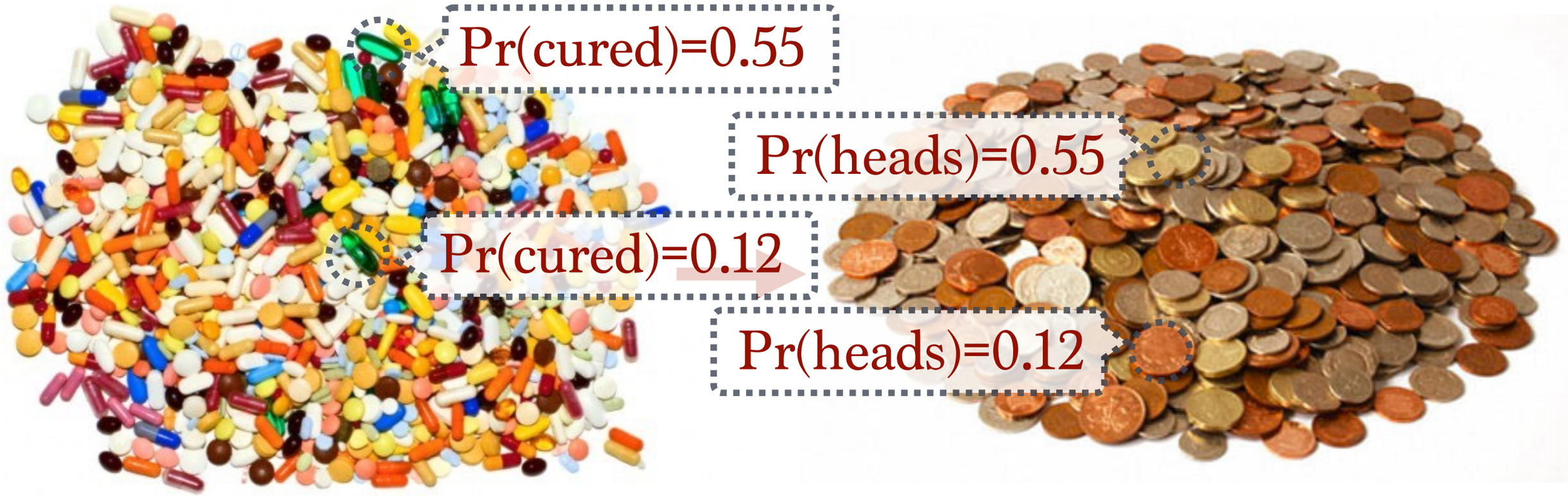
ON MULTI-ARMED BANDITS THEORY AND APPLICATIONS

Maryam Aziz

Advisor: Javed Aslam

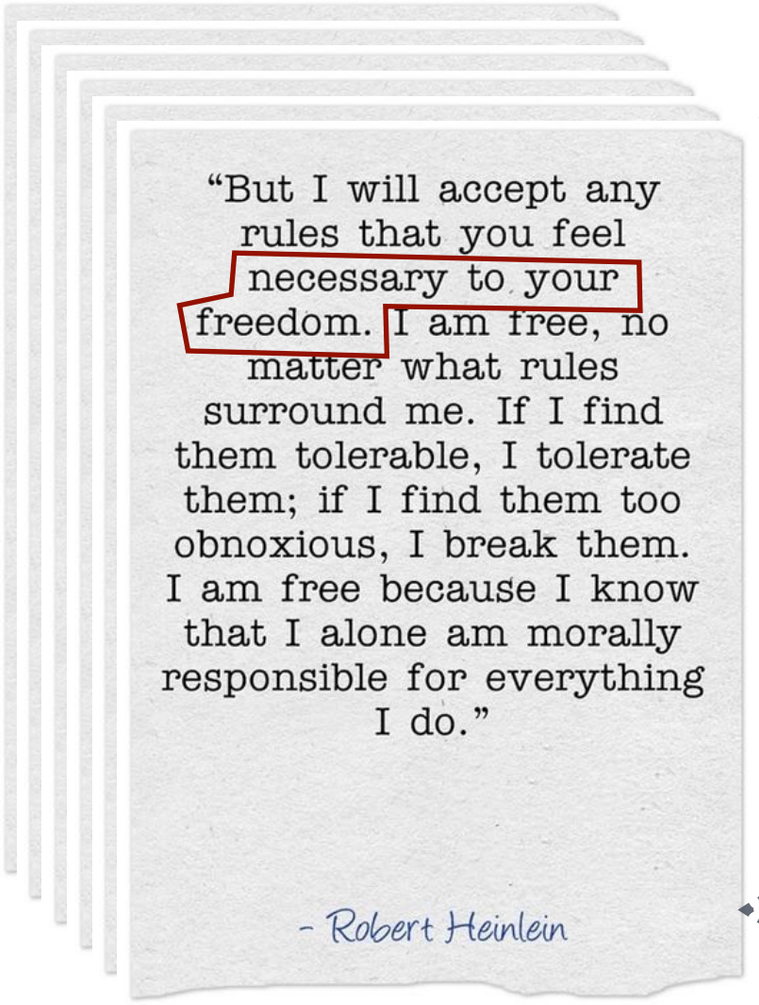
Committee: Emilie Kaufmann, Byron Wallace, Jonathan Ullman



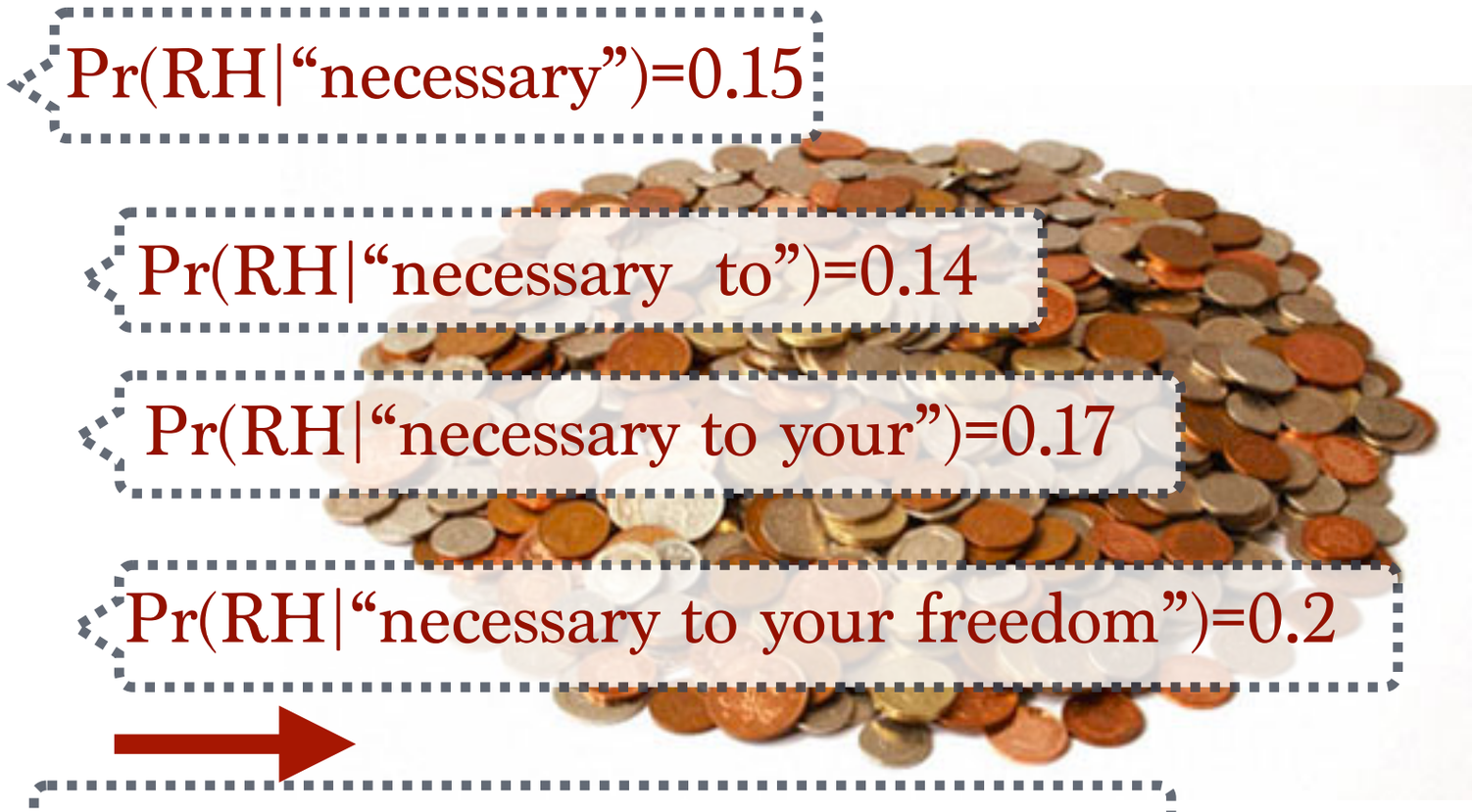
DRUG DISCOVERY

Given a small number (100) of patients and a large number (1000) of drugs, can I find an effective drug?

Text Feature Reservoir



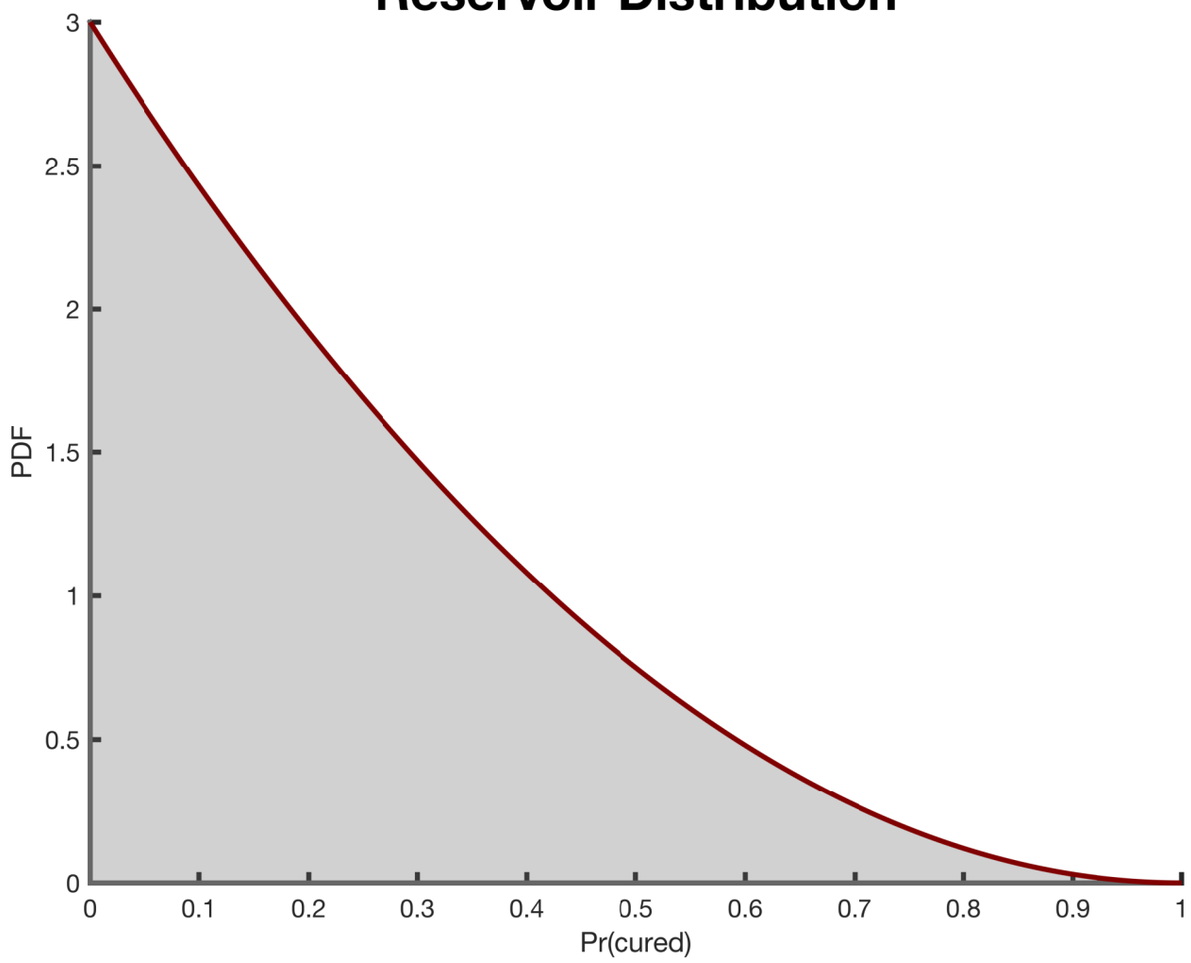
Coin Reservoir



Pr(RH|"necessary", "freedom")=0.28

TEXT CLASSIFICATION

Given a small of trials and a large number of "text features" (used to build classifiers), can I find a feature that accurately classifies whether a quote (document) belongs to Robert Heinlein?



Coin Reservoir

Given a finite number of tests and an infinite number of coins, can I find a high bias coin?

MULTI-ARMED BANDIT SETTING

VARIANTS OF THE MULTI-ARMED BANDIT PROBLEM

- cumulative regret vs. simple regret
- fixed budget vs. fixed confidence
- finite number of arms vs. infinite number of arms
- lower bounds vs. algorithms/upper bounds

E.g. Thompson, 1933; Even-Dar et al., 2006; Wang et al., 2009; Bonald and Prouti`ere, 2013; David and Shimkin, 2014; Carpentier and Valko, 2015; Kaufmann and Kalyanakrishnan, 2013

VARIANTS OF THE MULTI-ARMED BANDIT PROBLEM

- cumulative regret vs. simple regret
- fixed budget vs. fixed confidence
- finite number of arms vs. infinite number of arms
- lower bounds vs. algorithms/upper bounds

Theory

Application

E.g. Thompson, 1933; Even-Dar et al., 2006; Wang et al., 2009; Bonald and Prouti`ere, 2013; David and Shimkin, 2014; Carpentier and Valko, 2015; Kaufmann and Kalyanakrishnan, 2013

SUMMARY

Theory:

Pure Exploration for the Infinitely Armed Bandit Models		
	Lower Bound	**Algorithm/Upper Bound**
Fixed Confidence	Completed Work	Completed Work
Fixed Budget	Completed Work	Completed Work

Applications:

1) Dose Finding,
2) Boosted Decision Tree Training

ROAD MAP

Theory

Pure exploration in infinitely- armed bandit models with fixed-confidence. Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. ALT (Algorithmic Learning Theory) 18.

