Decision making based on model predictions: model quality and benefit curves

Course roadmap

1. Project valuation: valuation metrics, planning and rules
2. Model quality and decision making. Benefit curve
    - Simple threshold decisions
    - Benefit curve
    - Model quality metrics and benefit curve
    - Functional threshold decisions
    - Increment based threshold decisions
3. Estimating model risk discounts
4. A/B testing and financial result verification
5. Unobservable model errors, metalearning

Simple threshold decisions

Simple threshold-based decision making

Consider a binary classification model $X \longrightarrow Prob$, e.g. credit scoring.

$$X_1 \dots X_k \rightarrow Prob$$

34		0.94
20		0.25
21		0.16
50	…	0.86
41		0.51
25		0.33
19		0.27
82		0.12

$X_1 \dots, X_k$ – model factors

$Prob$ – model prediction

Simple threshold-based decision making

Simple threshold decision is a level $\boldsymbol{a}: if\ Prob > a$ then some action is undertaken, i.e. $\hat{Y} = 1\ and\ otherwise\ \hat{Y} = 0$

$X_1 \ldots$	$X_k \rightarrow$	$Prob$	$\hat{Y}$
34		0.94	1
20		0.25	0
21		0.16	0
50	…	0.86	1
41		0.51	1
25		0.33	0
19		0.27	0
82		0.12	0

Example: $a = 0.5$

if Prob > 0.5
then reject loan application

($\hat{Y}$=1)

Simple threshold-based decision making

Acceptance rate $\boldsymbol{c}$ is a percentage of observations that satisfy rule:

$if\ Prob \leq a$, i.e. $c = \sum_{i=1}^{N} I\{Prob_i \leq a\}/N$

X_1	$\dots X_k \rightarrow$	$Prob$	$\hat{Y}$	Y
34		0.94	1	
20		0.25	0	
21		0.16	0	
50	…	0.86	1	
41		0.51	1	
25		0.33	0	
19		0.27	0	
82		0.12	0	

N=8

Example: $a = 0.5$

$c = 5/8$

if Prob > 0.5

then 5/8 clients

are accepted

$(\hat{Y}=0)$

Simple threshold-based decision making

On historical data we can estimate errors of I and II type for each threshold level a (and also acceptance rate c)

$a = 0.5$

$c = 5/8$

$X_1 \dots X_k$		$\to$ *Prob*	$\hat{Y}$	Y	
34		0.94	1	1	→ True Positive
20		0.25	0	0	
21		0.16	0	1	→ False Negative (II type)
50	…	0.86	1	1	
41		0.51	1	0	→ False Positive (I type)
25		0.33	0	1	
19		0.27	0	0	
82		0.12	0	0	→ True Negative

Model prediction errors matrix

		Observed (Y)	
		Default	No-Default
Model prediction ($\hat{Y}$)	Reject	True Positive	FP *(I type)*
	Accept	FN *(II type)*	True Negative
		Positive (P) Total # of default (bad) clients	Negative (N) Total # of creditworthy (good) clients

Financial effect and model prediction errors

Financial result of model performance depends on FP and FN error types:

$$\underbrace{\overset{\text{\# of accepted good clients}}{\boxed{N - FP}} \times \overset{\text{Credit margin}}{\boxed{\text{margin}}}}_{\text{Interest Income}} - \underbrace{\overset{\text{\# of accepted bad clients}}{\boxed{FN}} \times \overset{\text{Loss given default}}{\boxed{LGD}}}_{\text{Credit loss}}$$

N - total # of creditworthy (good) clients

Financial effect and model prediction errors

$$(N - FP) \times \underbrace{e_{FP}}_{\text{I type error cost}} - FN \times \underbrace{e_{FN}}_{\text{II type error cost}}$$

Thus, financial result is negative to FP and FN model errors

Wrap-up

1. Financial result is negative to model prediction errors

Wrap-up

1. Financial result is negative to model prediction errors
2. Two types of errors FP and FN, generally, have different impact on financial result

Benefit curve

Data for benefit curve

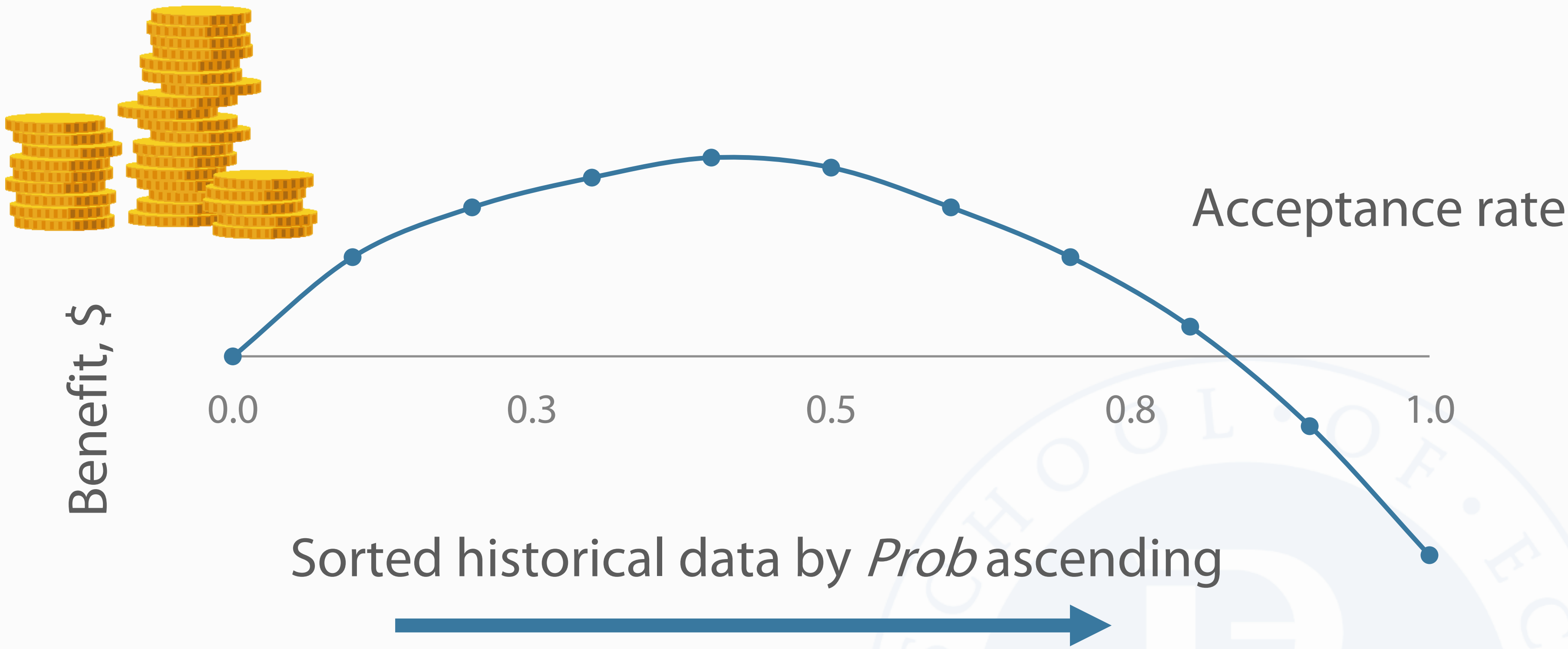
Sort historical data by *Prob* ascending

$X_1 \dots X_k$		$\rightarrow$ *Prob*	$\hat{Y}$	Y
82		0,12	0	0
21		0,16	0	1
20		0,25	0	0
19	…	0,27	0	0
25		0,33	0	1
41		0,51	1	0
50		0,86	1	1
34		0,94	1	1

