Probability Calculus (2021)

Bootstrap sampling: what is the fraction of the underlying observations ending up in the bootstrap sample when $N \to \infty$?
sample size.

$X = (X_1, \dots, X_N)$ initial sample (non-random)

$Y = (Y_1, \dots, Y_N)$ bootstraped sample

$$Y_j \sim \frac{1}{N} \sum_i \delta_{X_i} \quad \text{i.i.d.}$$

We are interested in $\mathbb{E}\xi$ where $\xi = \frac{1}{N} \sum_k \mathbb{1}_{\{\exists j : X_k = Y_j\}}$

symmetry of $\mathcal{L}(Y)$

$$\mathbb{E}\xi = \mathbb{P}(\exists j : X_1 = Y_j) = (1 - \mathbb{P}(\forall j \; Y_j \neq X_1)) =$$

Y_j i.i.d

$$= (1 - \mathbb{P}(Y_1 \neq X_1)^N) = \left(1 - \left(\frac{N-1}{N}\right)^N\right) = \Big(1 - \underbrace{\left(1 - \frac{1}{N}\right)^N}_{e^{-1}}\Big) \to$$

$N \to \infty$

$$\to 1 - e^{-1} \simeq 0.6321 \approx \frac{2}{3}$$

Gradient boosting (2022 & 2021)

$$Obj^{(t)} = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i^{(t-1)} + f_{R^{(t)}}(x_i)) + \Omega_{\gamma,\lambda}(f_{R^{(t)}}) + C$$

$$\Omega_{\gamma,\lambda}(f_{R^{(t)}}) = \gamma |T| + \frac{\lambda}{2} \sum_{j=1}^{|T|} w_j^2 \qquad w_j \text{ - value of } f_{R^{(t)}} \text{ on } R_j^{(t)}$$

For binary classification with

$$l(y,x) = -\left[y\log(\sigma(x)) + (1-y)\log(1-\sigma(x))\right]$$

and $\sigma(x) = \frac{1}{1+e^{-x}}$ compute the optimal tree weights $(w_j^t)_{j=1}^{|T|}$

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \Big(\underbrace{\partial_{\hat y}\, l(y_i, \hat y_i^{(t-1)})}_{g_i} f_{R^{(t)}}(x_i) + \frac{1}{2} \underbrace{\partial^2_{\hat y}\, l(y_i, \hat y_i^{(t-1)})}_{h_i} f_{R^{(t)}}(x_i)^2 \Big) +$$

$$\frac{\lambda}{2}\sum_{j=1}^{|T|} w_j^2 + \tilde{C} = \sum_{j=1}^{|T|} \underbrace{\Big(\sum_{i:x_i\in R_i^{(t)}} g_i\Big)}_{G_j} w_j + \frac{1}{2}\Big(\underbrace{\sum_{i:x_i\in R_i^{(t)}} h_i}_{H_j} + \lambda\Big) w_j^2 + \tilde{C} \longrightarrow \min_{(w_j)_{j=1}^{|T|}}$$

$|T|$ simple 1D quadratic optimisation problems:

$$\frac{1}{2}(H_j+\lambda)w_j^2 + G_j w_j \to \min_{w_j}$$

If $H_j > 0 \to w_j^* = -\frac{G_j}{H_j+\lambda}$

In our case $l(y,x) = -\left[y\log(\sigma(x)) + (1-y)\log(1-\sigma(x))\right]$

$$\partial_x l(y,x) = -\frac{y}{\sigma(x)}\sigma'(x) + \frac{1-y}{1-\sigma(x)}\sigma'(x) = -y(1-\sigma(x)) + (1-y)\sigma(x) =$$

$$\left\{ \sigma(x) = \frac{1}{1+e^{-x}} \qquad \sigma'(x) = \frac{1}{(1+e^{-x})^2}e^{-x} = \sigma(x)\cdot(1-\sigma(x)) \right\}$$