Pure-Exploration for Infinite-Armed Bandits with General Arm Reservoirs.
Maryam Aziz, Kevin Jamieson and Javed Aslam. Under review.

Application

On Multi-Armed Bandit Designs for Phase I Clinical Trials.
Maryam Aziz, Emilie Kaufmann, Marie-Karelle Riviere. In preprint.

Adaptively Pruning Features for Boosted Decision Trees.
Maryam Aziz, Jesse Anderton, Javed Aslam. In preprint.

ROAD MAP

Theory

Pure exploration in infinitely- armed bandit models with fixed-confidence.
Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. ALT (Algorithmic Learning Theory) 18.

Pure-Exploration for Infinite-Armed Bandits with General Arm Reservoirs.
Maryam Aziz, Kevin Jamieson and Javed Aslam. Under review.

Application

On Multi-Armed Bandit Designs for Phase I Clinical Trials.
Maryam Aziz, Emilie Kaufmann, Marie-Karelle Riviere. In preprint.

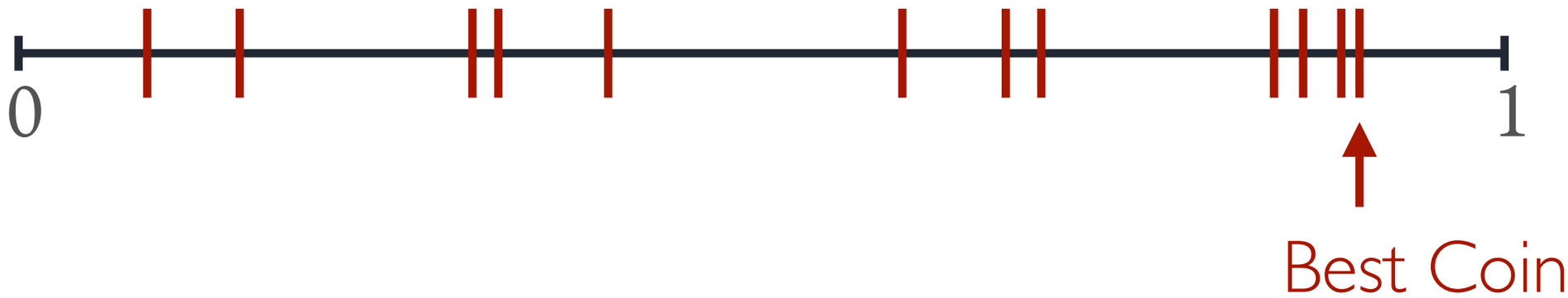
Adaptively Pruning Features for Boosted Decision Trees.
Maryam Aziz, Jesse Anderton, Javed Aslam. In preprint.

FIXED BUDGET SETTING

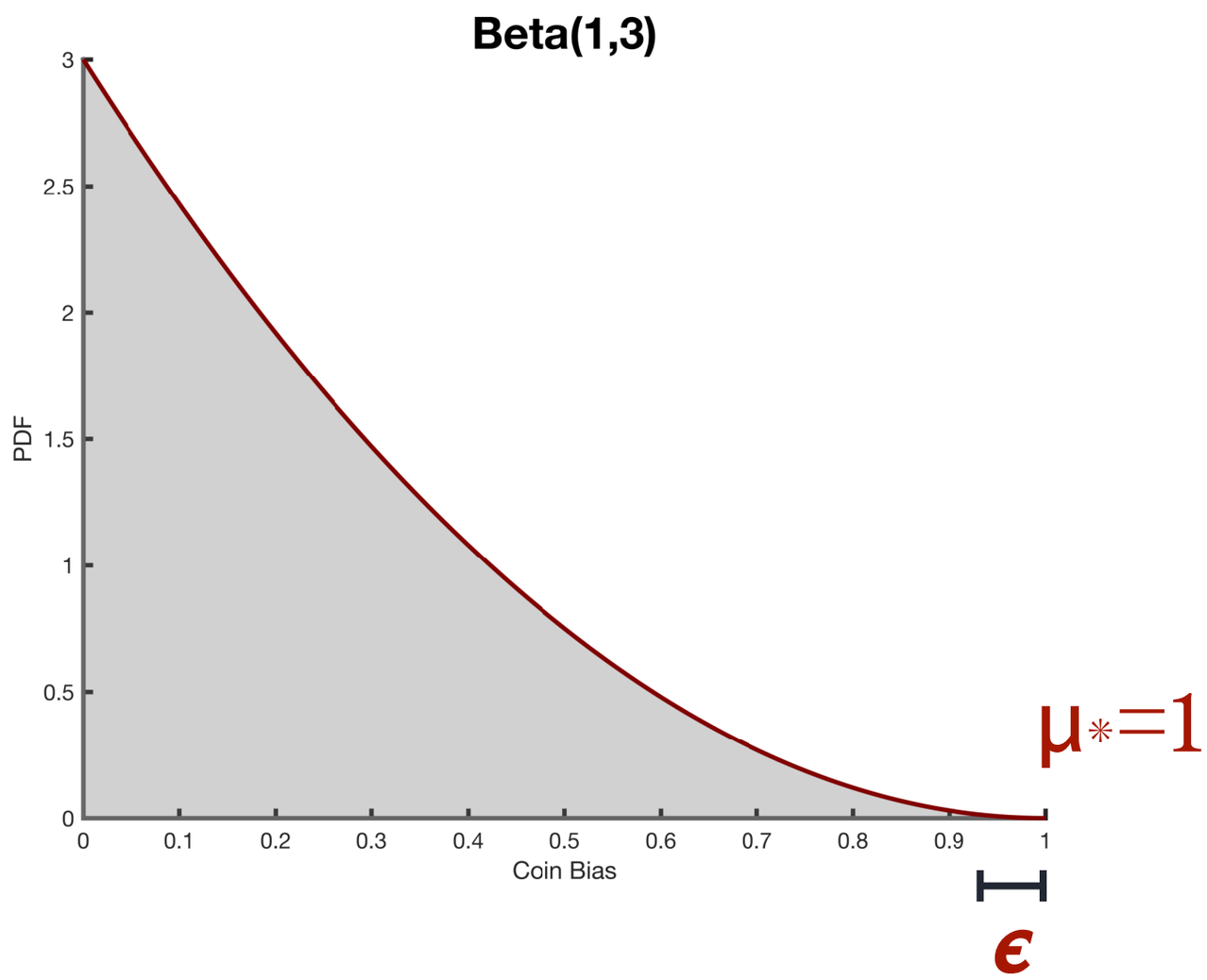
Pure Exploration for the Infinitely Armed Bandit Models		
	Lower Bound	**Algorithm/Upper Bound**
Fixed Confidence	Completed Work	Completed Work
Fixed Budget	Completed Work	Completed Work

Pure-Exploration for Infinite-Armed Bandits with General Arm Reservoirs.
Maryam Aziz, Kevin Jamieson and Javed Aslam. Under review.

TOP COIN (SIMPLE REGRET) IN THE FINITE CASE



TOP COIN (SIMPLE REGRET) IN THE INFINITE CASE



We do not require any assumption on the reservoir distribution that prior work relies on.

SIMPLE REGRET GOAL IN THE FIXED BUDGET SETTING

Given a small number (100) of patients and a large number (1000) of drugs, can I find an effective drug?

Given a fixed number of total coin flips T what's the smallest simple regret (ϵ) one can guarantee with probability $1 - \delta$?

SUCCESSIVE HALVING FOR FINITE # COIN

Total flips: 240

flips each 10 time, 80 flips

flips each 20 times, 80 flips

flips each 40 times, 80 flips

The coin (best) returned by the algorithm

EXPLORE-IDENTIFY DILEMMA

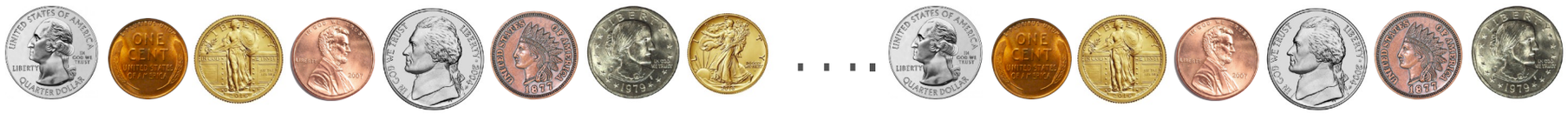
Total flips: 240. How to initialize SH?

2 coins, 120 flips/coin in 1st round

4 coins, 30 flips/coin in 1st round

8 coins, 10 flips/coin in 1st round

16 coins, ~3 flips/coin in 1st round



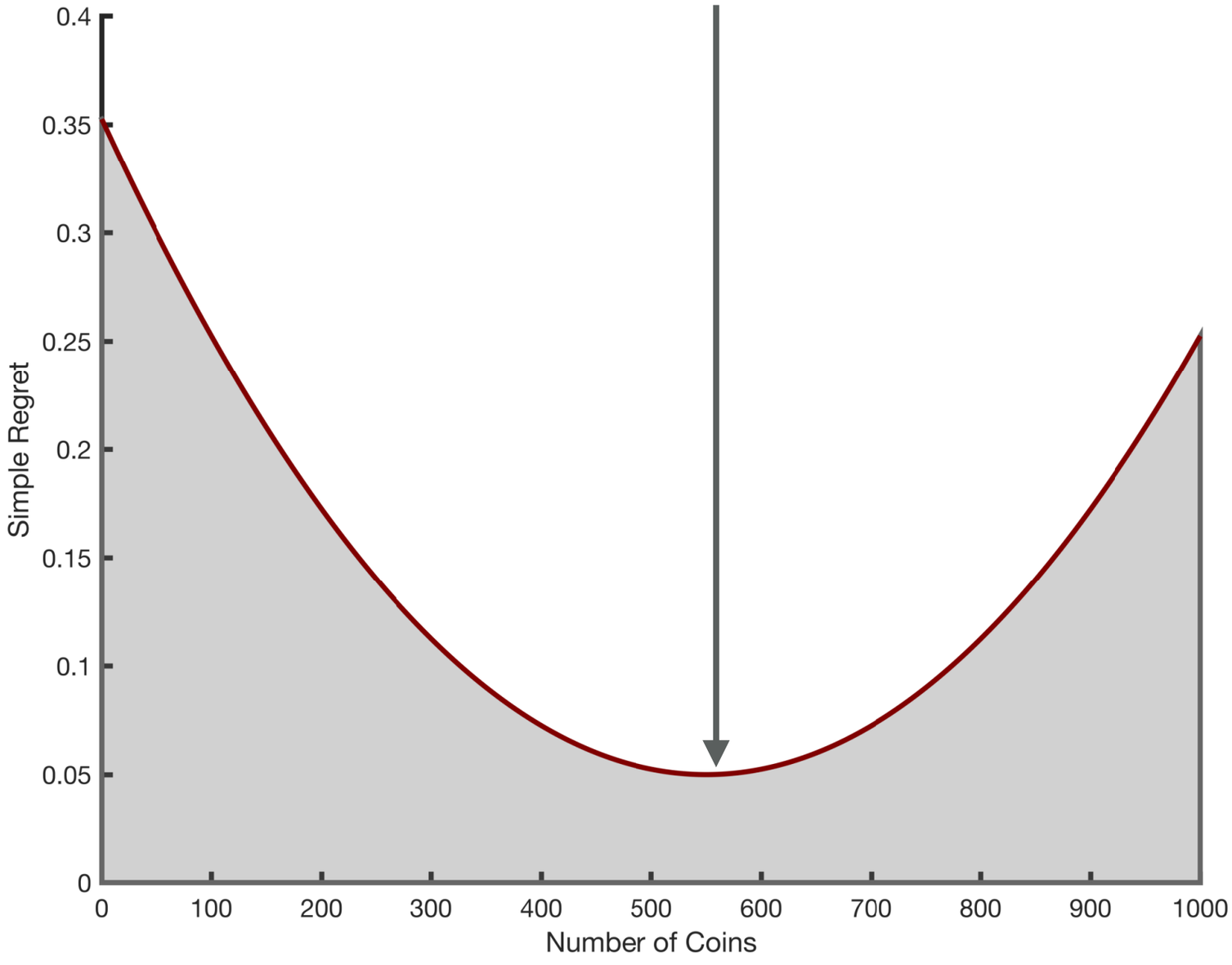
43 coins, 1 flip/coin in 1st round

HOW MANY COINS?