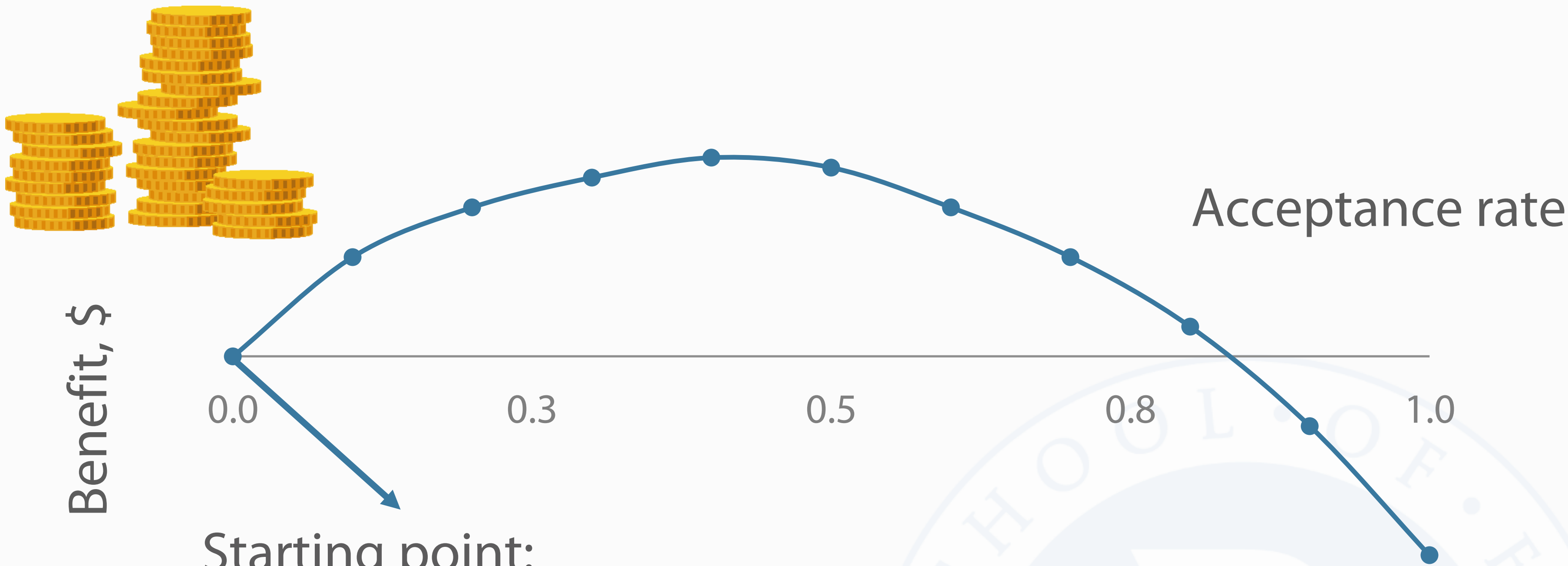
Acceptance rate $c = 5/8$

Threshold level $a = 0.5$

Benefit curve. Threshold vs financial result



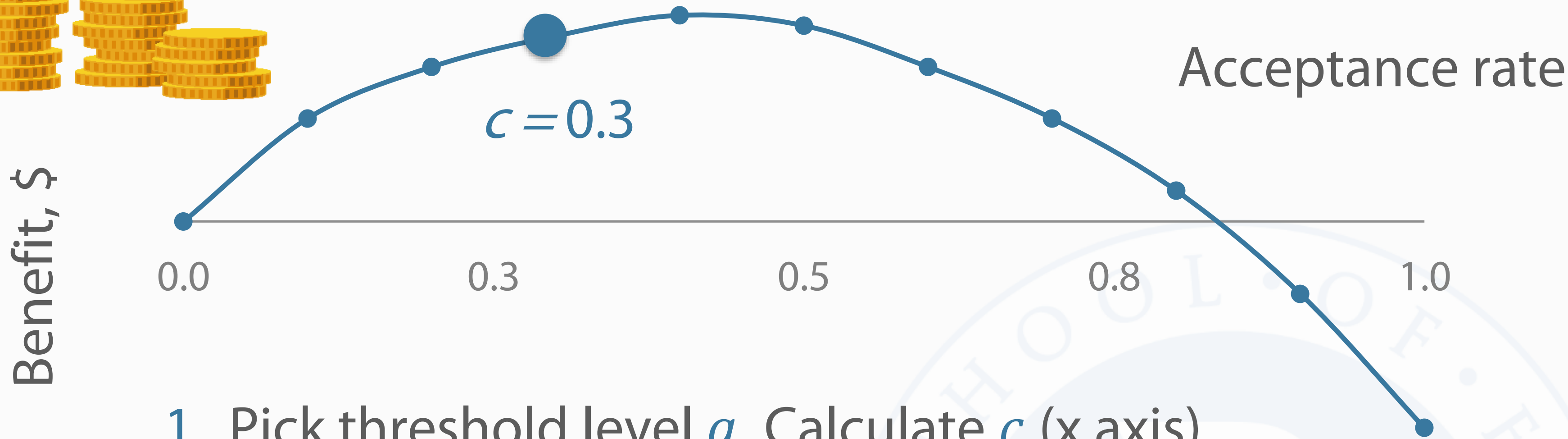
Benefit curve. Threshold vs financial result



Starting point:
Threshold level $a = 0$
Acceptance rate $c = 0$

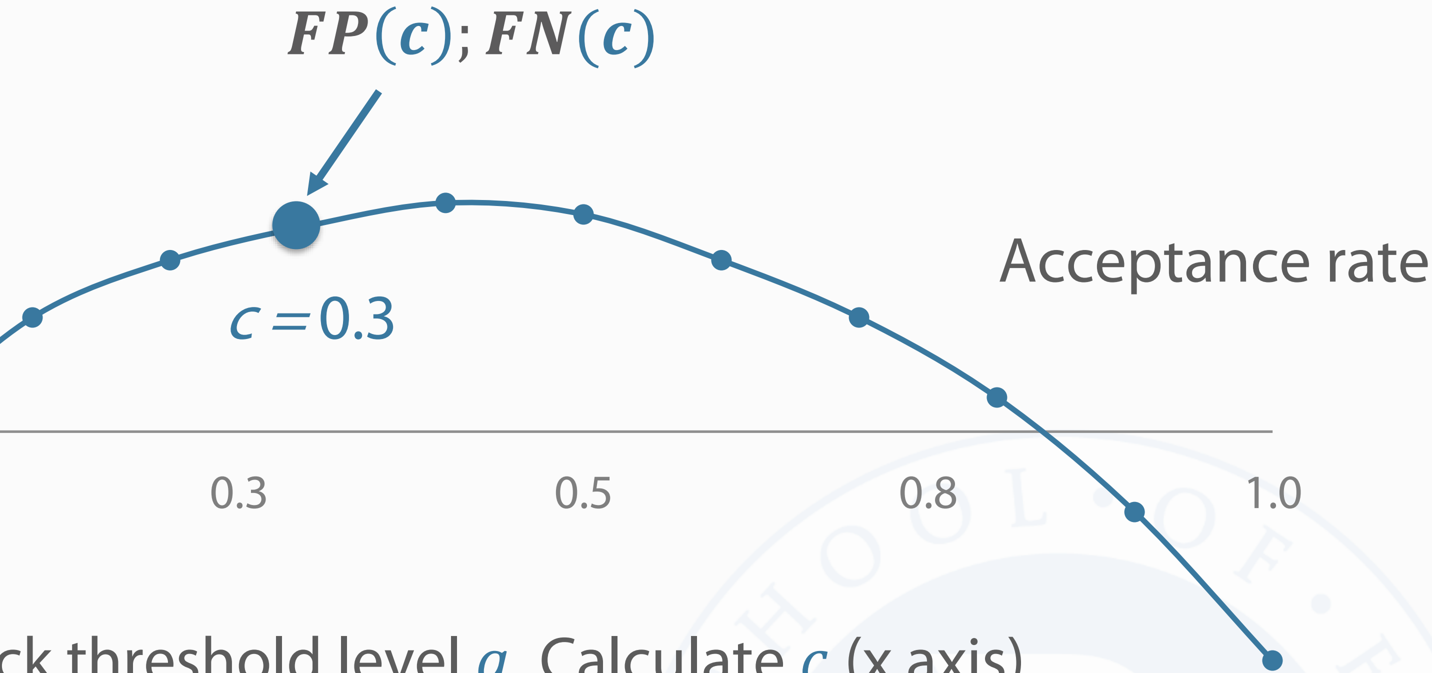
Reject if $Prob > 0$,
i.e. no loss and no gain
Zero benefit

