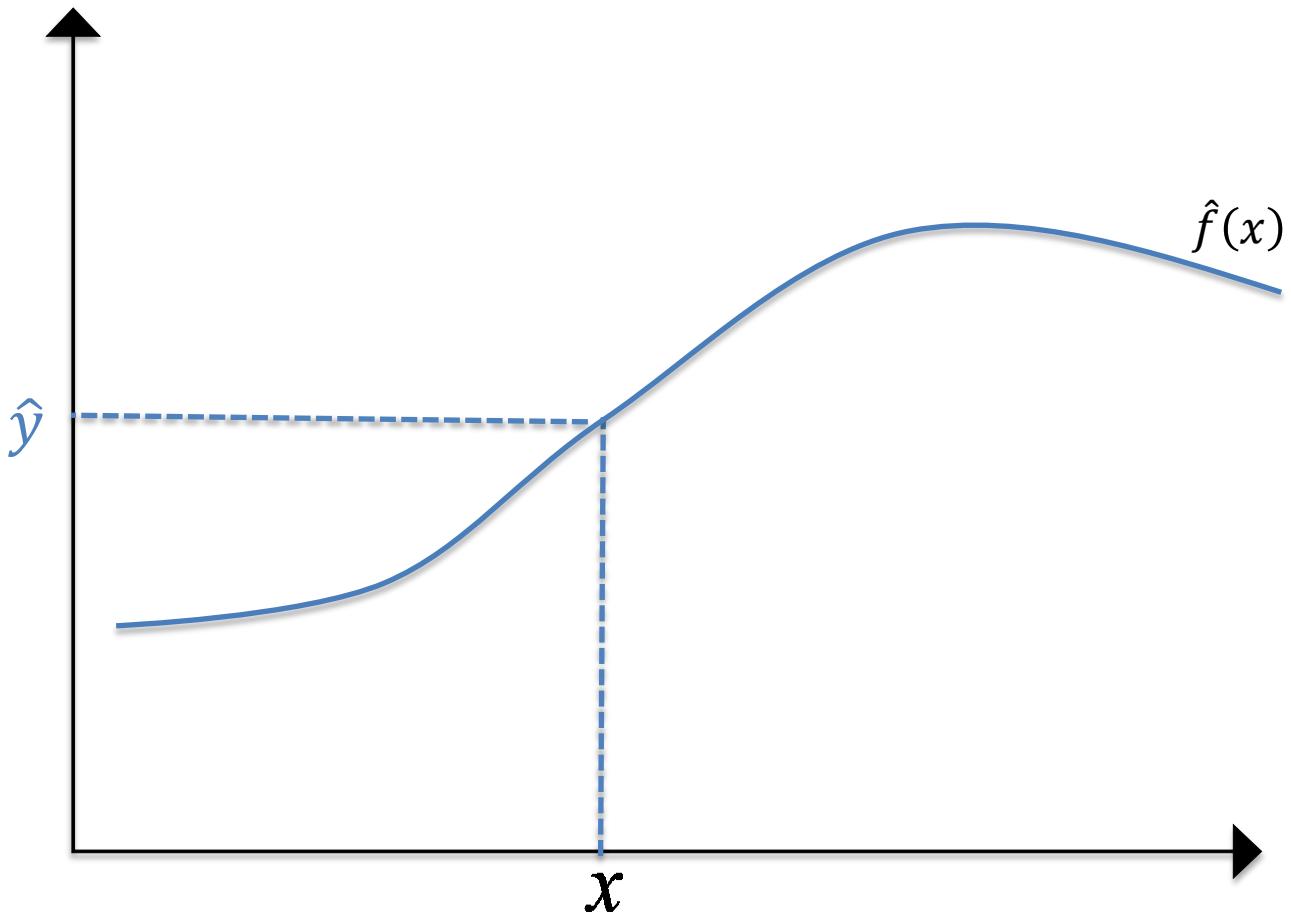