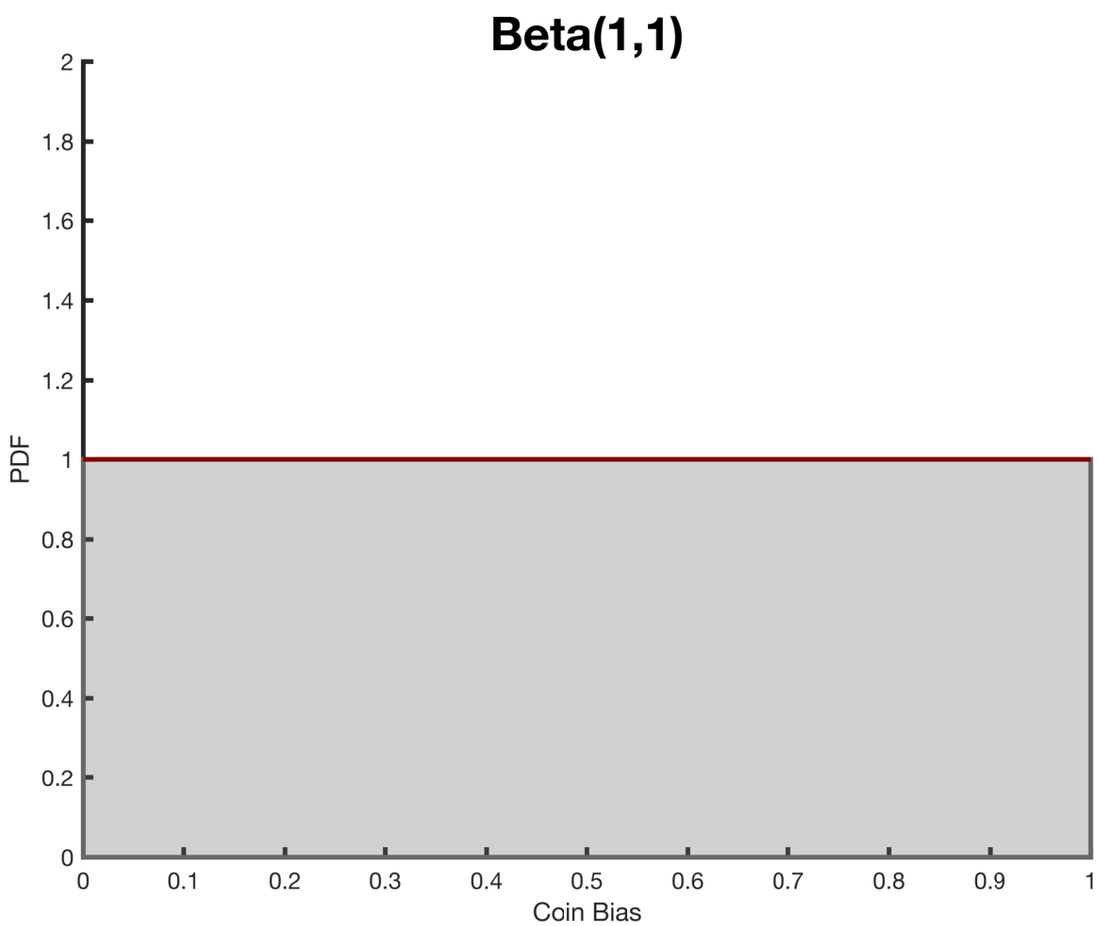
Identify Best Coin

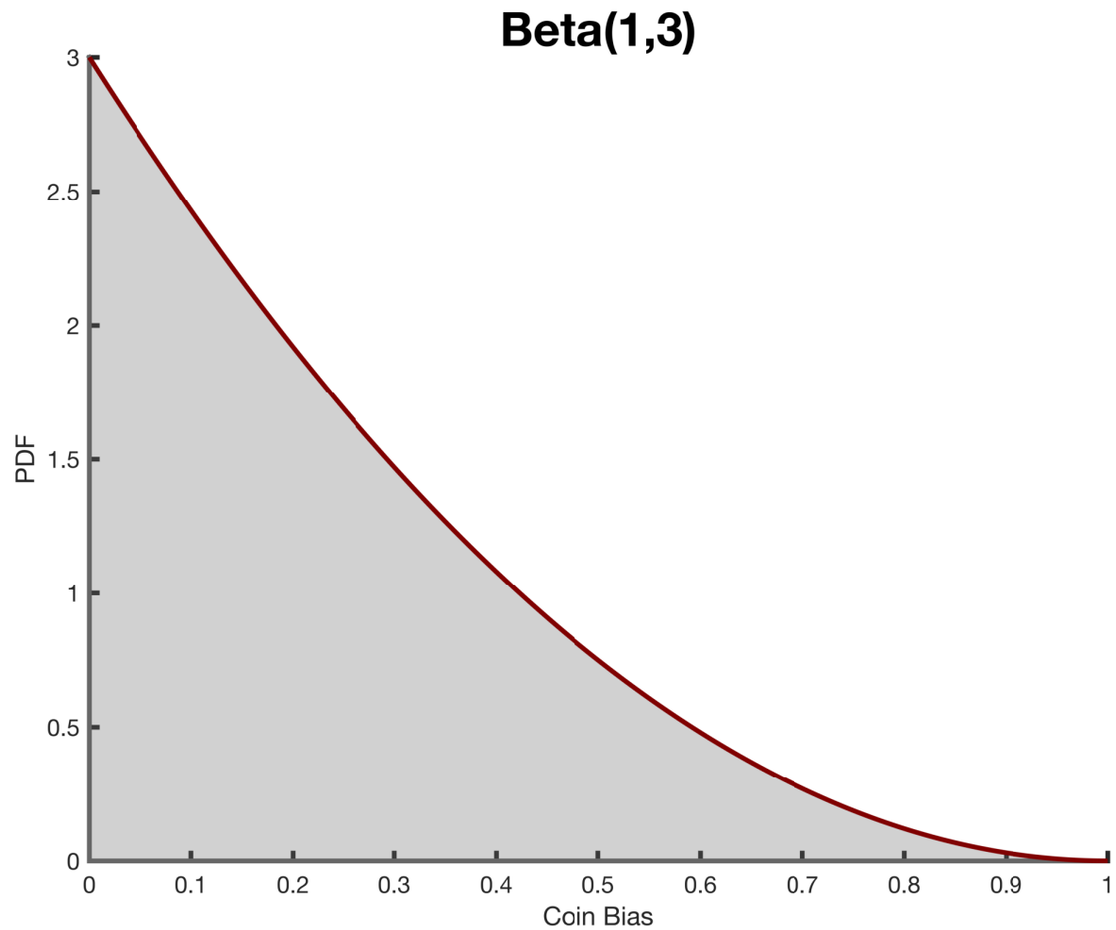
Sweet Spot

Explore Better Coins

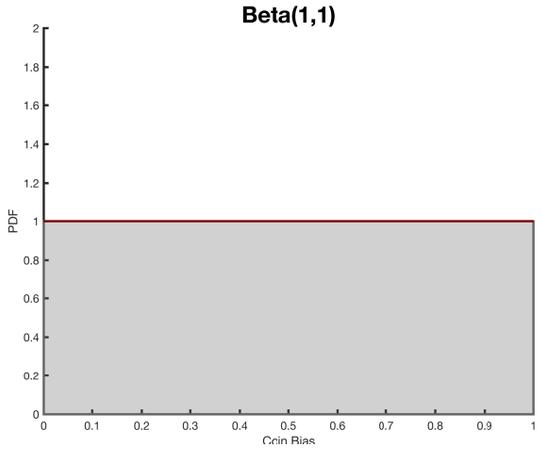


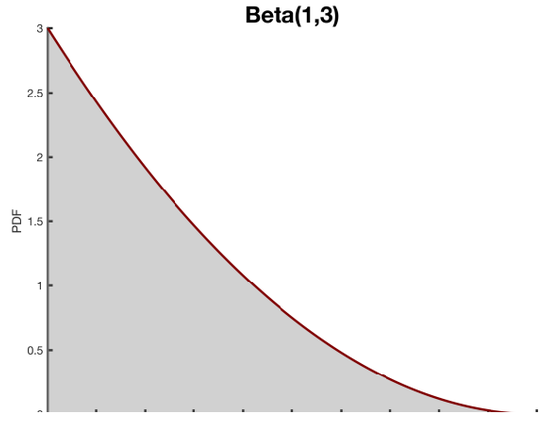
EXPLORE-IDENTIFY TRADEOFF

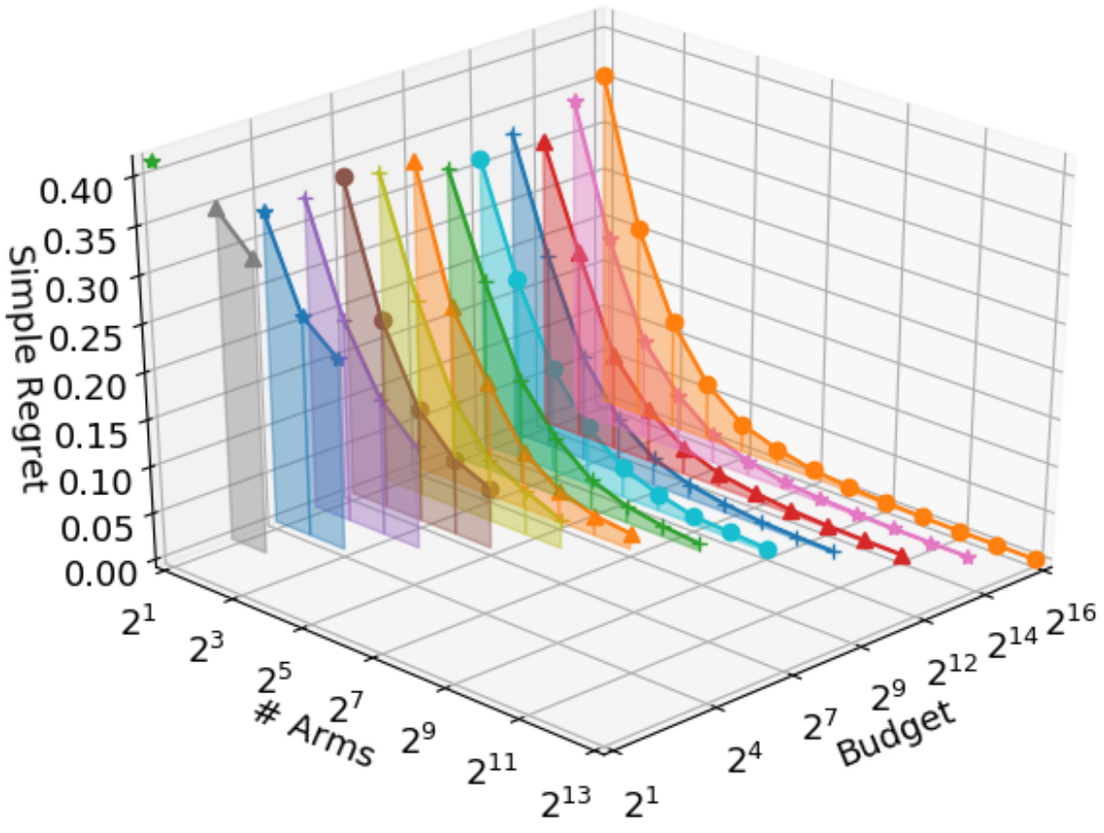


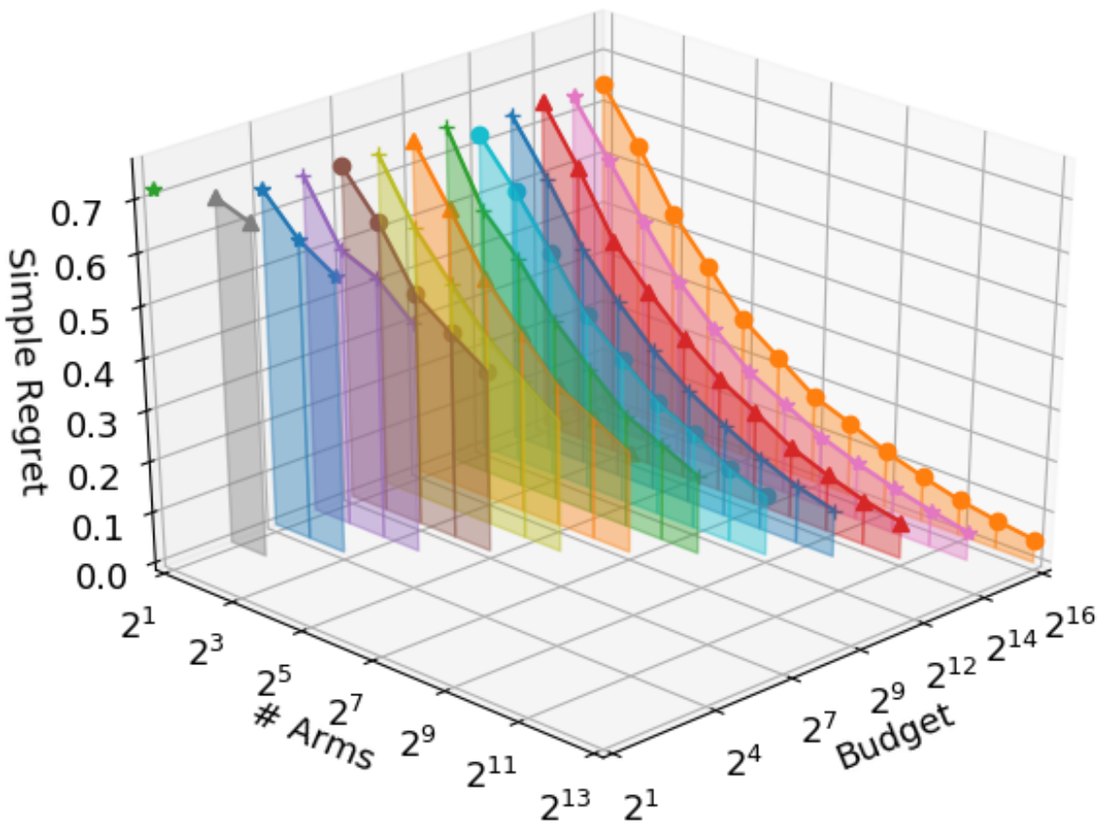


EXPLORE-IDENTIFY TRADEOFF



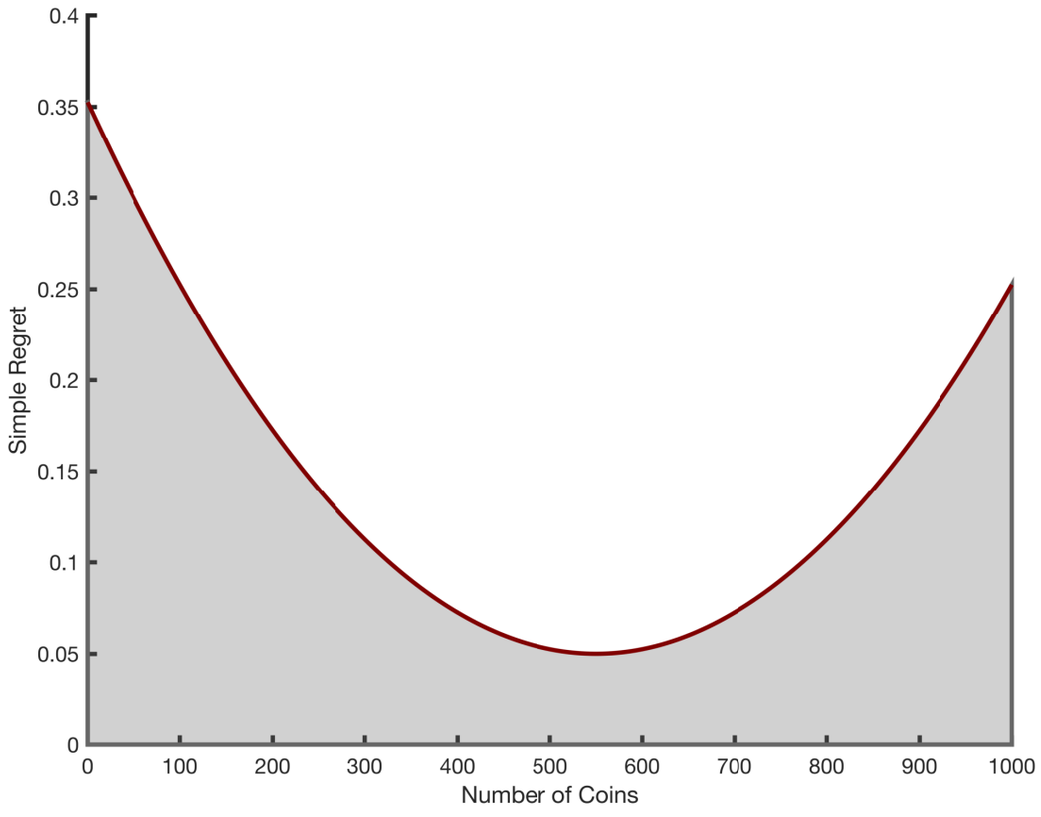




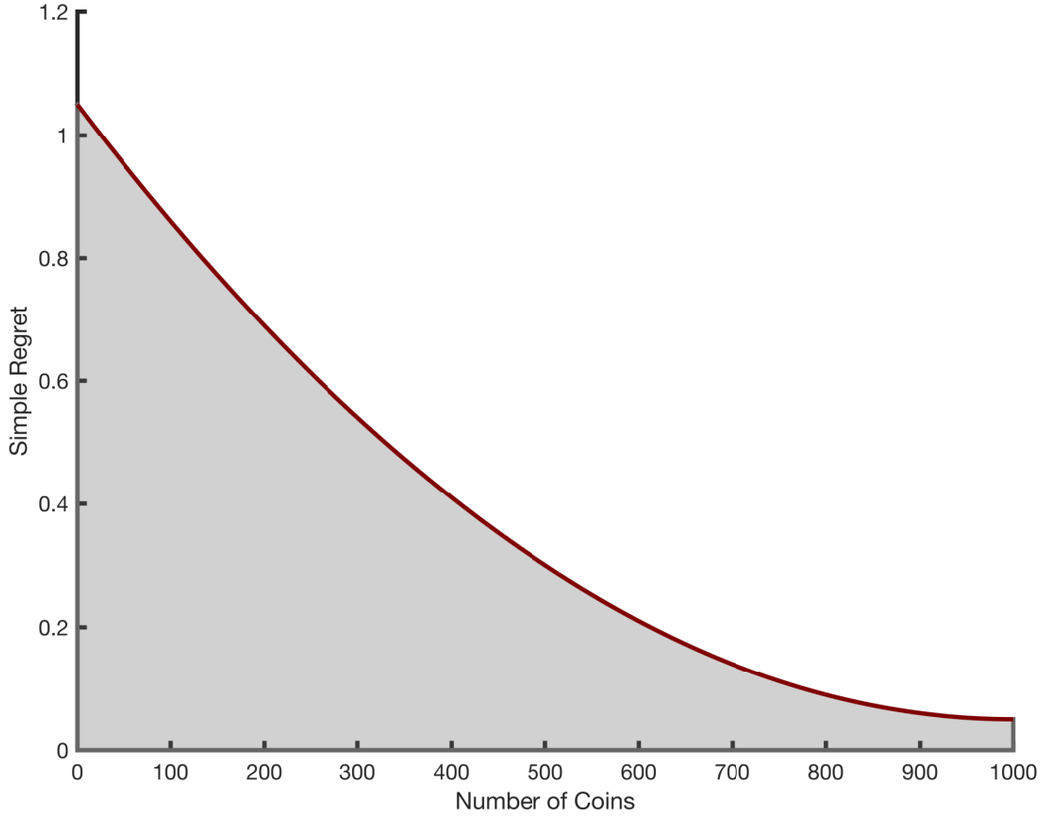


INTUITION VS. REALITY?