Benefit curve. Threshold vs financial result



1. Pick threshold level a. Calculate c (x axis)

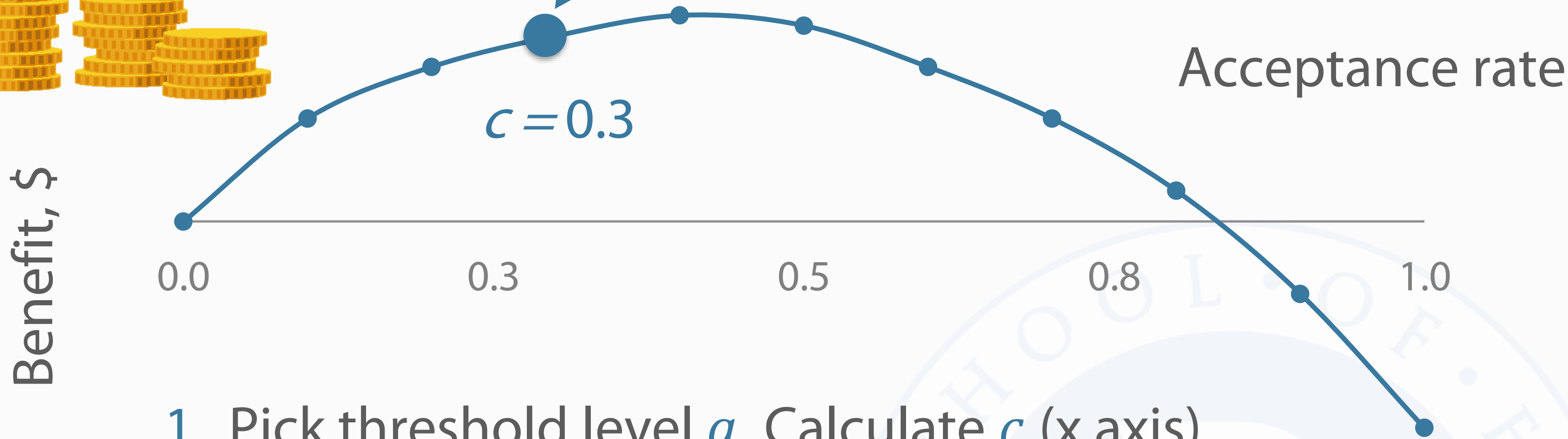
Benefit curve. Threshold vs financial result



1. Pick threshold level a. Calculate c (x axis)
2. Calculate *FP* and *FN* for given c

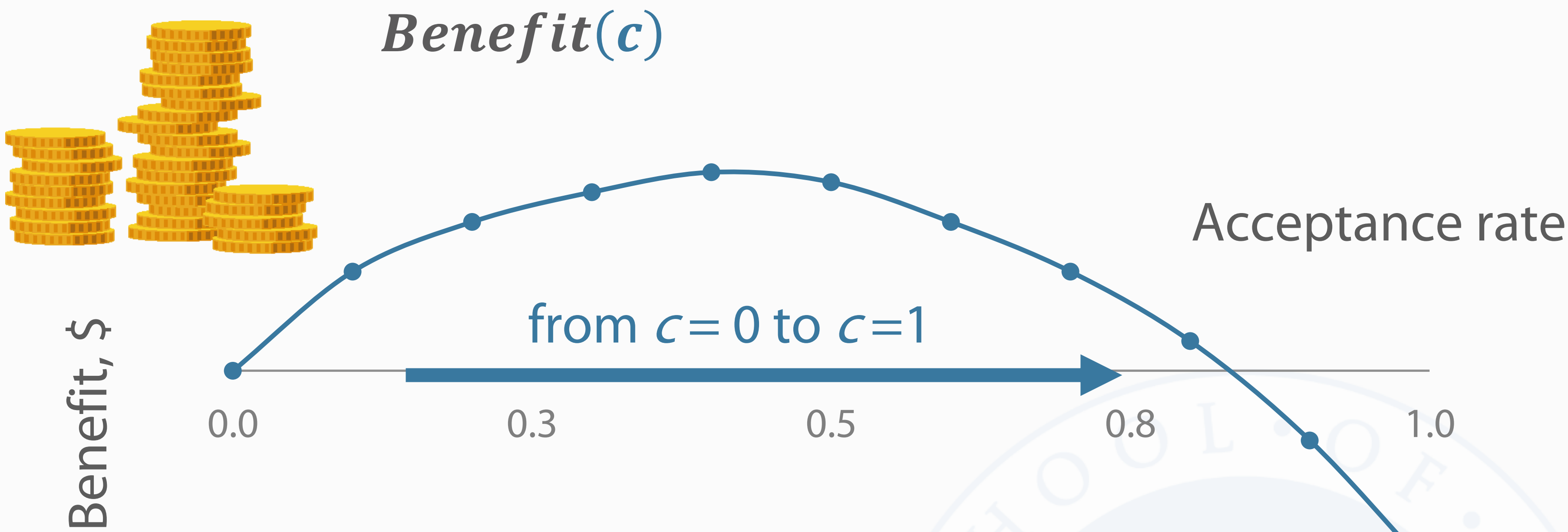
Benefit curve. Threshold vs financial result

$$\left(N - \boldsymbol{FP}(\boldsymbol{c})\right) \cdot e_{FP} - \boldsymbol{FN}(\boldsymbol{c}) \cdot e_{FN}$$



1. Pick threshold level a. Calculate c (x axis)
2. Calculate *FP* and *FN* for given c
3. Weigh *FP* and *FN* with error costs (e_{FP} and e_{FN}) and plot on y axis

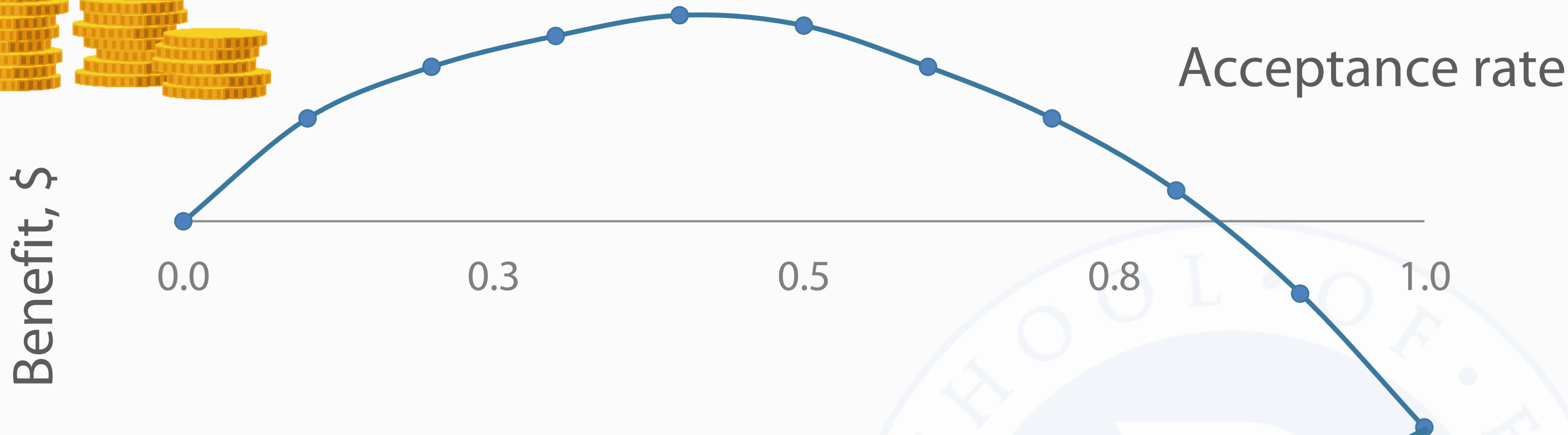
Benefit curve. Threshold vs financial result



1. Pick threshold level a. Calculate c (x axis)
2. Calculate FP and FN for given c
3. Weigh FP and FN with error costs (e_{FP} and e_{FN}) and plot on y axis

Reiterate 1-3 from $c = 0$ to $c = 1$

Benefit curve. Threshold vs financial result



Endpoint:
Threshold level $a = 1$
Acceptance rate $c = 1$
Reject if $Prob > 1$,
i.e. accept every loan application

Benefit curve. Threshold vs financial result

$$\boldsymbol{Benefit}(\mathbf{1}) = \left(N - \boldsymbol{FP}(\mathbf{1})\right) \cdot e_{FP} - \boldsymbol{FN}(\mathbf{1}) \cdot e_{FN} =$$