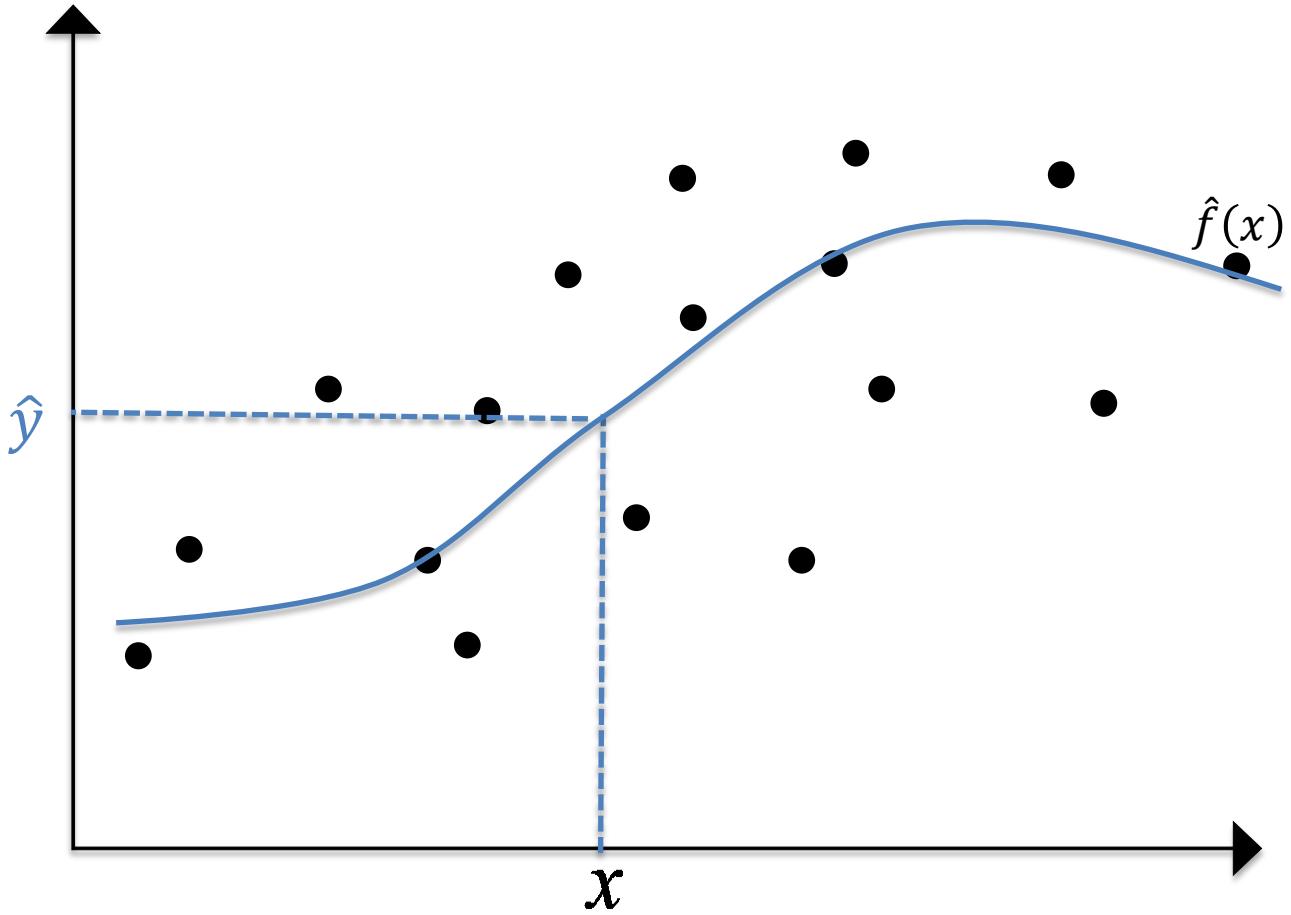


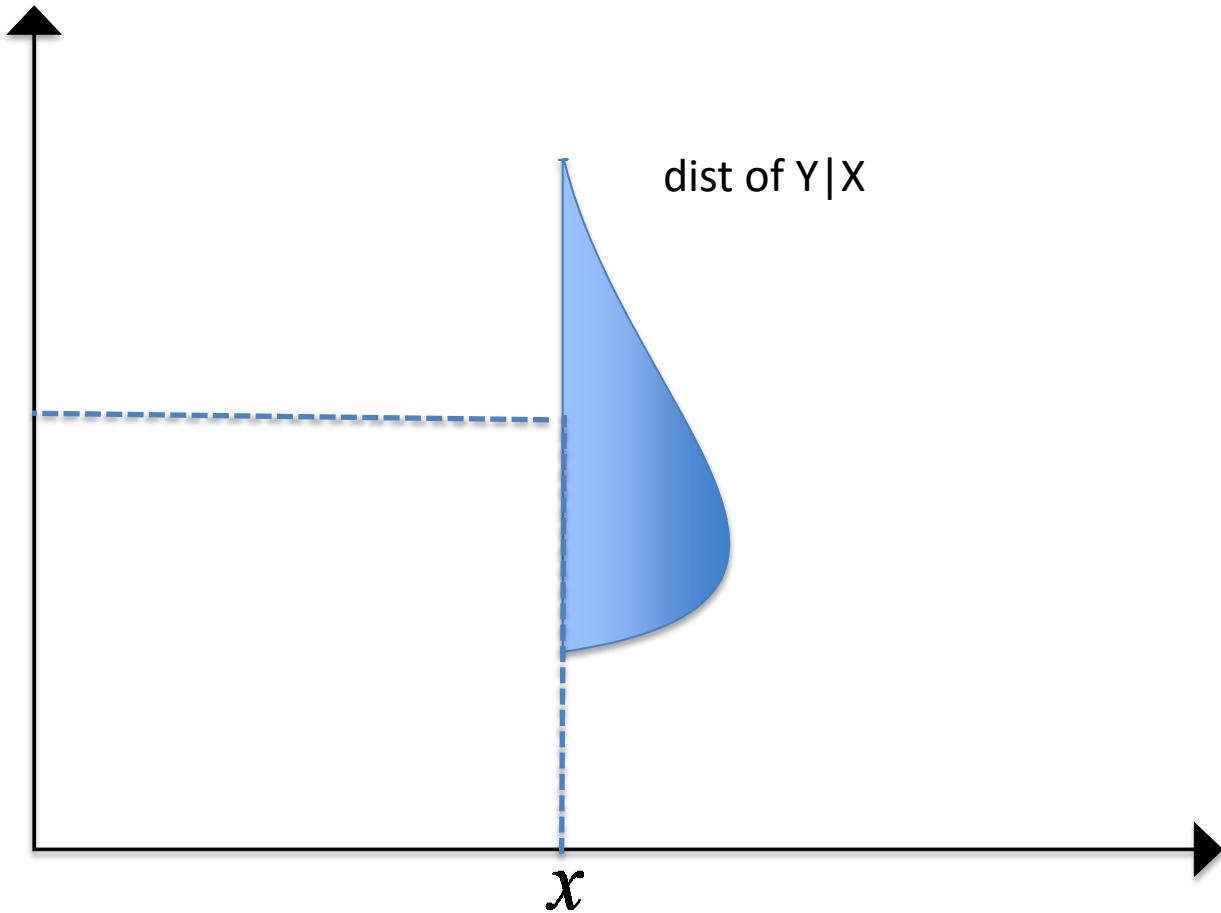