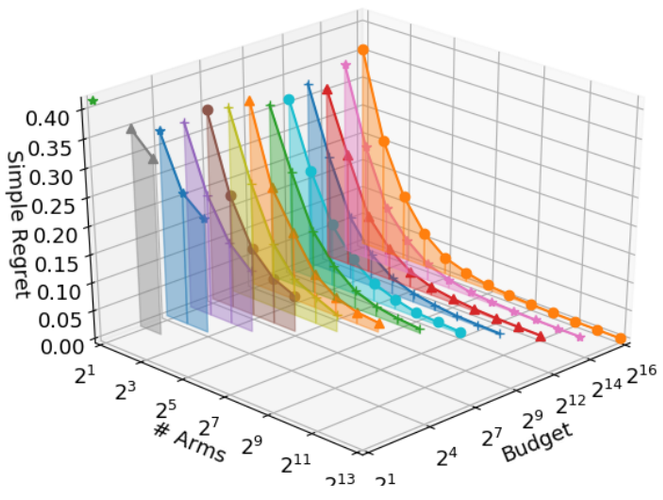
What We Expected

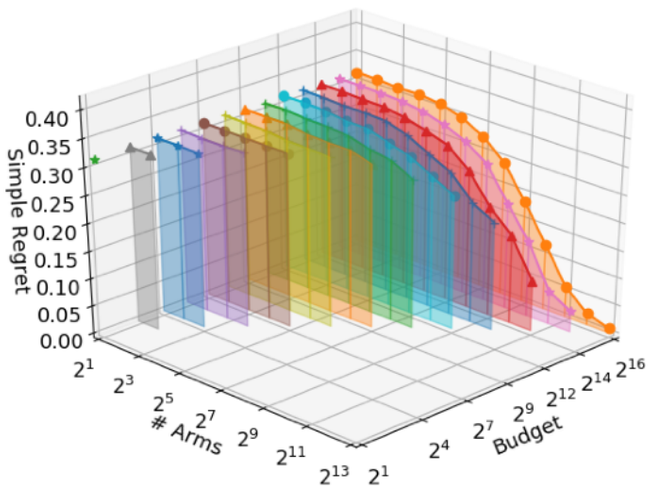


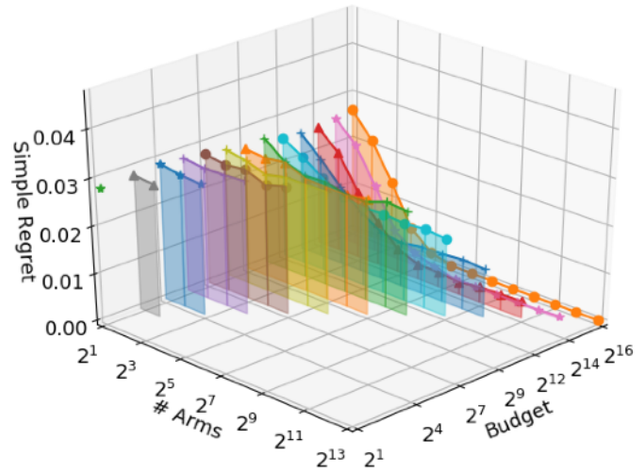
What We Got

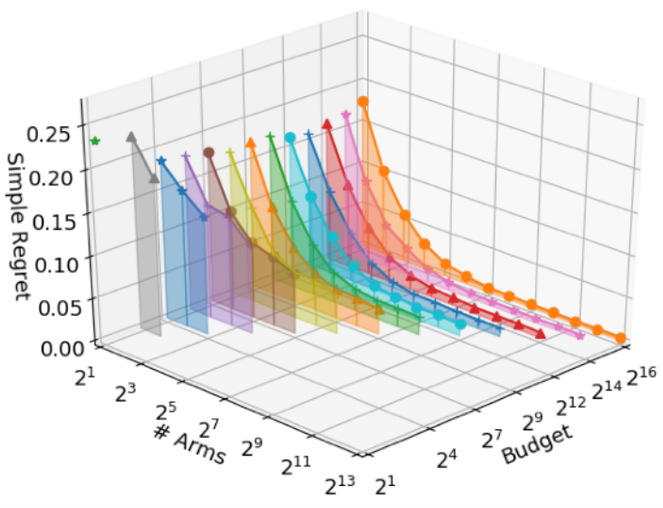


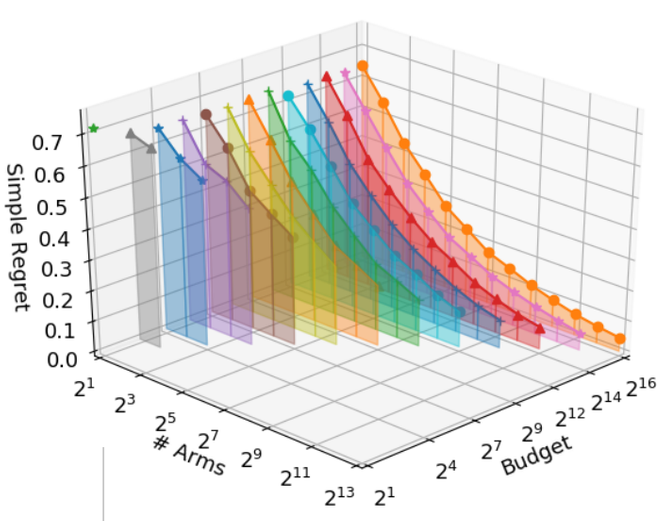
EXPLORE-IDENTIFY TRADEOFF

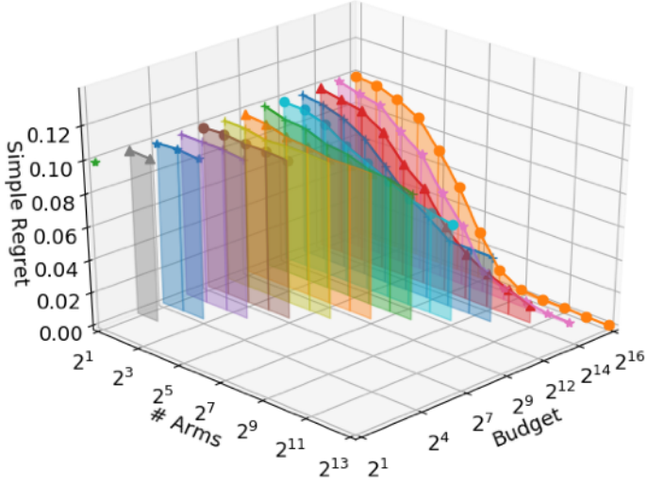


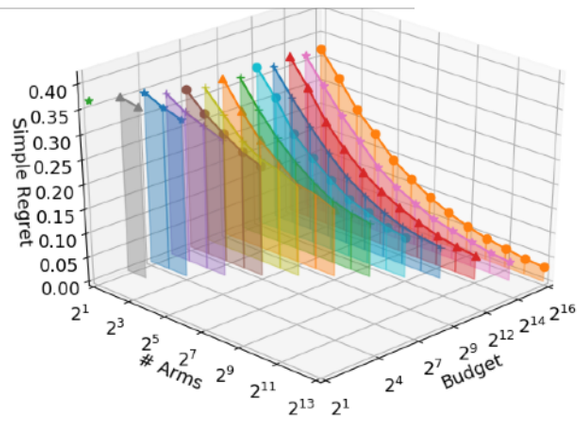


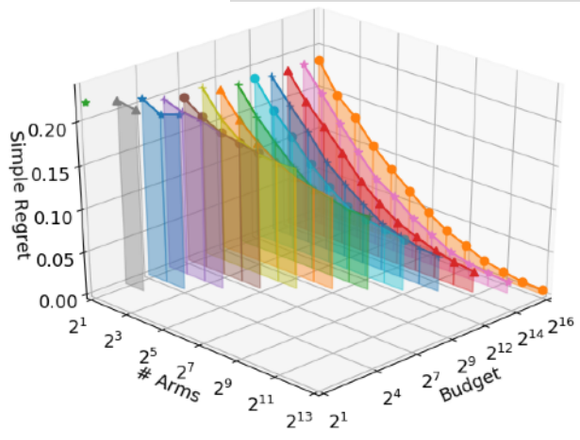












SUCCESSIVE HALVING FOR INFINITE
COIN (ISHA)