Benefit curve. Threshold vs financial result

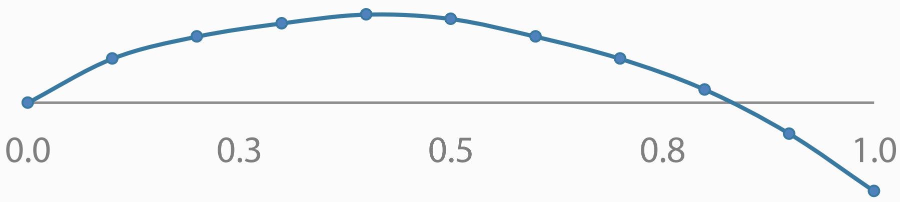
$$\boldsymbol{Benefit}(1) = \left(N - \boldsymbol{FP}(1)\right) \cdot e_{FP} - \boldsymbol{FN}(1) \cdot e_{FN} =$$

$$= (N - \mathbf{0}) \cdot e_{FP} - \boldsymbol{P} \cdot e_{FN} =$$

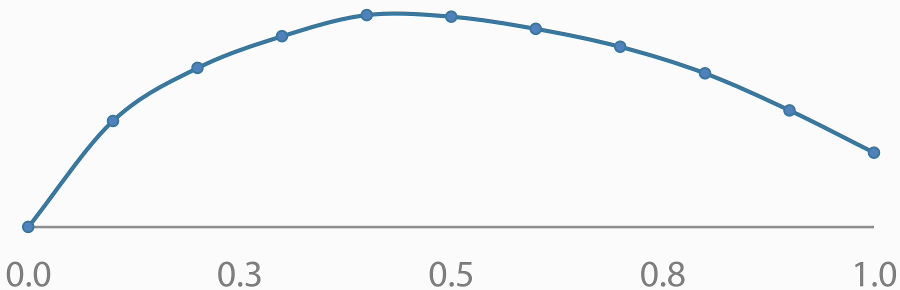
$$= N \cdot e_{FP} - P \cdot e_{FN}$$

Benefit curve. Endpoint values

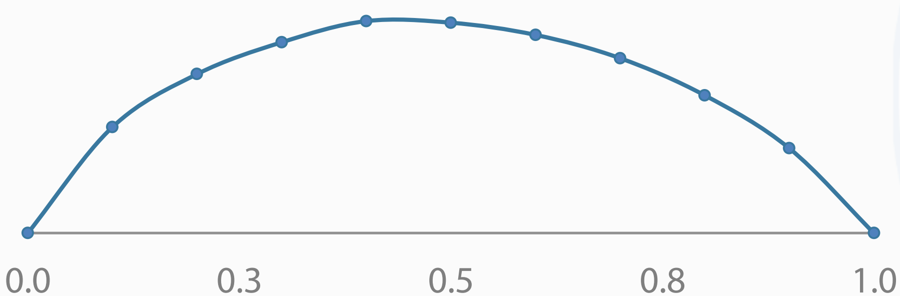
$$Benefit(1) = N \cdot e_{FP} - P \cdot e_{FN} \text{ vs } 0?$$



< 0 $\qquad \frac{N}{P} < \frac{e_{FN}}{e_{FP}}$



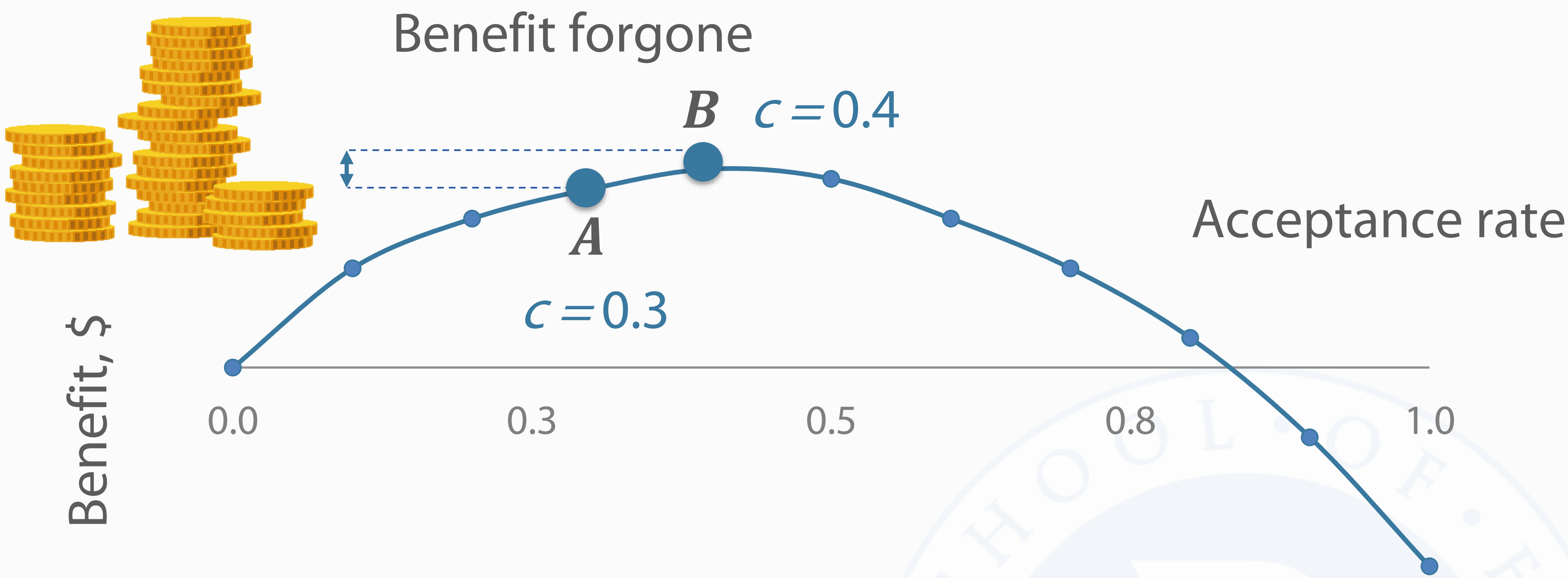
> 0 $\qquad \frac{N}{P} > \frac{e_{FN}}{e_{FP}}$



$= 0$ $\qquad \frac{N}{P} = \frac{e_{FN}}{e_{FP}}$

Class balance I and II errors ratio

Benefit curve. Sub-optimal model use



Sub-optimal decision making:

$Benefit(\boldsymbol{B}) - Benefit(\boldsymbol{A})$

Wrap-up

1. Decision making scheme has a huge impact on financial result

Wrap-up

1. Decision making scheme has a huge impact on financial result
2. It can easily ruin a model even with 100% quality metrics

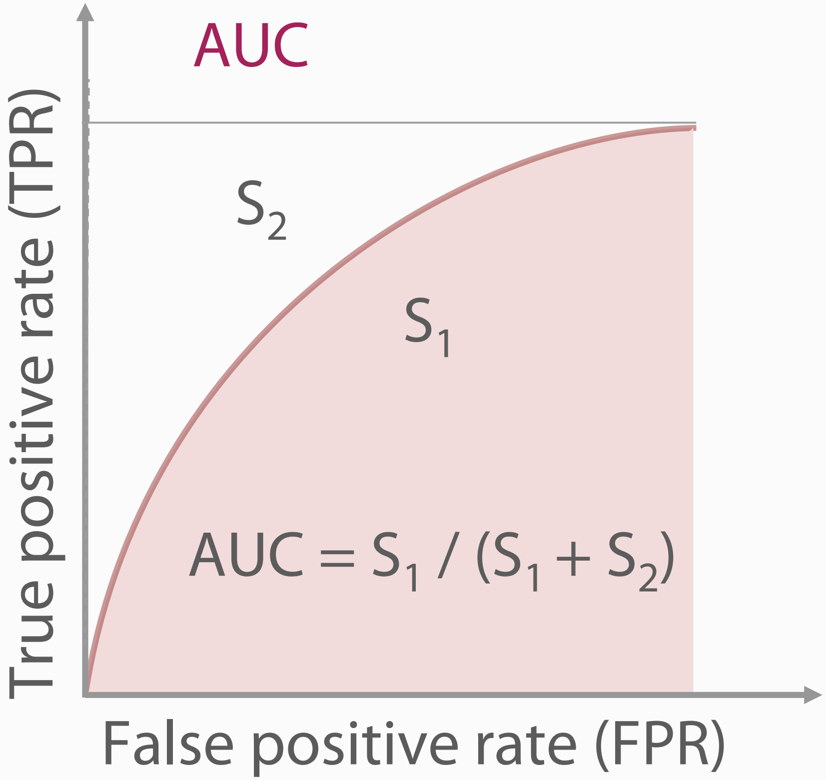
Wrap-up

1. Decision making scheme has a huge impact on financial result
2. It can easily ruin a model even with 100% quality metrics
3. Sometimes, it’s easier and more efficient to improve decision strategy than a model itself

Benefit curve and model quality metrics

Model quality metrics and prediction errors

Area under curve (AUC) aggregates all possible values of TP and FP rates



AUC – model quality metrics

$$AUC = \int_0^1 TPR\, dFPR$$

Model quality metrics and prediction errors

$$\mathrm{AUC} = \int_0^1 TPR\, dFPR \rightarrow$$

$$TPR = \frac{TP}{P} = \frac{P - FN}{P} = $$
$$= 1 - \frac{FN}{P} = 1 - FNR$$

$\rightarrow$

Model quality metrics and prediction errors

$$AUC = \int_0^1 TPR\, dFPR \rightarrow \left| \begin{aligned} TPR &= \frac{TP}{P} = \frac{P - FN}{P} = \\ &= 1 - \frac{FN}{P} = 1 - FNR \end{aligned} \right| \rightarrow$$