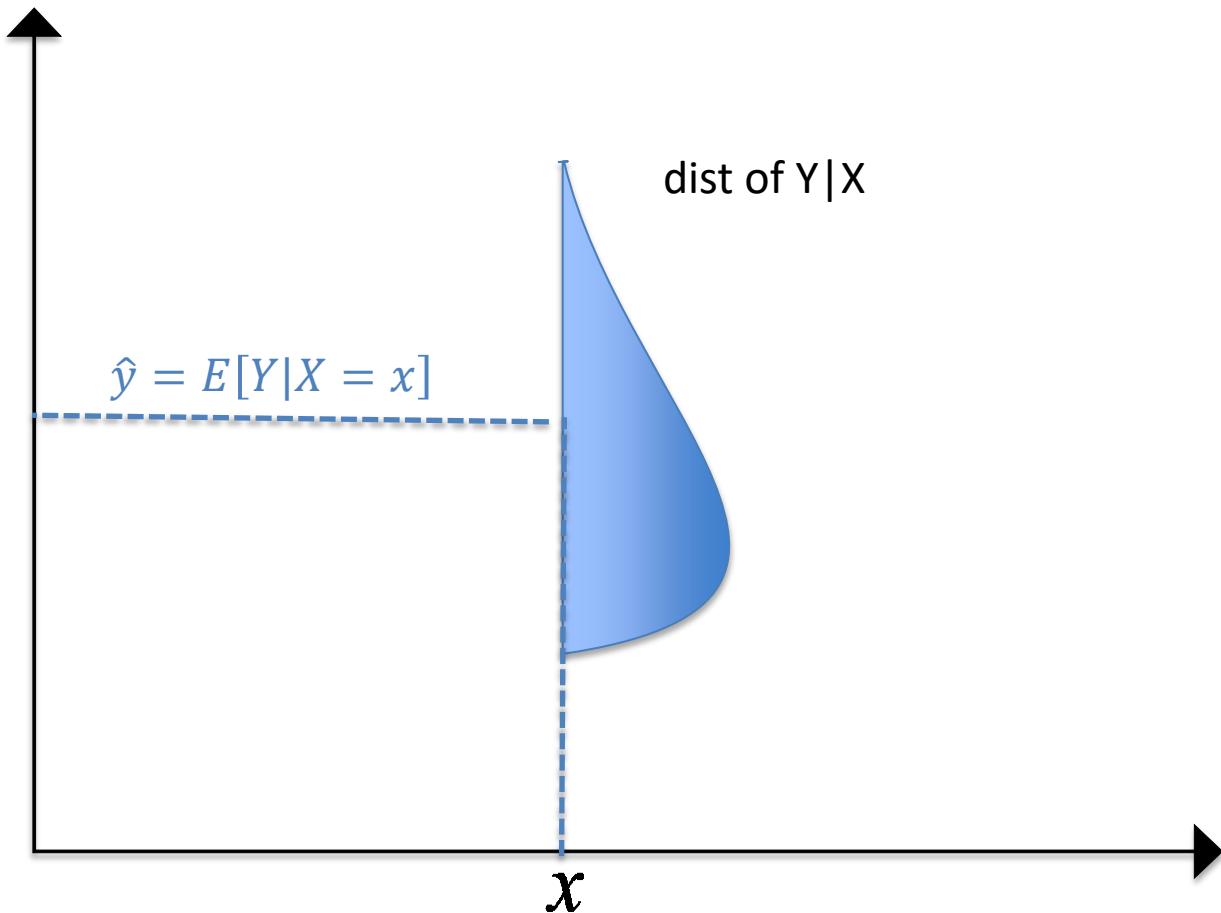