Total flips: 24

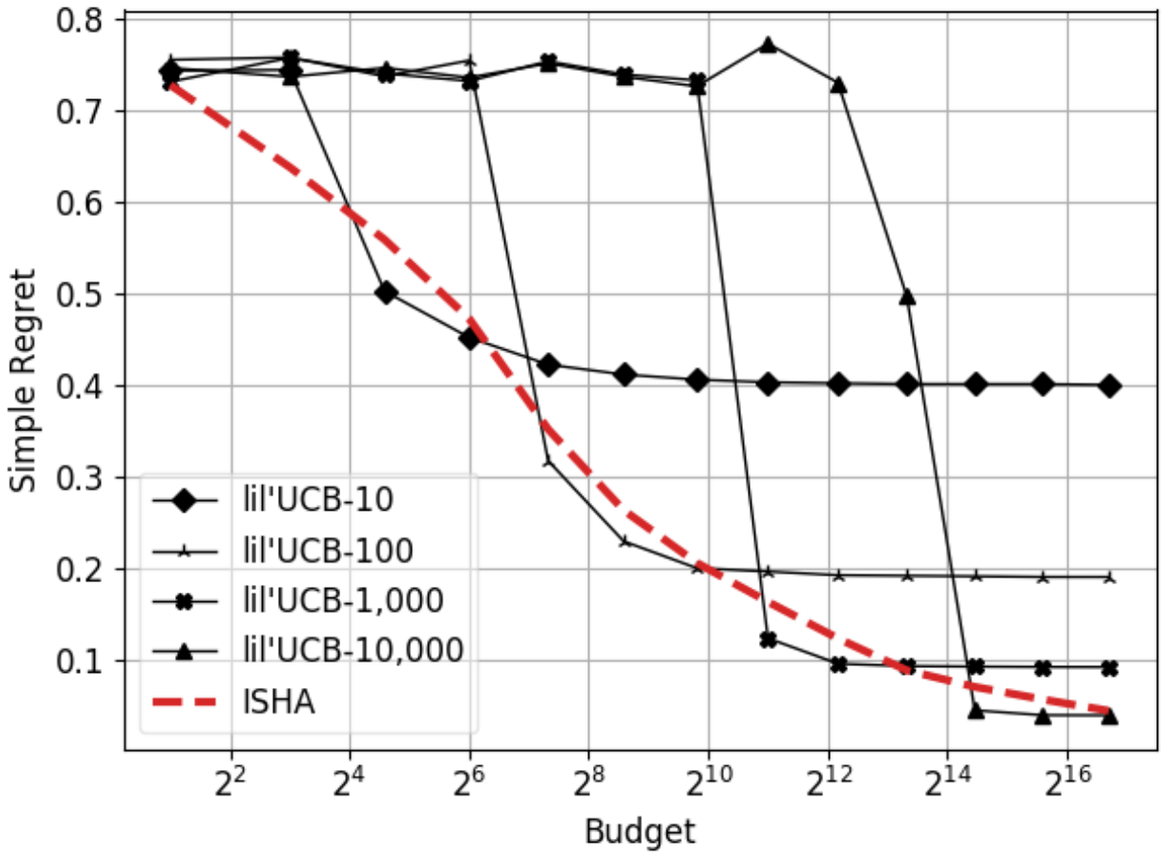
flips each 1 time, 8 flips

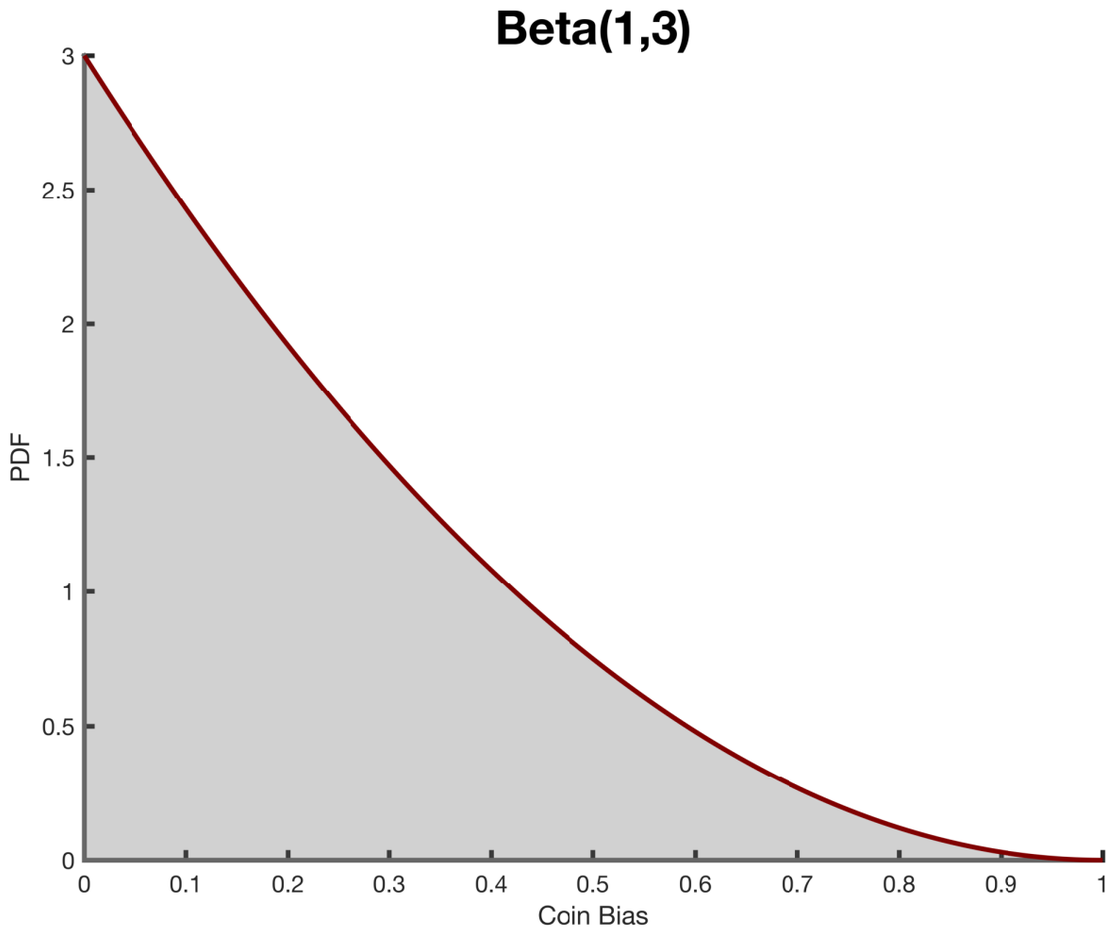
flips each 2 times, 8 flips

flips each 4 times, 8 flips

The coin (best) returned by the algorithm

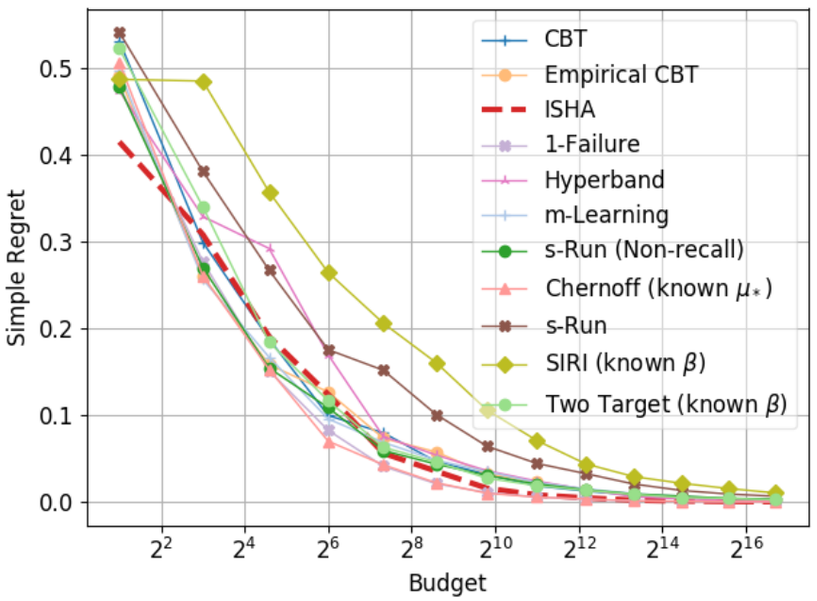
FIXED BUDGET ALGORITHM RESULTS FOR RESERVOIR BETA(1,3)

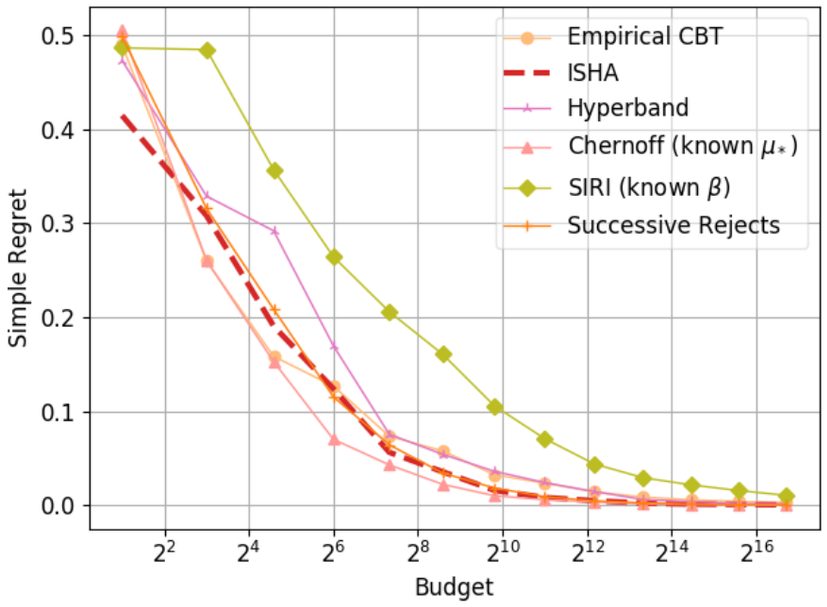


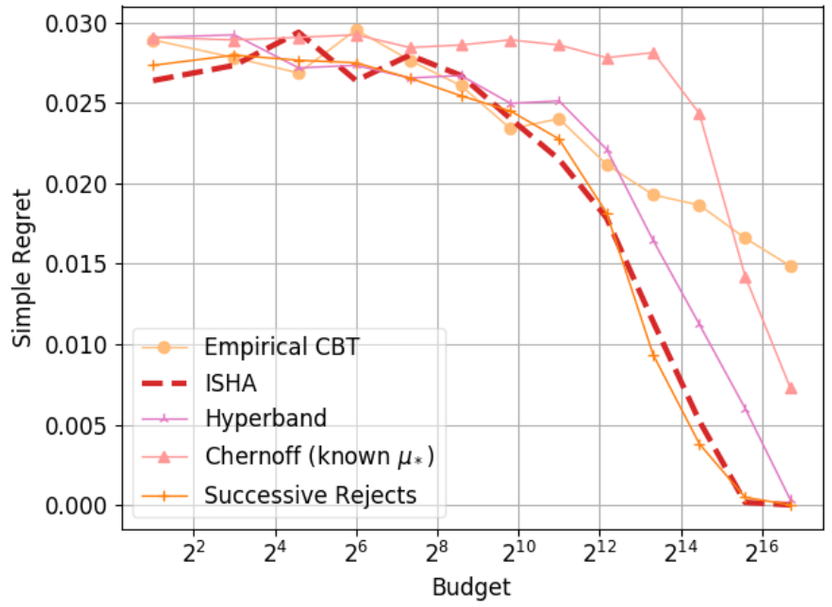


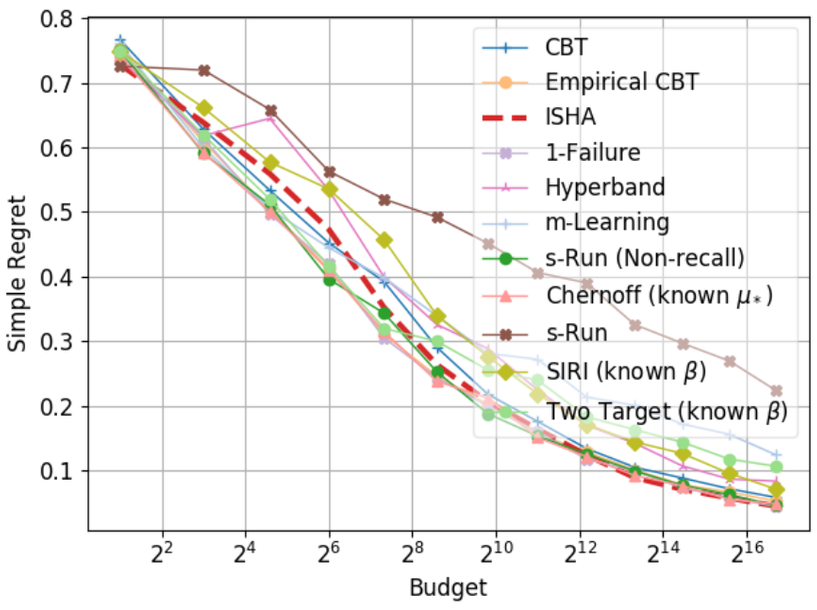
lil' ucb : An optimal exploration algorithm for multi-armed bandits. In COLT, 2014.
Kevin G. Jamieson, Matthew Malloy, Robert D. Nowak, and S´ebastien Bubeck.