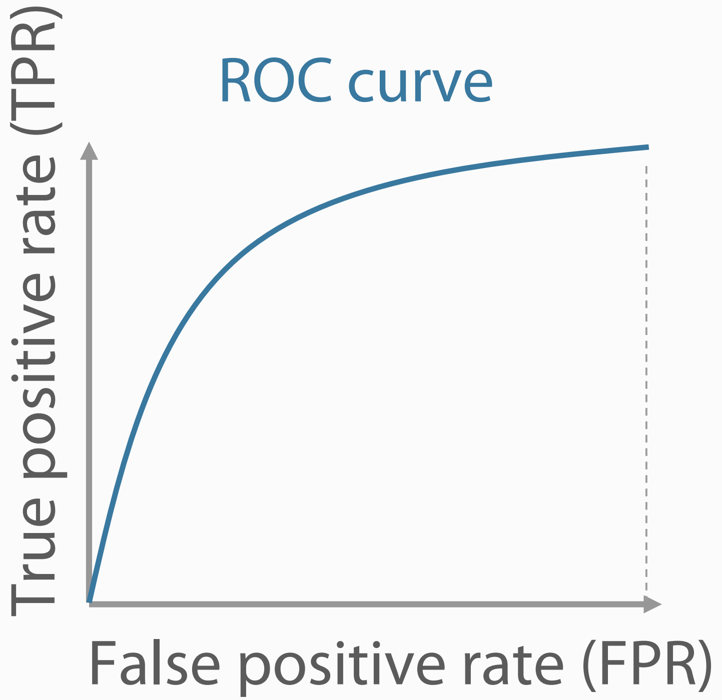
$$AUC = \int_0^1 (1 - FNR)\, dFPR =$$

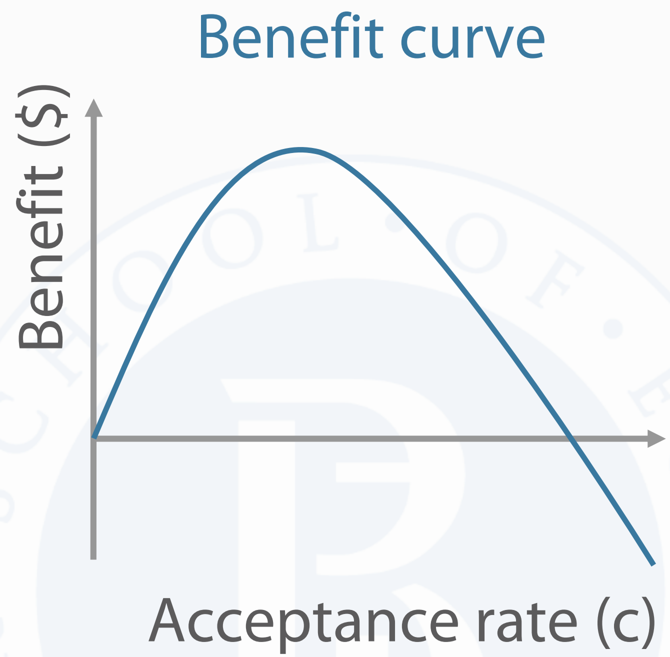
$$= 1 - \int_0^1 \boldsymbol{FNR} d\boldsymbol{FPR}$$

Thus, model quality metrics is negative to FP and FN model errors

Benefit curve and model quality metrics

$$\boldsymbol{Benefit}(c) \rightarrow max \quad w.r.t. \quad c$$
$$s.t. \ \boldsymbol{TPR}(c) \leq ROC\big(\boldsymbol{FPR}(c)\big)$$



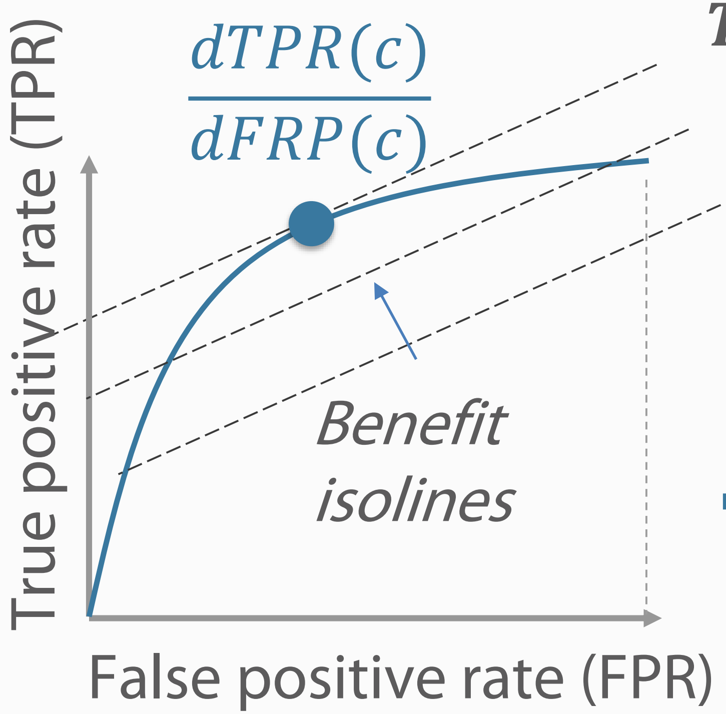


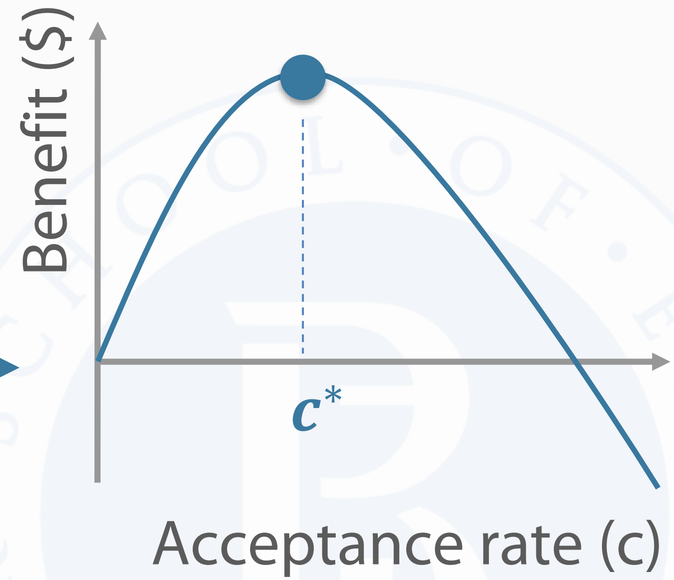
Benefit curve and model quality metrics

$$Benefit(c) = Ne_{FP} \cdot (1 - \boldsymbol{FPR}(c)) - Pe_{FN} \cdot (1 - \boldsymbol{TPR}(c))$$

$\rightarrow max$ w.r.t. c s.t. ROC curve

$$\boldsymbol{TPR}(c) = const + \frac{N/P}{e_{FN}/e_{FP}} \cdot \boldsymbol{FPR}(c)$$