$$= -y + \cancel{y\sigma(x)} + \sigma(x) - \cancel{y\sigma(x)} = \sigma(x) - y \qquad g_i = \sigma(\hat y_i^{(t-1)}) - y_i$$

$$\partial_{xx} l(y,x) = \sigma'(x) = \sigma(x)(1-\sigma(x)) \qquad h_i = \sigma(\hat y_i^{(t-1)})(1-\sigma(\hat y_i^{(t-1)}))$$

$$G_j = \sum_{i:x_i\in R_i^{(t)}} \left(\sigma(\hat y_i^{(t-1)}) - y_i\right)$$

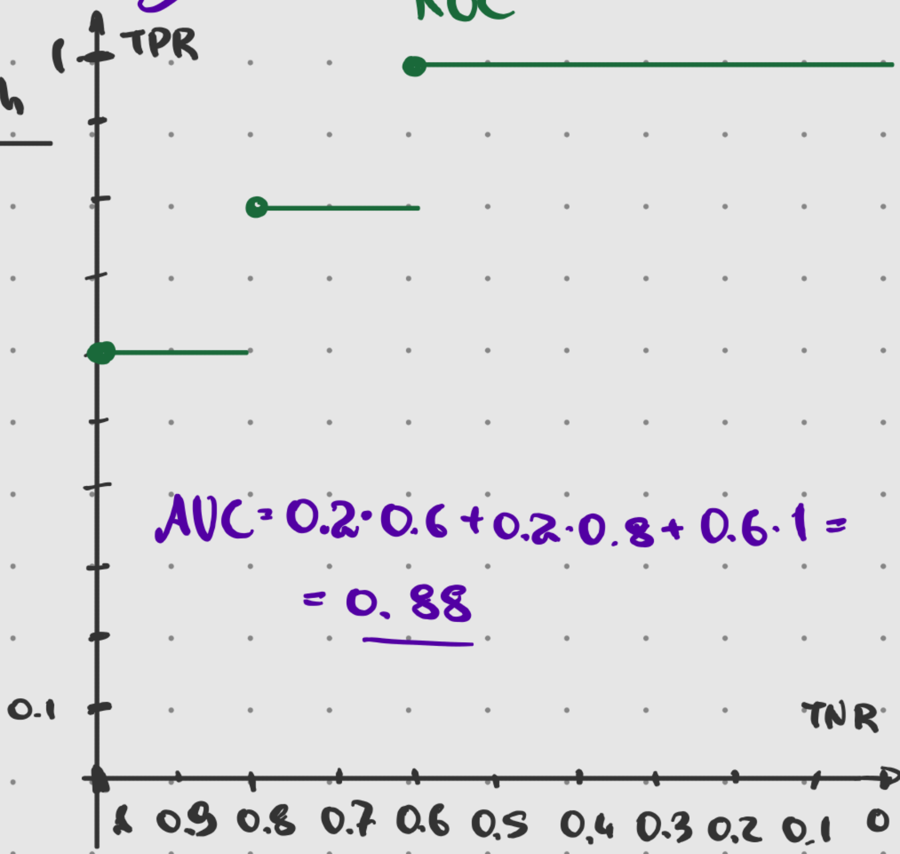
$$H_j = \left(\sum_{i: x_i \in R_j^{(t)}} \sigma(\hat{y}_i^{(t-1)})\left(1-\sigma(\hat{y}_i^{(t-1)})\right) + \lambda\right) \qquad w_j^* = -\frac{G_j}{H_j+\lambda} \quad j = 1, \dots, N$$

Evaluation of ML trainings (2023)