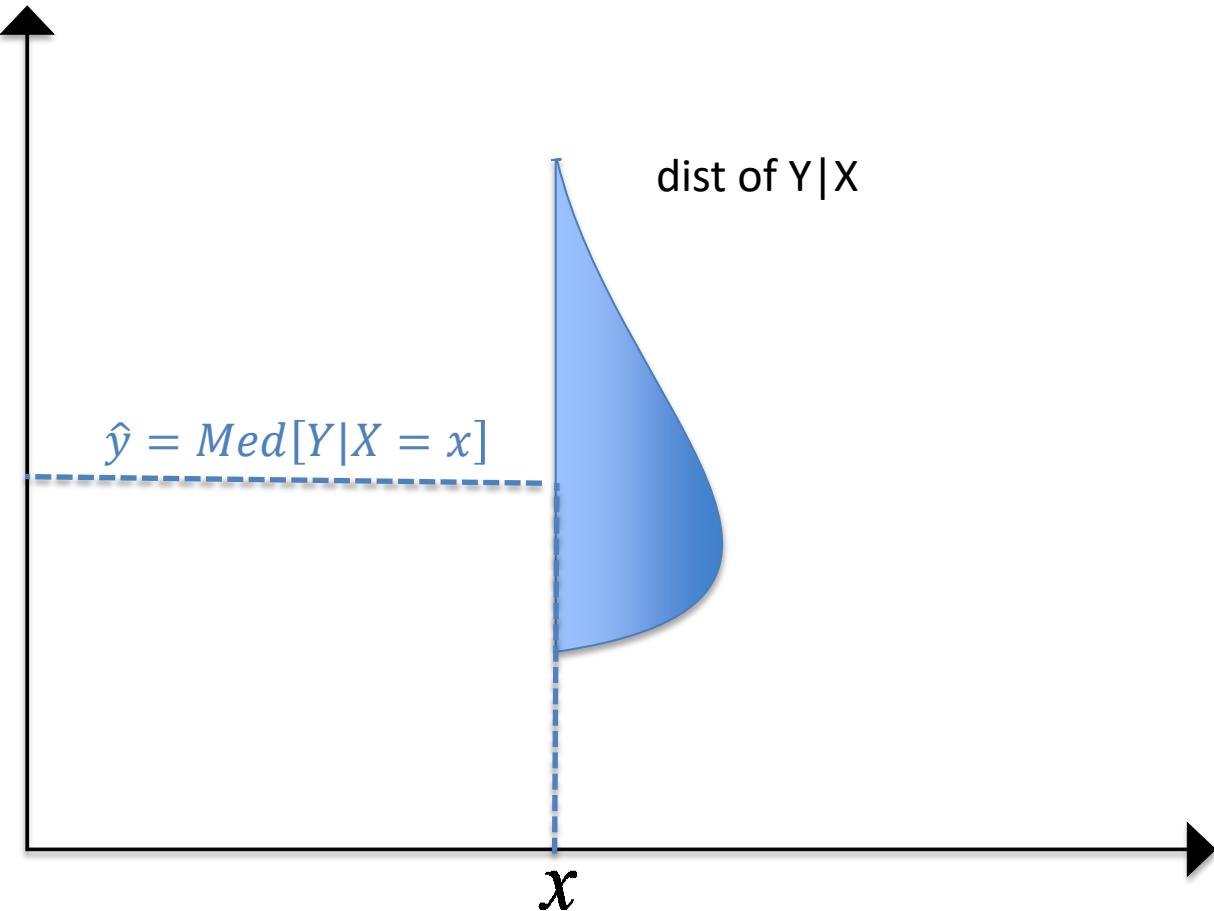


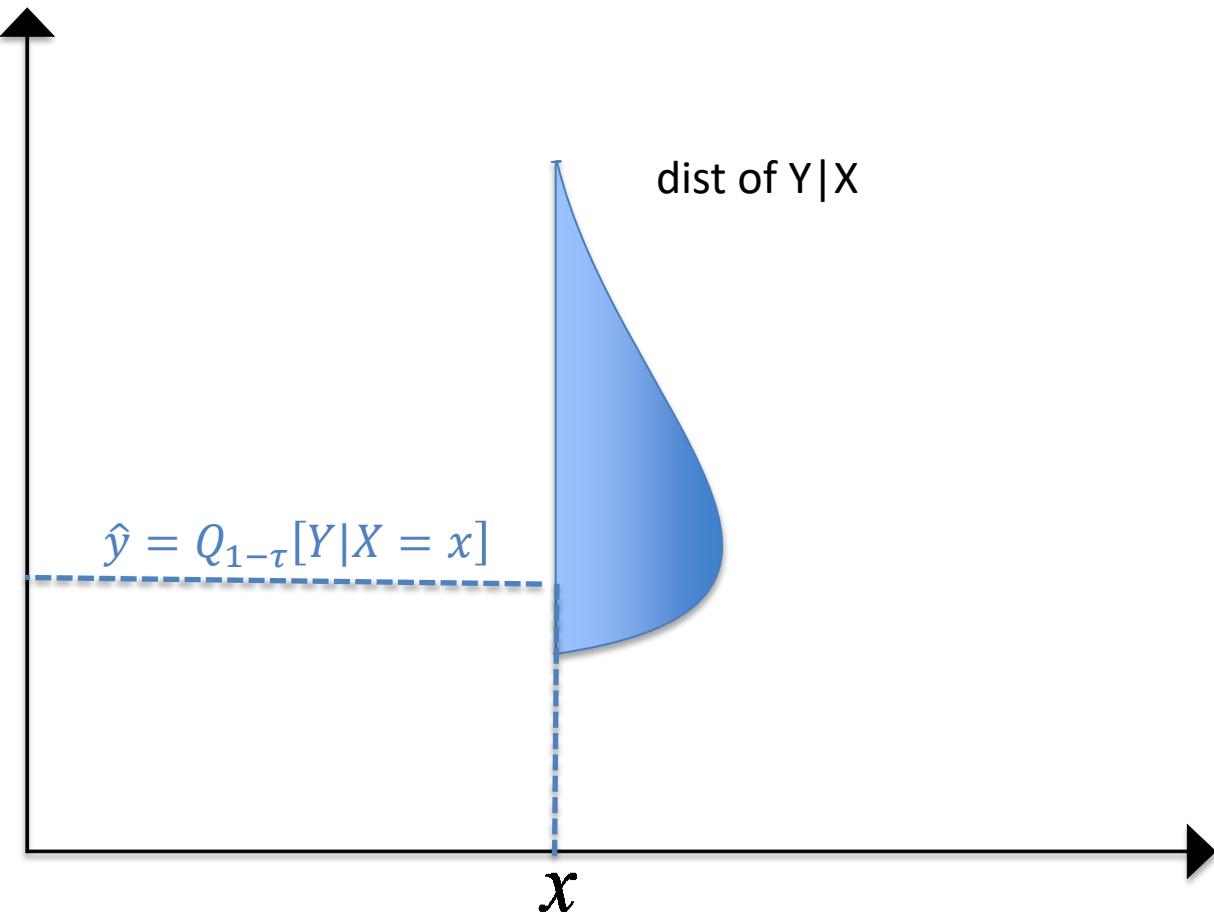