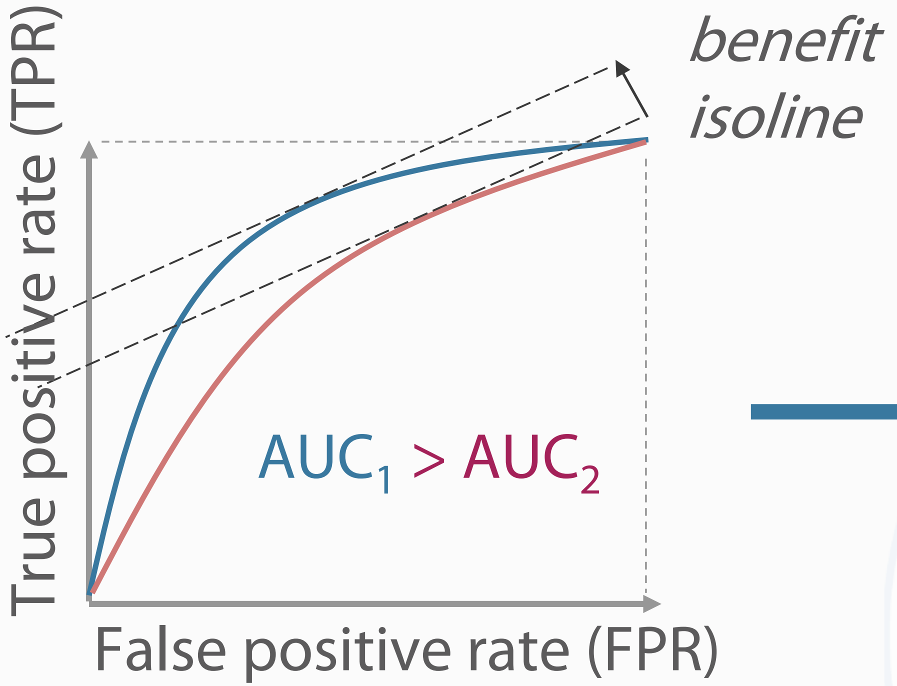


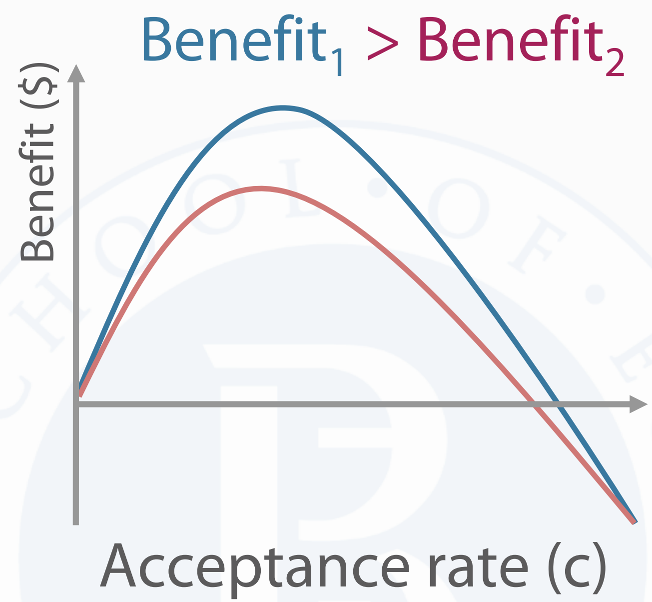


Optimal threshold level c^*: $\frac{dTPR(c)}{dFPR(c)} = \frac{N/P}{e_{FN}/e_{FP}}$