Probas	Ground truth
0.1	A
0.2	A
0.3	A
0.4	B
0.5	A
0.6	B
0.7	A
0.8	B
0.9	B
1.0	B

$p = \mathbb{P}(X_i = B)$

ROC



$$AUC = 0.2 \cdot 0.6 + 0.2 \cdot 0.8 + 0.6 \cdot 1 =$$
$$= \underline{0.88}$$

ROC curve = (TNR, TPR)

$$TNR = \frac{TN}{TN + FP}$$

$$TPR = \frac{TP}{TP + FN}$$

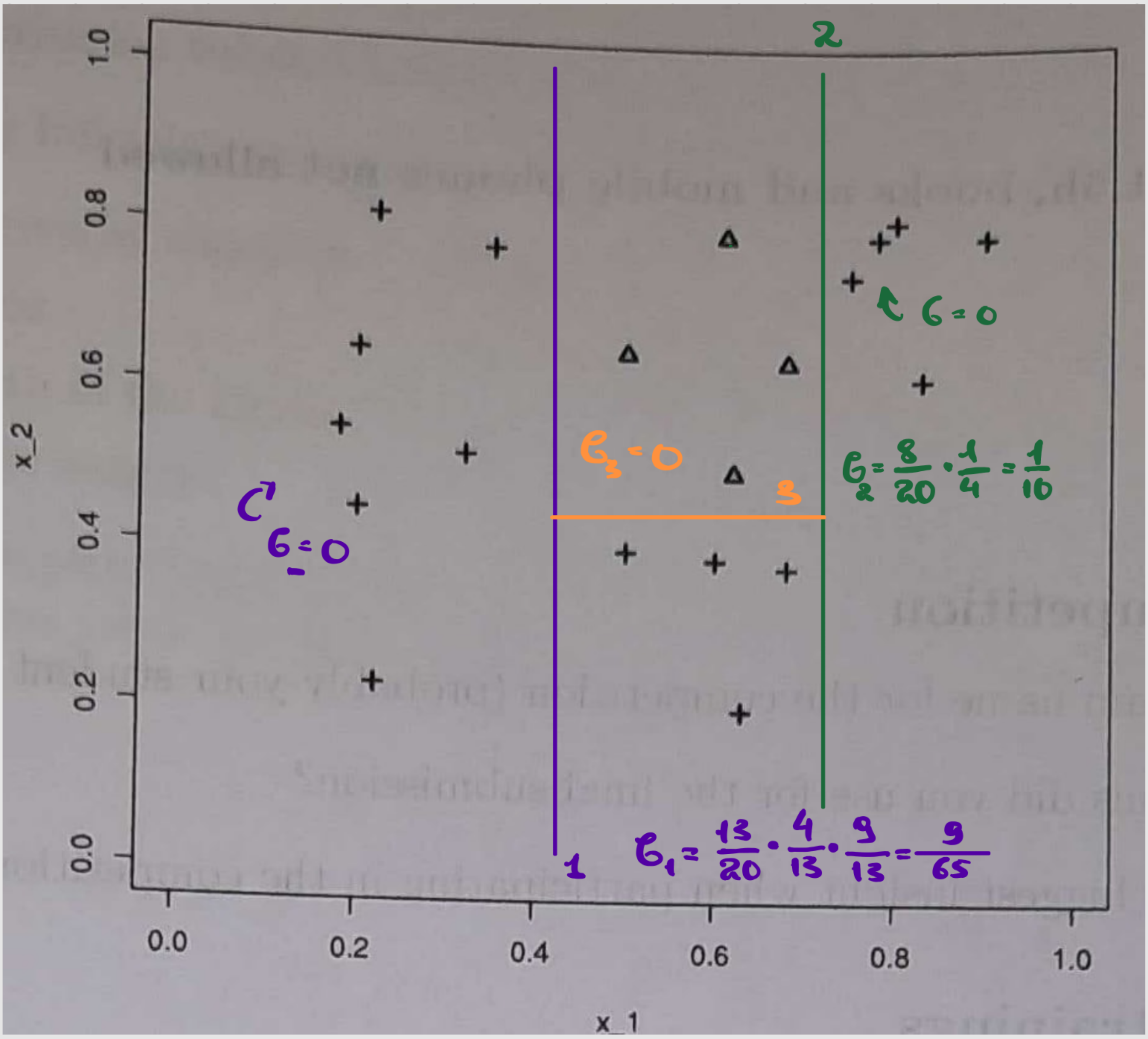
θ threshold : $p > \theta \Rightarrow B$