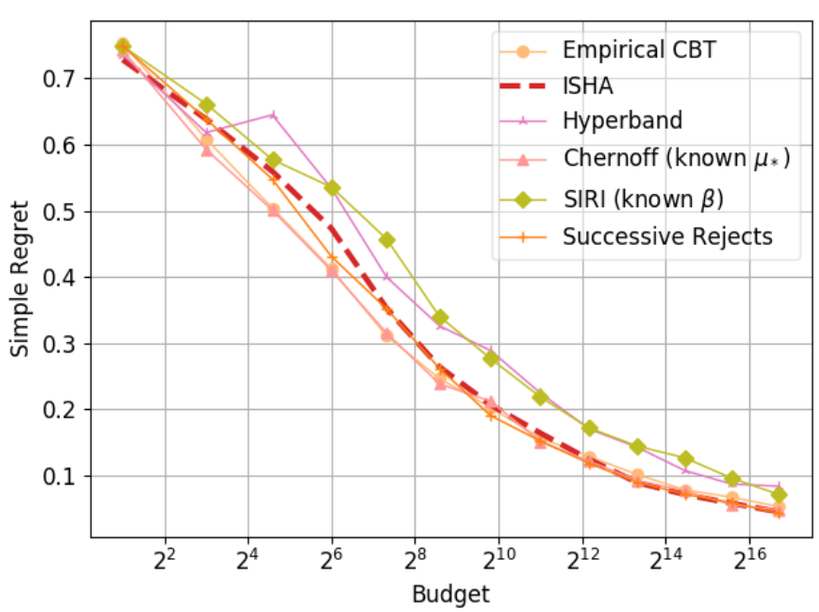
A SAMPLE SET OF RESULTS

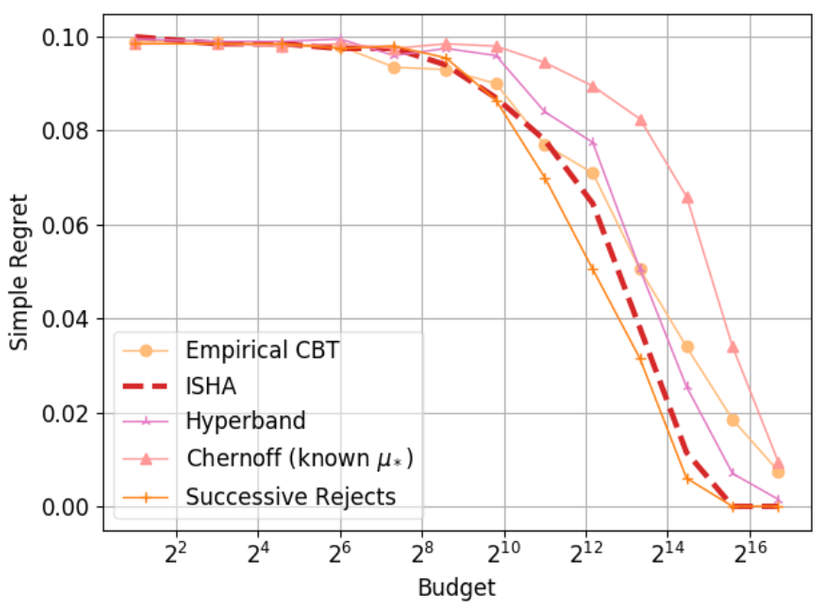




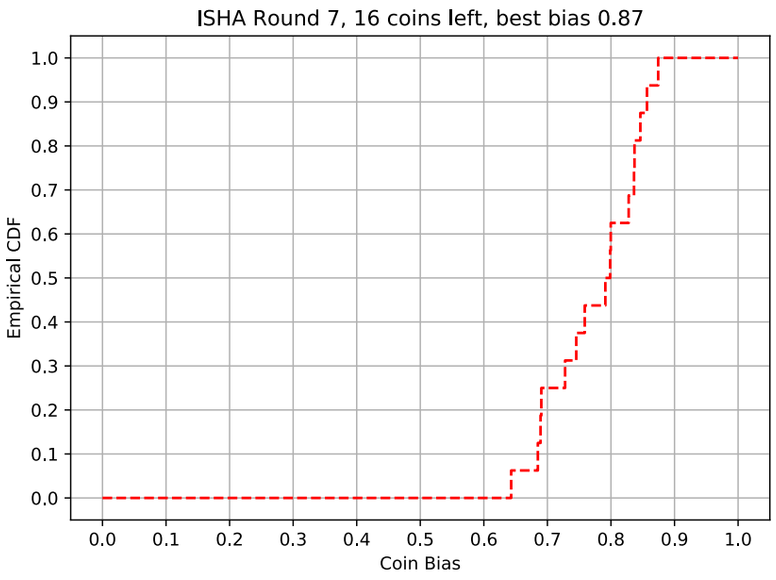


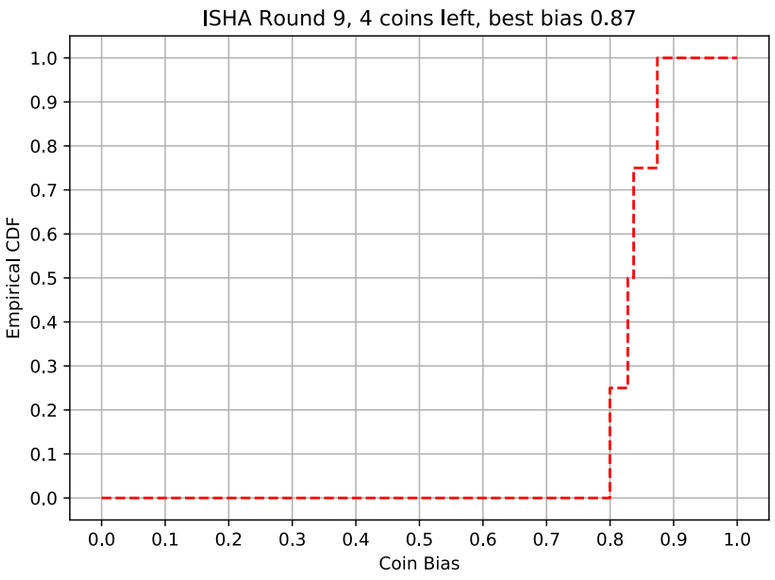


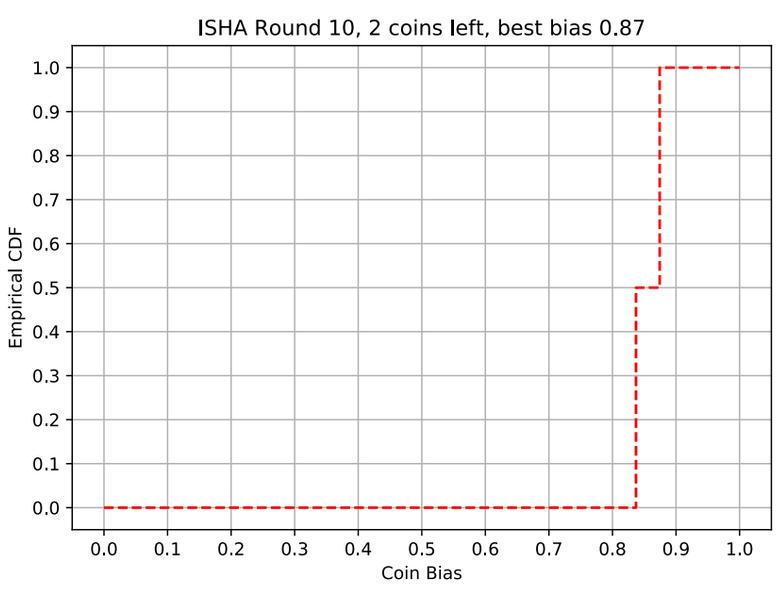




ISHA/UB PROOF KEY INSIGHTS







ISHA BOUNDS

Upper Bound: We return an ϵ-good arm with probability $1-\delta$ when $T = \lceil n \lg n \rceil$ and both:

$$n \geq \frac{1}{\nu(\mu_* + \epsilon)} \log\left(\frac{2\lg(n)}{\delta}\right)$$

$$n \geq 64R \log^2\left(\frac{4T}{\delta}\right)\left(\frac{1}{\epsilon^2} + \frac{1}{\nu(\mu_* + \epsilon)} \int_{x=\mu_*+\epsilon}^{\infty} \frac{1}{(x-\mu_*)^2} d\nu(x)\right)$$

Lower Bound: For any reservoir continuous around μ_* and (ϵ,δ)-correct non-recall bandit algorithm we have:

$$\mathbb{E}[\tau] \geq \frac{1}{\epsilon^2} \log\left(\frac{1}{\delta\nu(\mu_* + \epsilon)}\right) + \frac{1}{\nu(\mu_* + \epsilon)} \int_{x=\mu_*+\epsilon}^{\infty} \frac{1}{(x-\mu_*)^2} d\nu(x)$$

ROAD MAP

Theory

Pure exploration in infinitely- armed bandit models with fixed-confidence.
Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. ALT (Algorithmic Learning Theory) 18.

Pure-Exploration for Infinite-Armed Bandits with General Arm Reservoirs.
Maryam Aziz, Kevin Jamieson and Javed Aslam. Under review.

Application

On Multi-Armed Bandit Designs for Phase I Clinical Trials.
Maryam Aziz, Emilie Kaufmann, Marie-Karelle Riviere. In preprint.

Adaptively Pruning Features for Boosted Decision Trees.
Maryam Aziz, Jesse Anderton, Javed Aslam. In preprint.

DOSE FINDING SETUP

larger dose, more effective, Pr(toxicity)=0.55

smaller dose, less effective, Pr(toxicity)=0.3

How to find best dose while treating as many trial patients as possible with best dosage?

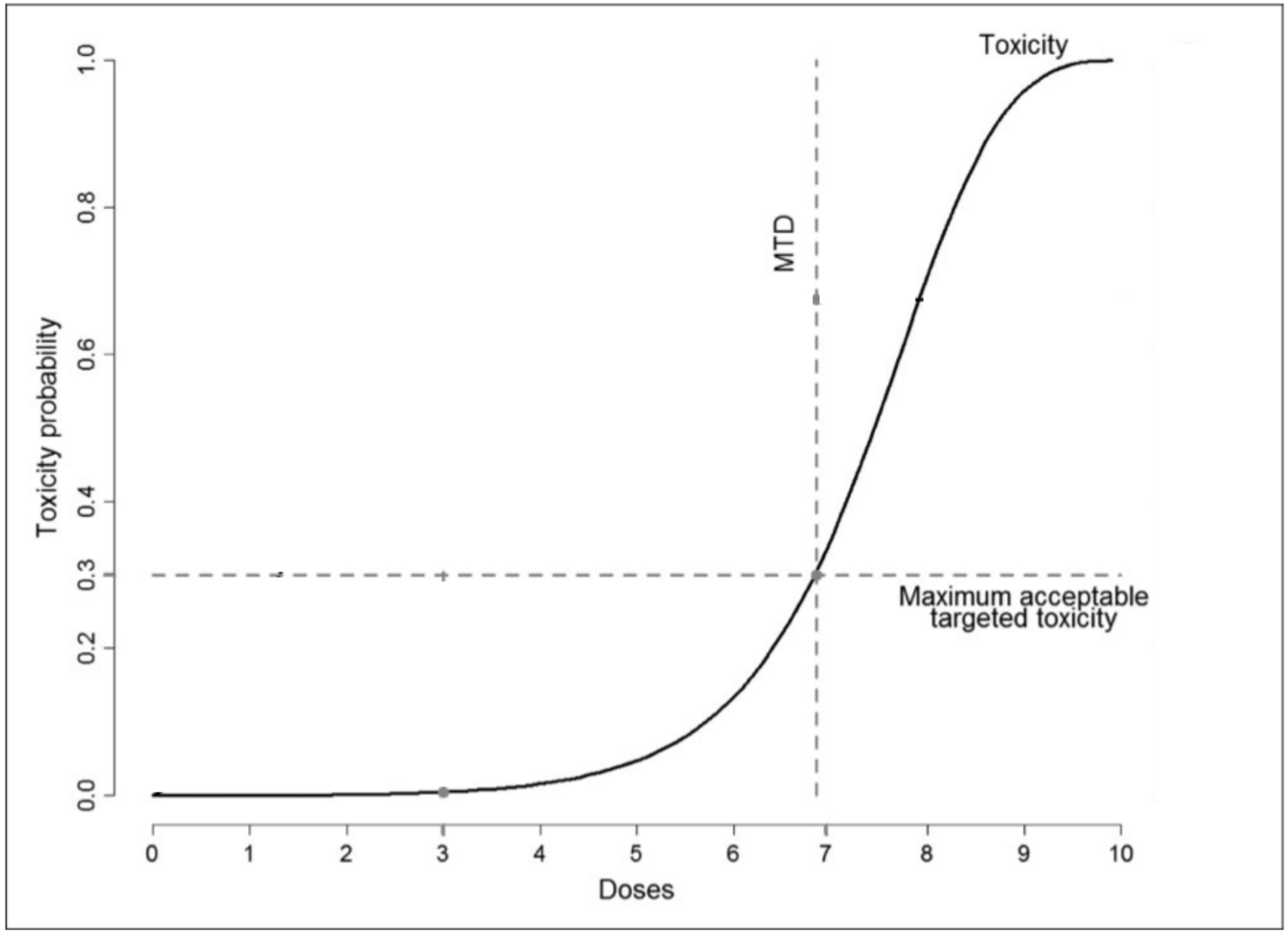
Unethical to test each arm thoroughly. We add assumptions about underlying structure.

DOSE FINDING SETUP

- Before, # coins >> budget. Now, #coins < budget but budget too small to test each coin thoroughly.