Benefit curve and model quality metrics

Typical case is that with higher AUC one obtains higher benefit

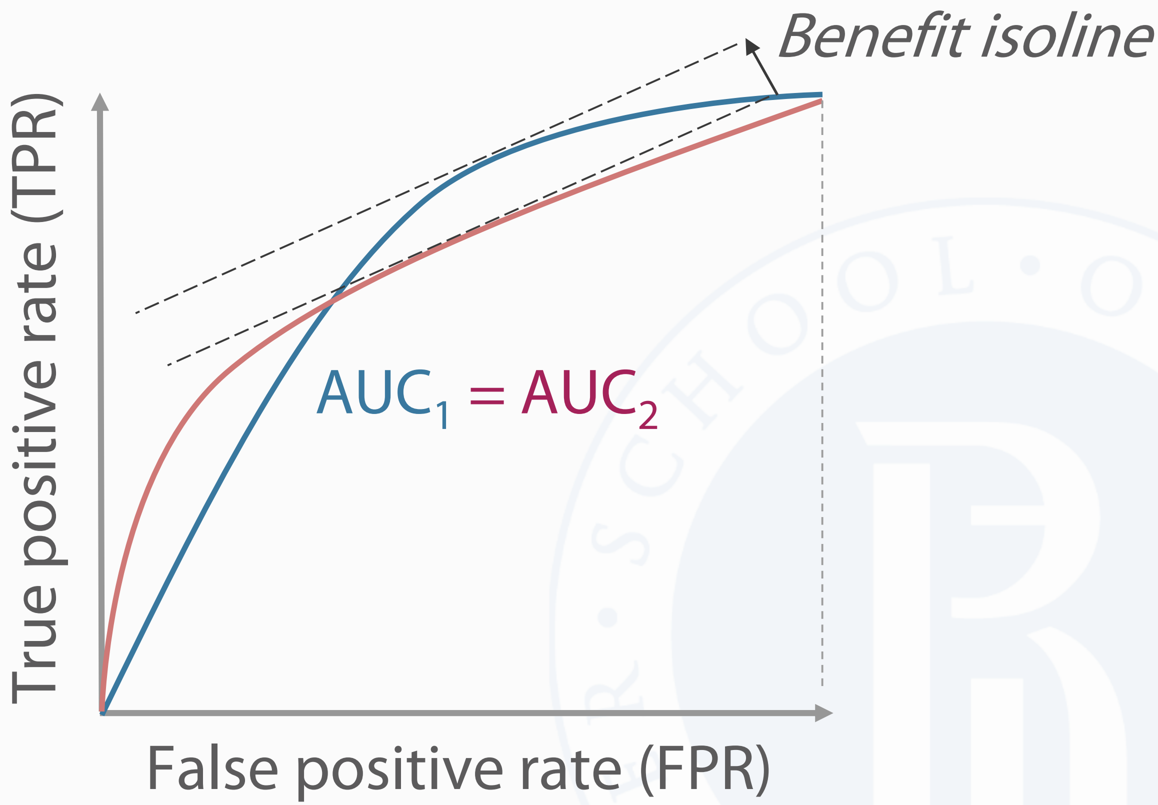




Benefit curve and model quality metrics

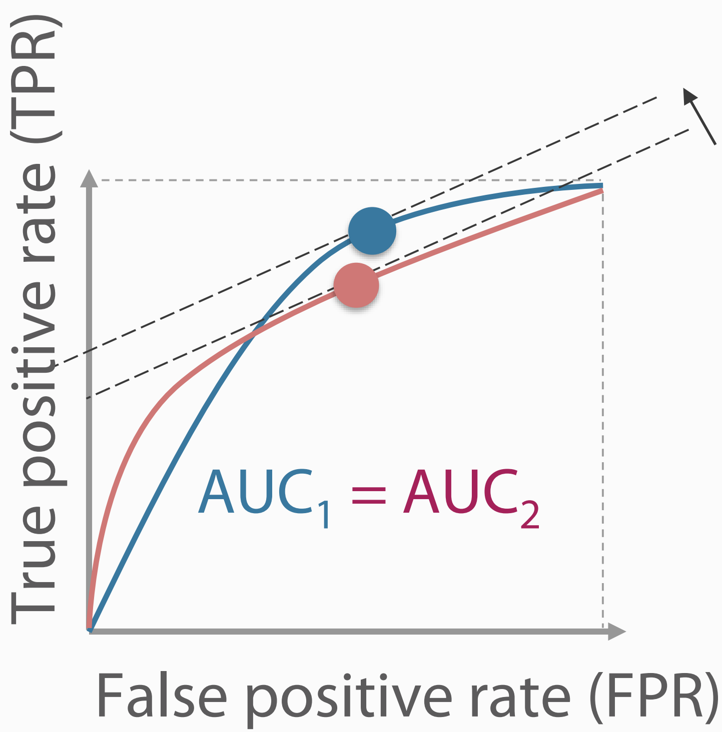
Same AUC but different benefits

However, not only AUC value matters, but also ROC curve itself

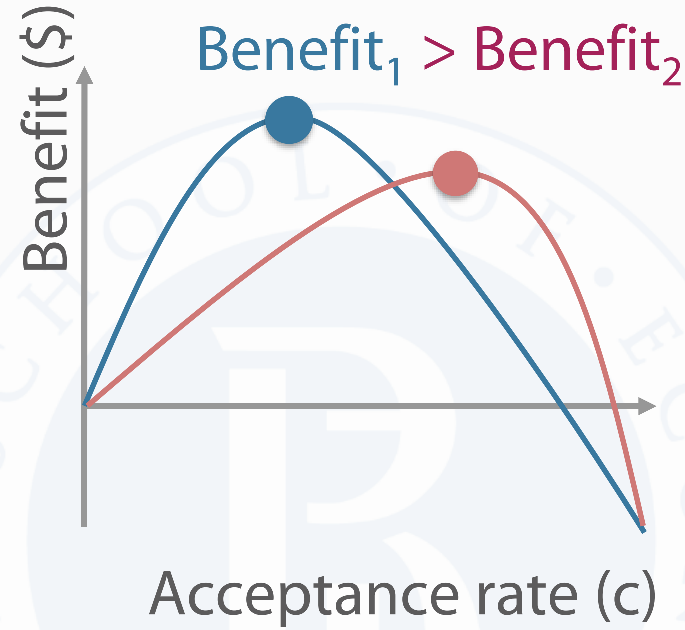


Same AUC but different benefits

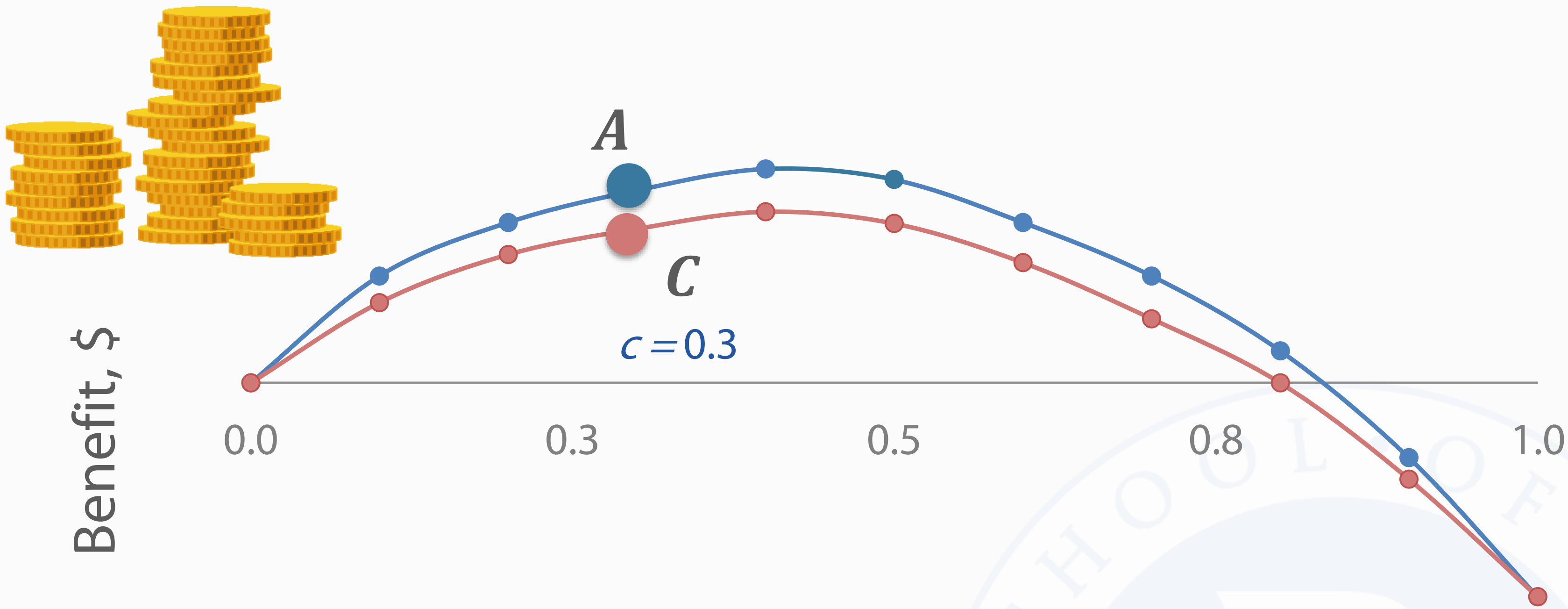
ROC curves as well as benefit curves may not majorize each other



However,



Benefit curve. Model decay



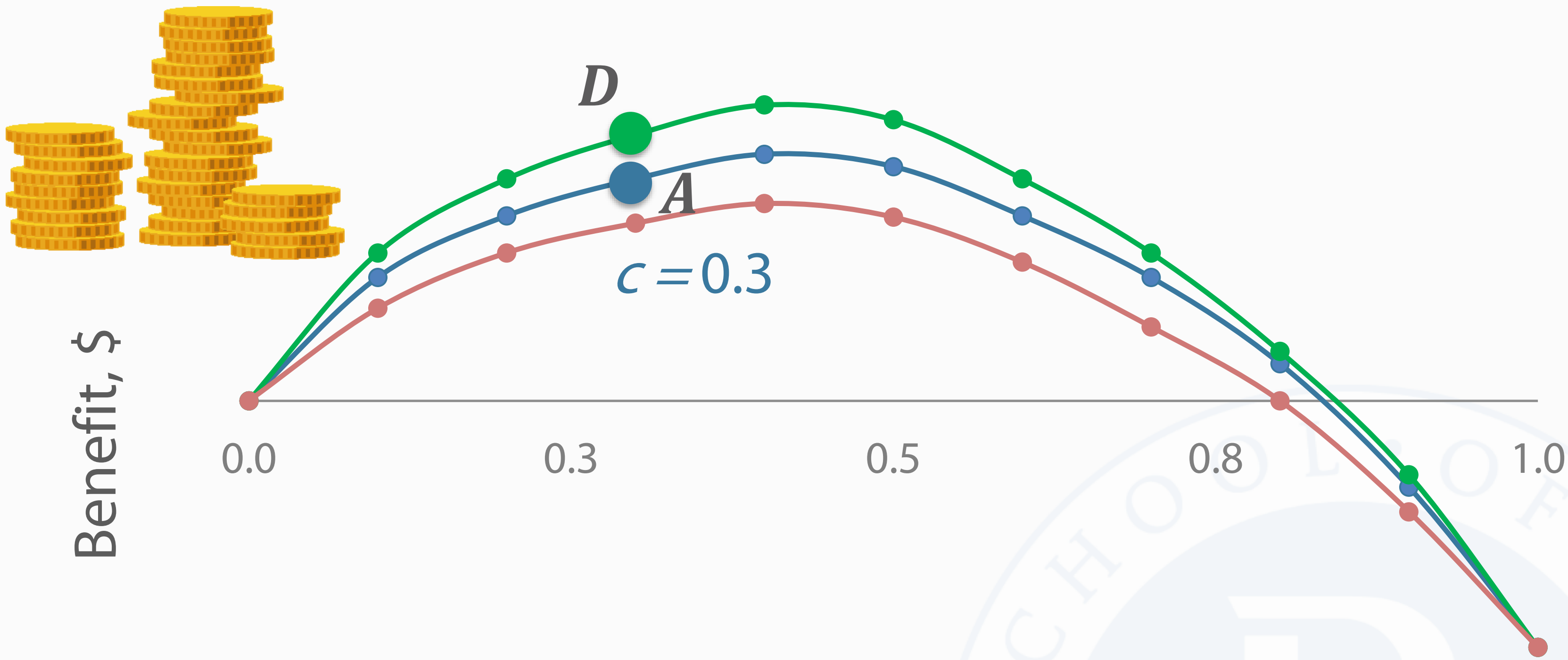
Model decay:

$$Benefit(\boldsymbol{A}) - Benefit(\boldsymbol{C})$$

– Model, AUC = 80%

– Model, w decr. AUC = 65%

Benefit curve. Better feasible model



Benefit from better feasible model:

$$Benefit(\boldsymbol{D}) - Benefit(\boldsymbol{A})$$

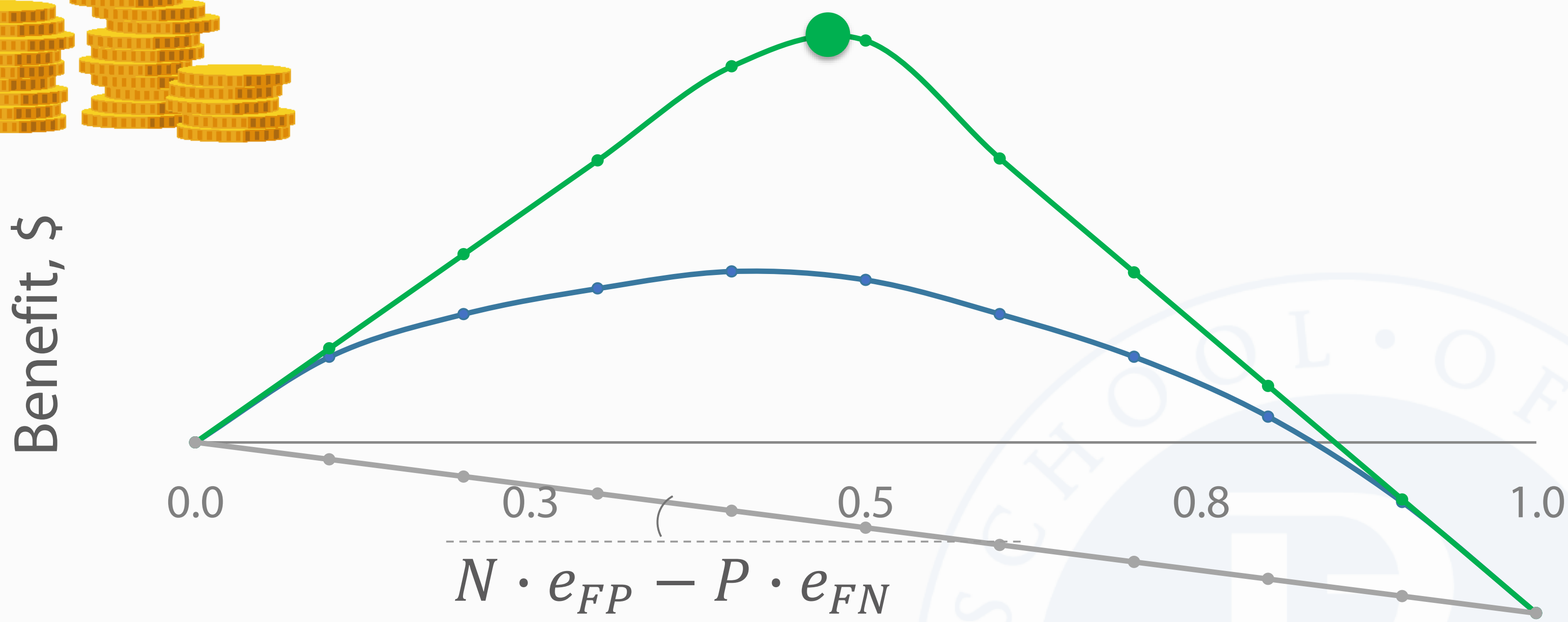
– Model, AUC = 80%

– Model, w better AUC = 90%

– Model, w decr. AUC = 65%

Benefit curve. Random and ideal model

$$Benefit_{max} = N \cdot e_{FP}$$



– Model, AUC = 80%

– Ideal, AUC = 100%

– Random, AUC = 65%

Wrap-up

1. Benefit curves are a powerful tool to assess expected financial result, they are dual to ROC curves

Wrap-up

1. Benefit curves are a powerful tool to assess expected financial result, they are dual to ROC curves
2. In general, the higher model quality metrics are, the higher is financial result

Wrap-up

1. Benefit curves are a powerful tool to assess expected financial result, they are dual to ROC curves

2. In general, the higher model quality metrics are, the higher is financial result

3. However, models with same AUCs can have different financial results, or even have opposite relation between financial result

Functional threshold decisions

Functional threshold decisions

- Instead of: "$Prob > a$" rule we make decision based on: if "$f(Prob) > a$" then some action is undertaken

Functional threshold decisions

- Instead of: "$Prob > a$" rule we make decision based on: if "$f(Prob) > a$" then some action is undertaken

- It can be useful when individual objects (clients) deliver different value for us

Functional threshold decisions

Resume to loan application problem:

$X_1 \dots X_k$	$Prob$	$margin_i$	lgd_i
34	0.94	100	380
20	0.25	140	500
21	0.16	90	280
50	0.86	85	360
41	0.51	250	1000
25	0.33	120	260
19	0.27	150	340
82	0.12	170	720

Functional threshold decisions

One can use following functional rule:

$if\ (1 - Prob) \cdot margin_i - Prob \cdot lgd_i \geq a$ then accept a loan, reject otherwise

Naturally, a controls for benefit appetite - required return

i.e. reject if:

$$\boxed{Prob > \frac{margin_i - a}{margin_i + lgd_i},} \qquad a \in [-lgd_i, margin_i\,]$$

Thus, functional threshold is equivalent to an individual simple threshold for each object (client)

Functional threshold decisions

Resume to loan application problem:

$X_1 \dots X_k$	$Prob$	$margin_i$	lgd_i	$f(Prob)$
34	0.94	100	380	-351.2
20	0.25	140	500	-20.0
21	0.16	90	280	30.8
50	0.86	85	360	-297.7
41	0.51	250	1000	-387.5
25	0.33	120	260	-5.4
19	0.27	150	340	17.7
82	0.12	170	720	63.2

Functional threshold decisions

Resume to loan application problem:

$X_1 \dots X_k$	$Prob$	$margin_i$	lgd_i	$f(Prob)$
82	0.12	170	720	63.2
21	0.16	90	280	30.8
19	**0.27**	150	340	17.7
25	0.33	120	260	-5.4
20	**0.25**	140	500	-20
50	0.86	85	360	-297.7
34	0.94	100	380	-351.2
41	0.51	250	1000	-387.5

Wrap-up

1. If the FP and FN error costs are different for every object (client), one should apply functional threshold decisions or, equally, individual thresholds for probability of event on client level

Increment based threshold decisions

Increment based threshold decisions

Again, consider a binary classification model $X \longrightarrow Prob$.
But now, our decision has influence on Prob.

Example: response, collection, churn models

$X_1 \dots X_k$	$Pr(Y = 1 \vert \boldsymbol{u_i} = 0)$	$Pr(Y = 1 \vert \boldsymbol{u_i} = 1)$
34	0.94	0.96
20	0.25	0.87
21	0.16	0.8
50	0.86	0.65
41	0.51	0.43
25	0.33	0.35
19	0.27	0.32
82	0.12	0.13

Denote $\boldsymbol{u_i} = 1$, if we apply an action to *i*-th object (client) and $u_i = 0$, otherwise

Increment based threshold decisions

Simple threshold decision rule is not appropriate anymore.

If we sort by Prob, we likely undertake action for clients who would do the same thing on their own

$X_1 \dots X_k$	$Pr(Y = 1 \mid \boldsymbol{u_i} = 0)$	$Pr(Y = 1 \mid \boldsymbol{u_i} = 1)$
34	**0.94**	**0.96**
50	**0.86**	**0.87**
41	0.51	0.8
20	0.25	0.65
21	0.16	0.43
25	0.33	0.35
19	0.27	0.32
82	0.12	0.13

Small increment

Increment based threshold decisions

So, let us set the problem

$$\sum_{i=1}^{N} Pr\,(Y_i = 1|u_i) \cdot margin - u_i \cdot cost \underset{u_i=\{0,1\}}{\rightarrow} max$$

$i = 1, \dots N$ – a set of objects (clients)

Y_i – target event (e.g, cross-sell / retention / reimbursement)

$margin$ – benefit from one target event

u_i – whether we undertake action (communication with client)

$cost$ – action costs (communication expenses)

Increment based threshold decisions

Denote $Pr_1 = Pr(Y_i = 1|u_i = 1)$ and
$$Pr_0 = Pr(Y_i = 1|u_i = 0)$$

Increment based threshold decisions

Denote $Pr_1 = Pr(Y_i = 1|u_i = 1)$ and
$Pr_0 = Pr(Y_i = 1|u_i = 0)$

On client level $u_i = 1$ should be chosen if it leads to higher financial result, i.e.:

$$Pr_1 \cdot margin - cost > Pr_0 \cdot margin$$

Increment based threshold decisions

Denote $Pr_1 = Pr(Y_i = 1|u_i = 1)$ and

$$Pr_0 = Pr(Y_i = 1|u_i = 0)$$

On client level $u_i = 1$ should be chosen if it leads to higher financial result, i.e.:

$$Pr_1 \cdot margin - cost > Pr_0 \cdot margin$$

So,

$$\boxed{Pr_1 - Pr_0 > \frac{cost}{margin}}$$

Prob increment due to our action

Increment based threshold decisions

Now, we see, if one applies simple decision rule, then top clients tend to have small increment ΔPr.

Next step is to resort our clients based on ΔPr

$X_1 \dots X_k$	Pr_0	Pr_1	ΔPr
34	**0.94**	**0.96**	**0.02**
50	**0.86**	**0.87**	**0.01**
41	0.51	0.8	0.29
20	0.25	0.65	0.4
21	0.16	0.43	0.27
25	0.33	0.35	0.02
19	0.27	0.32	0.05
82	0.12	0.13	0.01

Small increment

Increment based threshold decisions

So, we apply communication only to top clients with large increase in probability of target event

$X_1 \dots X_k$	Pr_0	Pr_1	ΔPr
20	0.25	0.65	0.4
41	0.51	0.8	0.29
21	0.16	0.43	0.27
19	0.27	0.32	0.05
34	0.94	0.96	0.02
25	0.33	0.35	0.02
50	0.86	0.87	0.01
82	0.12	0.13	0.01

Large increment

$$a_{min} = \frac{cost}{margin}$$

Wrap-up

1. When our decision has influence on probability of target event one should use incremental threshold decision strategy instead of simple thresholds

Wrap-up

1. When our decision has influence on probability of target event one should use incremental threshold decision strategy instead of simple thresholds

2. It allows one to maximize probability of desired target event and, therefore, bring more benefit at less action cost