$\theta = 0.75$ TN = 1
TPR = 0.6

$\theta = 0.55$ TNR = 0.8
TPR = 0.8

$\theta = 0.35$ TNR = 0.6
TPR = 1

Decision tree (2023)



$$G = \sum_{k=1}^{K} p_{mk}(1 - p_{mk})$$

proportion of observations of class k

In our case $G = 2p\,(1-p)$

prop. of observations of class 1.

$N^{+} = 16$ - class 0

$N^{\Delta} = 4$ - class 1

$p = 0.2 \rightarrow G = 0.4 \cdot 0.8 = 0.32$

When spleating region R into R^+ and R^-

we compare G with $\frac{n^+}{n^R} G^+ + \frac{n^-}{n^R} G^- =$

$$\frac{7}{20} \cdot 0 + \frac{13}{20} \cdot \frac{9}{13} \, \frac{4}{13}$$

$$= 2\left(\frac{n^+}{n^R} \cdot \frac{n^{+,\Delta}}{n^+} \cdot \frac{n^{+,*}}{n^+} + \frac{n^-}{n^R} \cdot \frac{n^{-,*}}{n^-} \cdot \frac{n^{-,\Delta}}{n^-} \right) = \frac{2}{n^R} \left(\frac{n^{+,\Delta} \cdot n^{+,*}}{n^+} + \frac{n^{-,\Delta} \cdot n^{-,*}}{n^-} \right)$$

If $n^{+,\Delta} = 0$, $\quad \frac{n^{-,\Delta}(n^{-,*}+1)}{(n^-+1)} > \frac{n^{-,\Delta} \cdot n^{-,*}}{n^-} \rightarrow$

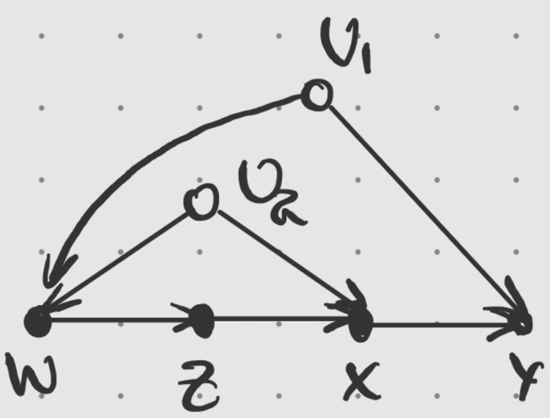
$\rightarrow$ one should separate as much '*' as possible

$$G_0 - G_1 = \frac{32}{100} - \frac{9}{65} \simeq 0.1815$$

$$G_1 - G_2 = \frac{9}{65} - 0.1 \simeq 0.0385$$

$$G_2 - G_3 = 0.1$$

Causal inference (2023)



U_1, U_2 unobserved

1) Why we cannot use back-/front-door criter?

No front-door because no intermediates between

X and Y.

No back-door because one should block $U_2 \to W$ but W is a collider $\to$ creates a path $U_1 \text{---} U_2$ when conditioning on W.

Do-calculus rules

graph without arrows to X

Insertion/deletion of observations

1) $P(Y \mid do(X), Z, W) = P(Y \mid do(X), W)$ if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$

without arrows from Z

Action/Observation exchange

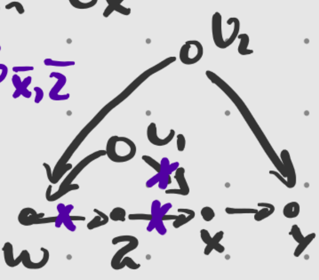
2) $P(Y \mid do(X), do(Z), W) = P(Y \mid do(X), Z, W)$ if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$