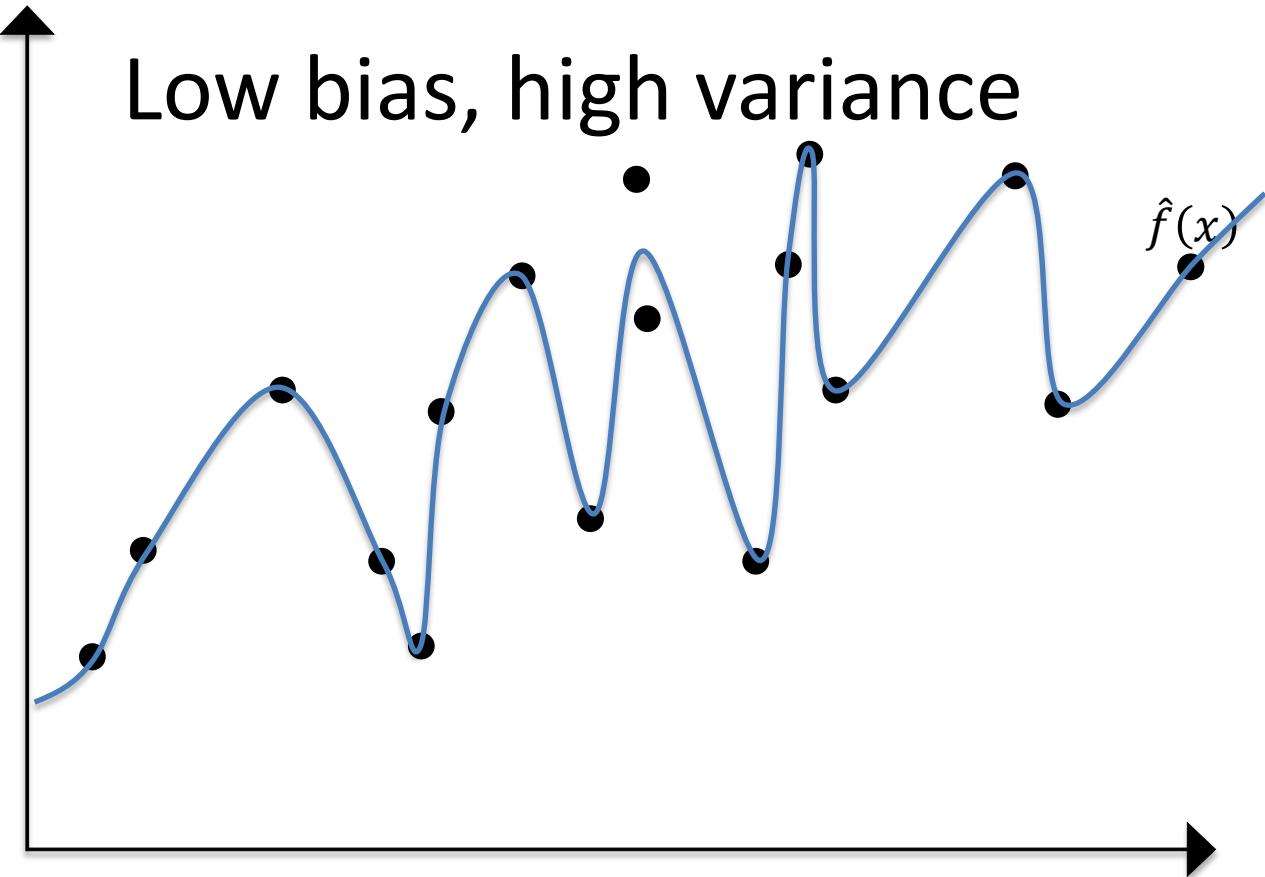