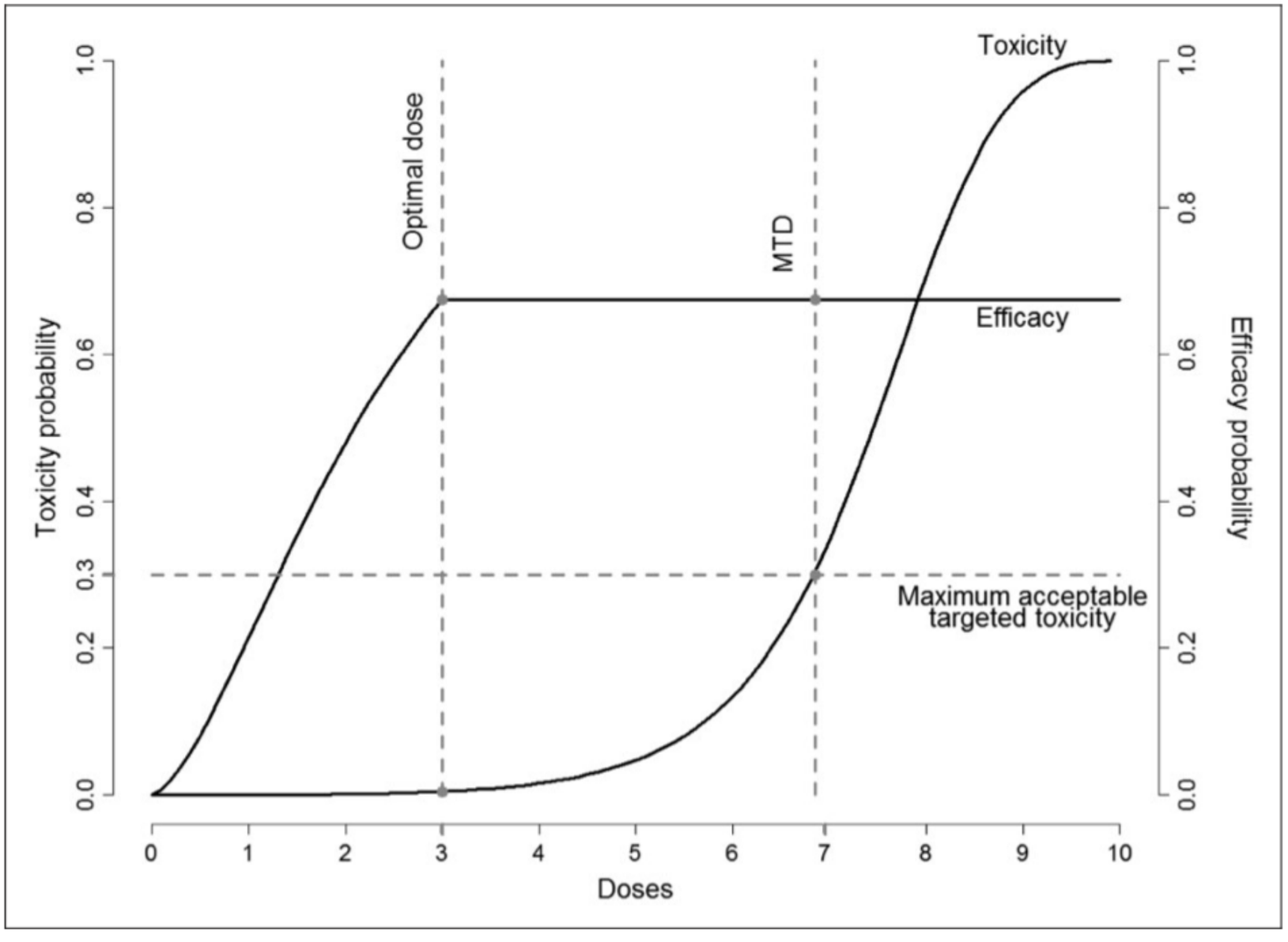
- Before, we cared about simple regret. Now, we care about both simple regret (dose recommendation) and cumulative regret (dose allocation).

DOSE FINDING WHEN TOXICITY INCREASES WITH EFFICACY



Unethical to test each arm thoroughly. We add assumptions about underlying structure.

DOSE FINDING WHEN TOXICITY INCREASES AND EFFICACY PLATEAUS



Unethical to test each arm thoroughly. We add assumptions about underlying structure.

DOSE FINDING ALGORITHM

- We developed variants of the Thompson Sampling algorithm to handle certain constraints related to cancer research, focusing on finding an optimal drug dosage with minimal trials.

- Our algorithms beat the state-of-the-art algorithm. We also provided theoretical insights on the problem.

TOXICITY-ONLY RESULT

Max Tolerable Tox: 0.3, 36 patients, 2k simulations

Dose:		1	**2**	3	4	5	6
Toxicity:		0.1	**0.25**	0.4	0.5	0.65	0.75
CRM	Rec	4.8	49.7	39.0	6.5	0.1	0.0
	Alloc	**17.8**	38.3	30.9	9.0	2.4	1.7
TS_A	Rec	3	**50.8**	36.4	7	1.6	1.1
	Alloc	29.6	**40.1**	**23.4**	**6.1**	**0.8**	**0.1**

TOX+EFFICACY RESULT

Max Tolerable Tox: 0.35, 60 patients, 2k simulations

Dose:		1	2	**3**	4	5	6
Toxicity:		0.01	0.05	**0.15**	0.2	0.45	0.6
Efficacy:		0.1	0.35	**0.6**	0.6	0.6	0.6
MTA-RA	Rec	0.4	7.0	54.9	29.1	7.4	0.7
	Alloc	**7.1**	**14.2**	**37.9**	24.9	12.9	2.5
TS_A	Rec	0.3	9.6	**59.4**	26.1	3.5	0.3
	Alloc	10.7	20.7	35.7	**23.9**	**7.3**	**0.9**

ROAD MAP

Theory

Pure exploration in infinitely- armed bandit models with fixed-confidence.
Maryam Aziz, Jesse Anderton, Emilie Kaufmann, and Javed Aslam. ALT (Algorithmic Learning Theory) 18.

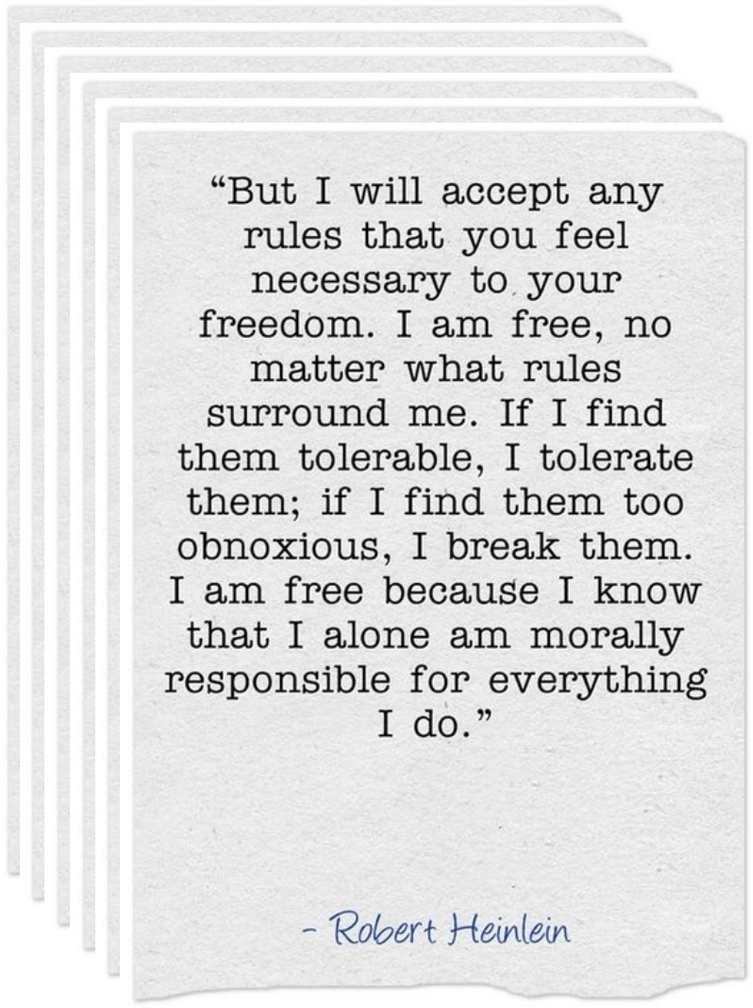
Pure-Exploration for Infinite-Armed Bandits with General Arm Reservoirs.
Maryam Aziz, Kevin Jamieson and Javed Aslam. Under review.

Application

On Multi-Armed Bandit Designs for Phase I Clinical Trials.
Maryam Aziz, Emilie Kaufmann, Marie-Karelle Riviere. Under review.

Adaptively Pruning Features for Boosted Decision Trees.
Maryam Aziz, Jesse Anderton, Javed Aslam. In preprint.

Text Feature Reservoir

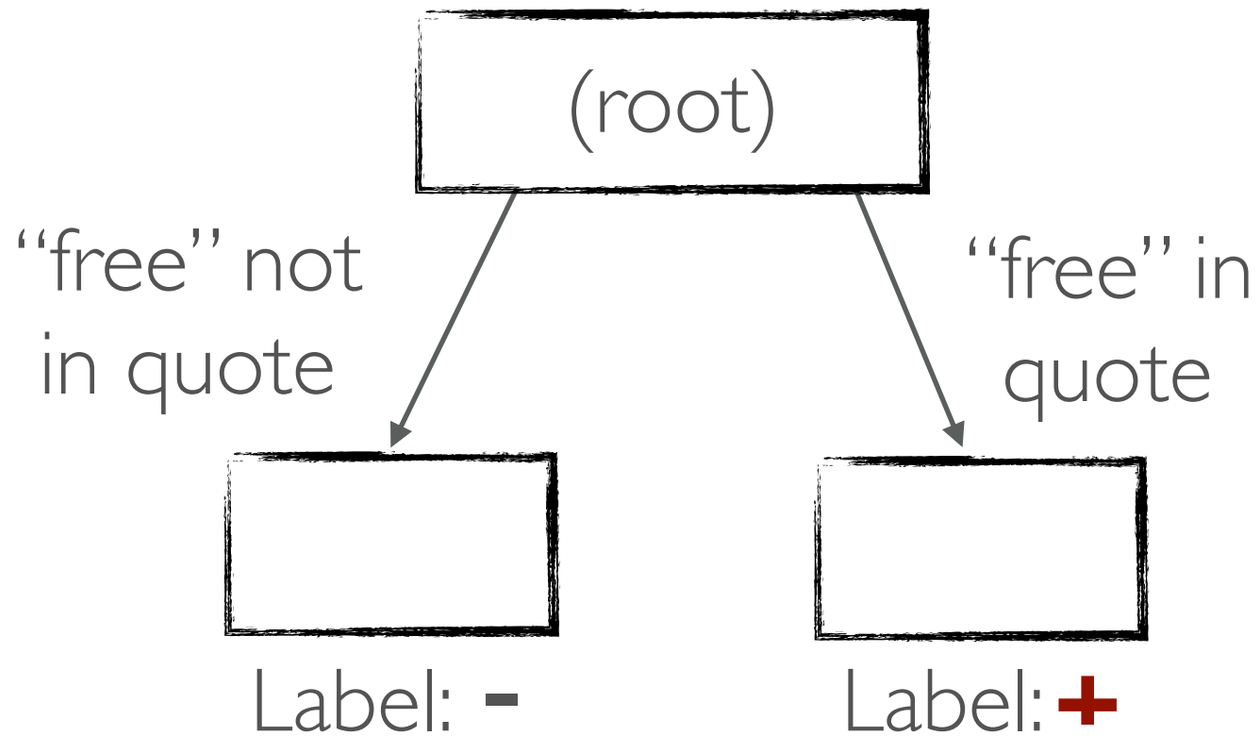


Coin Reservoir

TEXT CLASSIFICATION

Given a small of trials and a large number of "text features" (used to build classifiers), can I find a feature that accurately classifies whether a quote (document) belongs to Robert Heinlein?

DECISION TREES



Quote By:

+ Robert Heinlein
- Anyone Else

- Tree split rule: if a skip-gram appears in a document send it right, otherwise left.
- We assign a label to a quote based on which leaf it lands on.

BOOSTED DECISION TREES

For each round $\mathbf{t = 1, 2, ..., T}$:

You have a set of documents $\mathbf{x_i}$ and weights $\mathbf{w_i}$

1. A tree $\mathbf{h_t(x)}$ is chosen to optimize (e.g.) weighted accuracy

2. Boosting evaluates $\mathbf{h_t(x)}$ and chooses a confidence score $\boldsymbol{\alpha_t}$

3. Update weights $\mathbf{w_i}$ of document $\mathbf{x_i}$

Boosted Classifier

Decision tree

$$\mathcal{H}_T(x) = \sum_t^T \alpha_t h_t(x)$$

WEIGHTED ACCURACY

$$\sum_i w_i \, 1\{y_i = h_t(x_i)\}$$

document ($\mathbf{i}$)	Weight ($\mathbf{w_i}$)	Label ($\mathbf{y_i}$)	$\mathbf{h_t(x_i)}$	Acc
1	0.15	+	-	0.00
2	0.15	-	-	0.15
3	0.4	+	+	0.40
4	0.3	+	-	0.00
				0.55

FEATURE MATRIX

m features (e.g. was skip-gram **j** in doc **i**?)

n documents

	f_1	f_2	.	.	f_m
x_1	Y	N	N	Y	Y
x_2	Y	N	N	Y	Y
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_n	Y	N	Y	Y	Y

REWARD MATRIX

Trees are "coins," reward if correct

	h_1	h_2	.	.	h_{2m}
x_1	1	0	0	1	1
x_2	1	0	0	1	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_n	1	0	1	1	1

Documents

Each cell is a "coin flip"

CHOOSING TREES

Decision rule: doc goes right if it predicted +, otherwise goes left.

feature 1

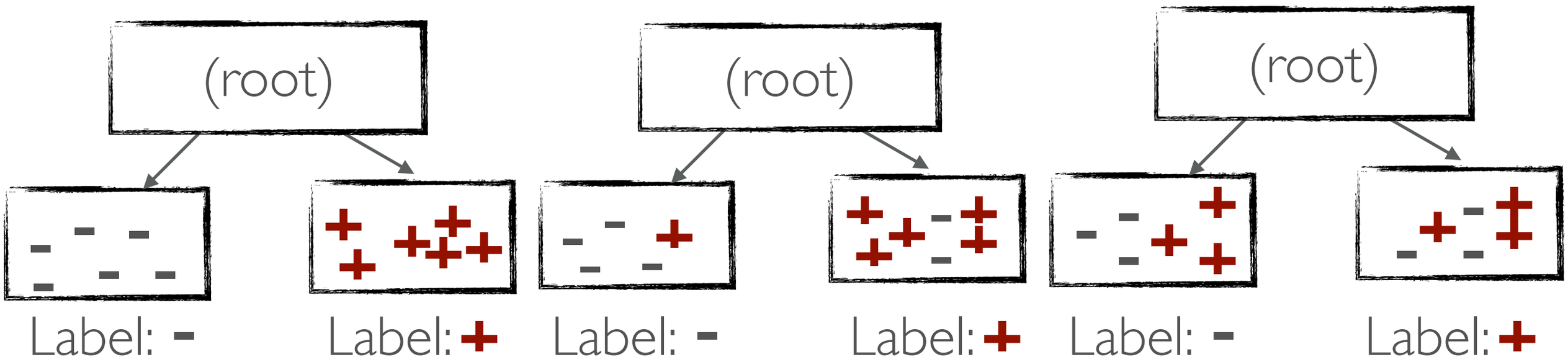
acc=1

feature 2

acc=9/12 =0.75

feature 3

acc=6/12=0.5



The simplest way to choose a tree is to calculate the weighted accuracy for each, using all training documents, and pick a winner.