Insertion/deletion of actions

3) $P(Y \mid do(X), do(Z), W) = P(Y \mid do(X), W)$ if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}}$

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$

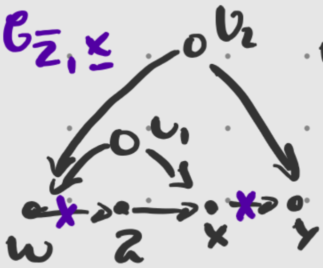
$G_{\overline{X}, \overline{Z}}$



(3) $(Z \perp\!\!\!\perp Y \mid X)$, $Z(W) = Z$

$P(Y \mid do(X)) \overset{(3)}{=}$

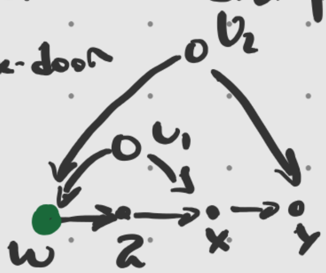
$G_{\overline{Z}, \underline{X}}$



$= P(Y \mid do(X), do(Z)) \overset{(2)}{=}$

$= P(Y \mid X, do(Z)) = \{$ def. of condit. probas $\}$

$= \dfrac{P(Y, X \mid do(Z))}{P(X \mid do(Z))} \overset{\text{Back-door}}{=}$



Back-door for $Z \to (X, Y)$
W

$= \dfrac{\sum_w P(Y, X \mid Z, w)}{\sum_w P(X \mid Z, w)}$

Causal inference (2021/2022)

$(Z \to Y)$

Fraction r of indiv. suffer from a fatal syndrome Z. $(P(Z=1)=r)$

It makes it uncomf. to take life prolonging drug X $(Z \to X)$

$Y = \begin{cases} 1, \text{ death} \\ 0, \text{ survival} \end{cases}$ $\quad X = \begin{cases} 1, \text{ taking the drug} \\ 0, \text{ otherwise} \end{cases}$

$$P(\underbrace{Y=1}_{y_1} \mid \underbrace{Z=0}_{z_0}, \underbrace{X=0}_{x_0}) = p_1$$

$$P(Y=1 \mid Z=0, X=1) = p_2$$

$$P(Y=1 \mid Z=1, X=0) = p_3$$

$$P(Y=1 \mid Z=1, X=1) = p_4$$

$p_1 > p_2$ (life-prolonging)

$$q_1 = P(X=1 \mid Z=0)$$

$$q_2 = P(X=1 \mid Z=1)$$

$q_2 < q_1$

(1) Draw the causal diagram

$$\begin{tikzcd} X \arrow[rr] & & Y \\ & Z \arrow[ul] \arrow[ur] & \end{tikzcd}$$

(2) Compute the average causal effect

$$P(Y=1 \mid do(X=1)) - P(Y=1 \mid do(X=0))$$

Z closes back-door $X \leftarrow Z \rightarrow Y$ $\rightarrow$

$$P(Y \mid do(X), Z) = P(Y \mid X, Z)$$

$$P(Y=1 \mid do(X=1)) = \overbrace{P(Z=1)}^{r}\, \underbrace{P(Y=1 \mid do(X=1), Z=1)}_{P(Y=1 \mid \text{"}X=1, Z=1)} +$$

$$+ \overbrace{P(Z=0)}^{1-r}\, P(Y=1 \mid do(X=1), Z=0) = r p_4 + (1-r) p_2$$

$$P(Y=1 \mid do(X=0)) = \overbrace{P(Z=1)}^{r}\, \underbrace{P(Y=1 \mid do(X=0), Z=1)}_{P(Y=1 \mid \text{"}X=0, Z=1)} +$$

$$+ \overbrace{P(Z=0)}^{1-r}\, P(Y=1 \mid do(X=0), Z=0) = r p_3 + (1-r) p_1$$

Average causal effect $= r(p_4 - p_3) + (1-r)(p_2 - p_1)$