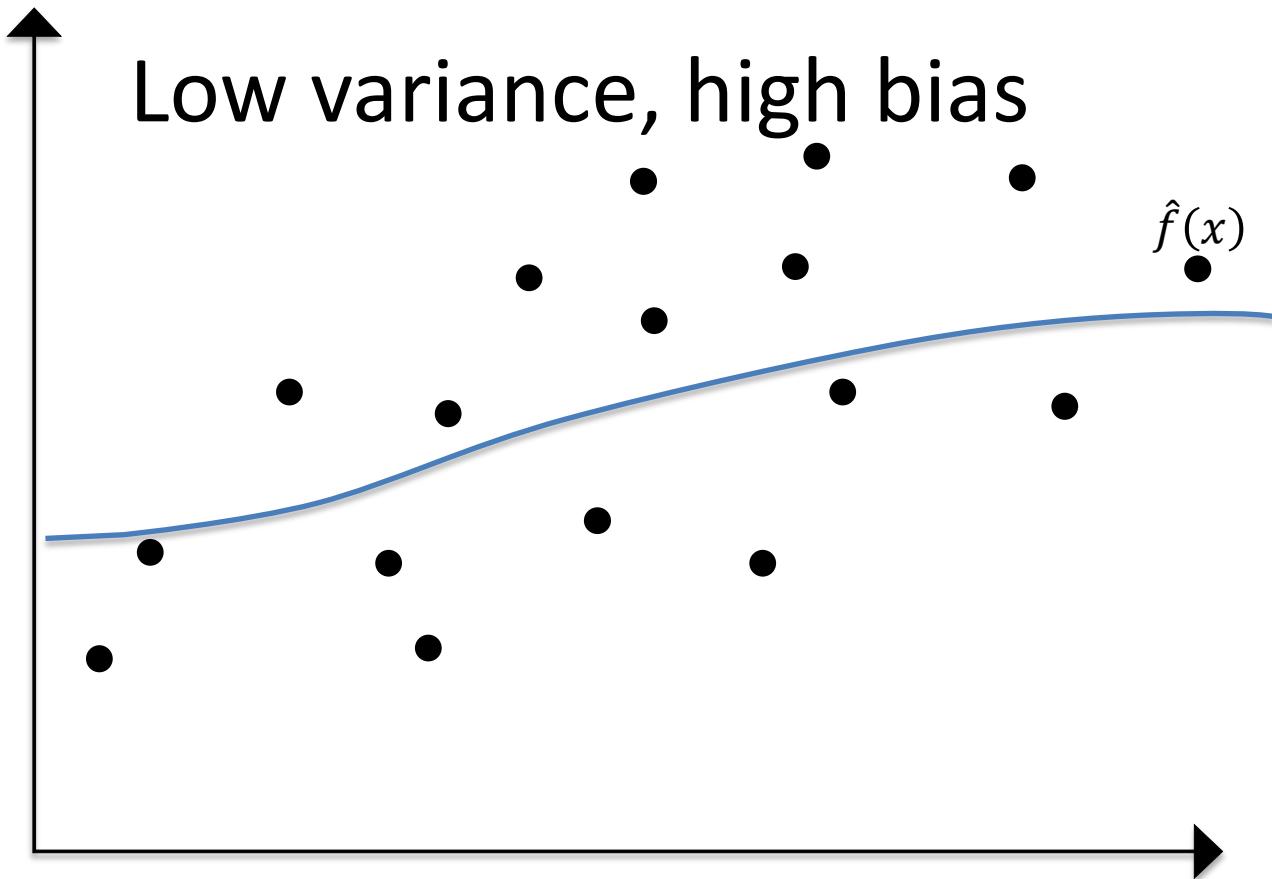






Low bias, high variance





	Y	X1	X2	X3
Fold 1				
Fold 2				
Fold 3				

Initial node