OUR GOAL

Decision rule: doc goes right if it predicted + , otherwise goes left.

feature 1

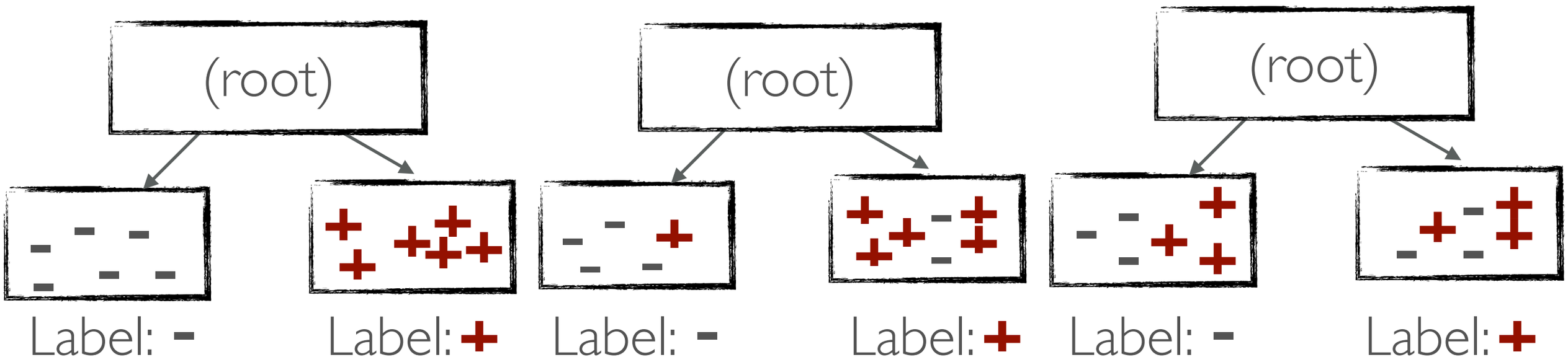
acc=1

feature 2

acc=9/12 =0.75

feature 3

acc=6/12=0.5

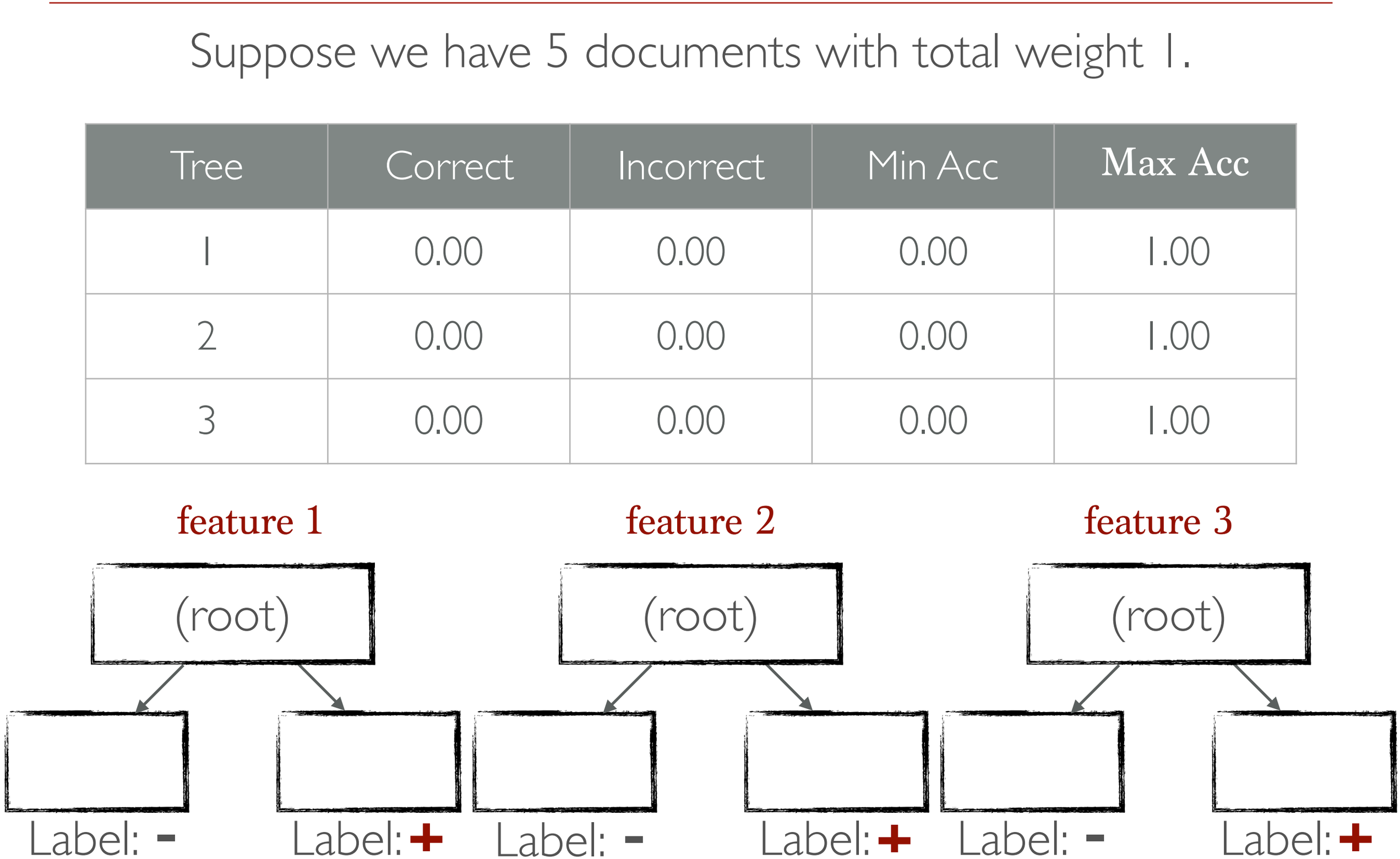


AP-Boost (our algorithm) finds the best tree (coin) while minimizing the number of documents (budget) we train on.

AP-BOOST INTUITION

Suppose we have 5 documents with total weight 1.

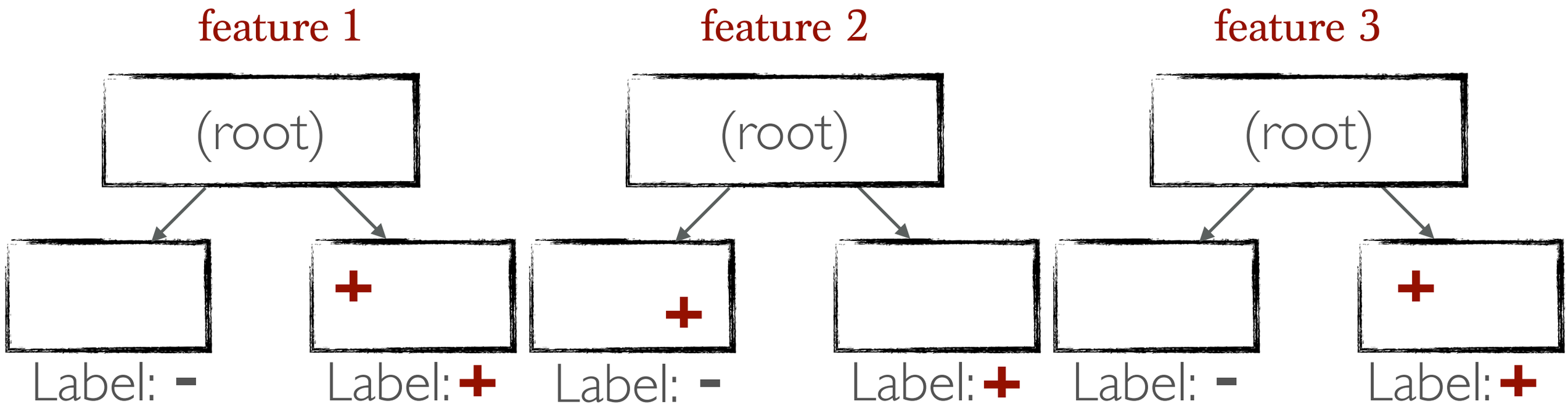
Tree	Correct	Incorrect	Min Acc	Max Acc
1	0.00	0.00	0.00	1.00
2	0.00	0.00	0.00	1.00
3	0.00	0.00	0.00	1.00



AP-BOOST INTUITION

Add document 1, weight 0.20

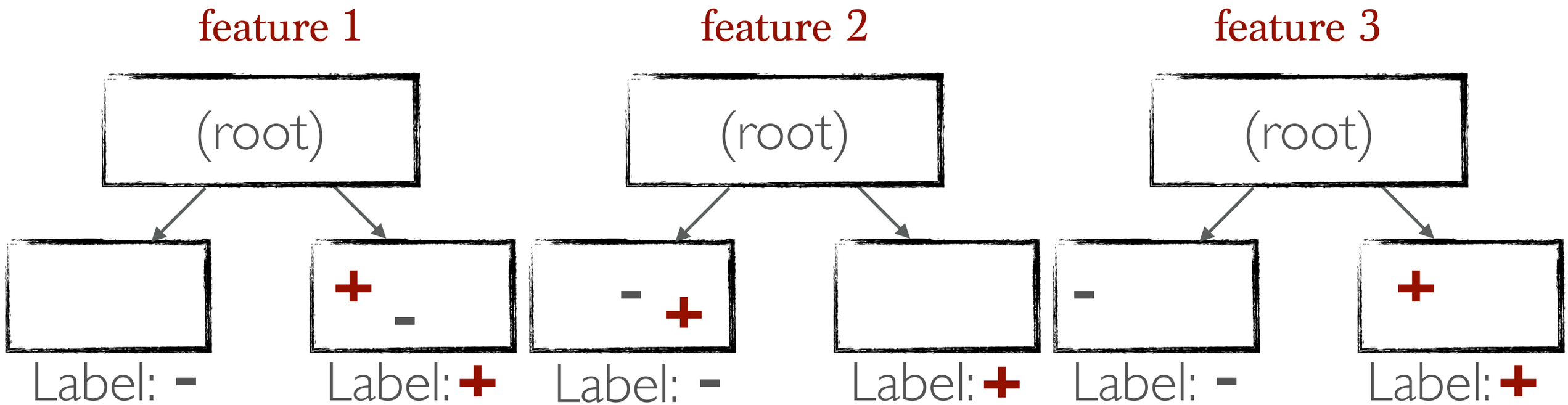
Tree	Correct	Incorrect	Min Acc	Max Acc
1	0.20	0.00	**0.20**	1.00
2	0.00	0.20	0.00	0.80
3	0.20	0.00	**0.20**	1.00



AP-BOOST INTUITION

Add document 2, weight 0.20

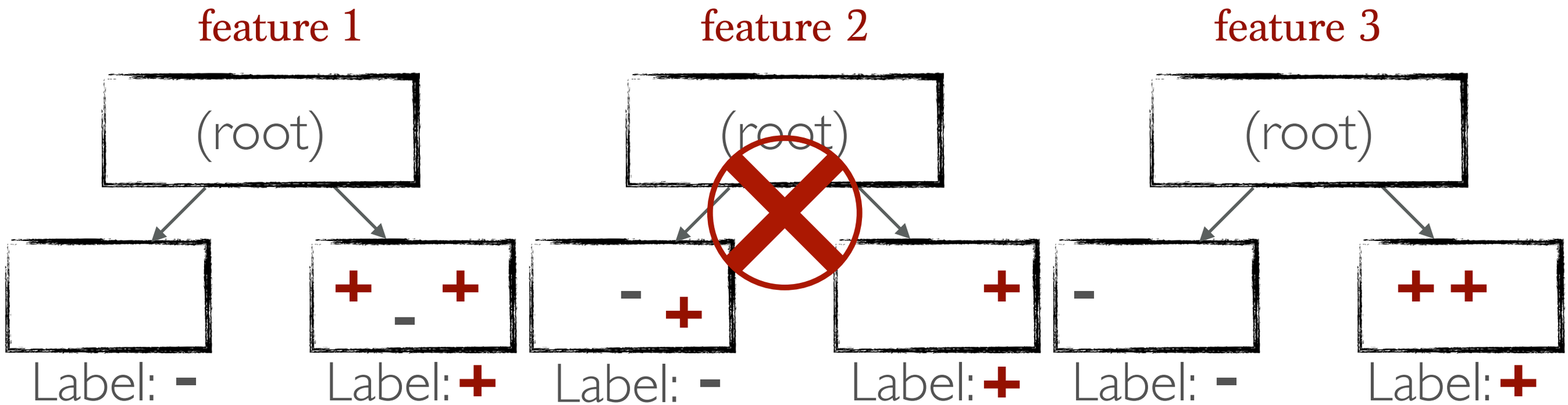
Tree	Correct	Incorrect	Min Acc	Max Acc
1	0.20	0.20	0.20	0.80
2	0.20	0.20	0.20	0.80
3	0.40	0.00	**0.40**	1.00



AP-BOOST INTUITION

Add document 3, weight 0.20, rule out tree 2.

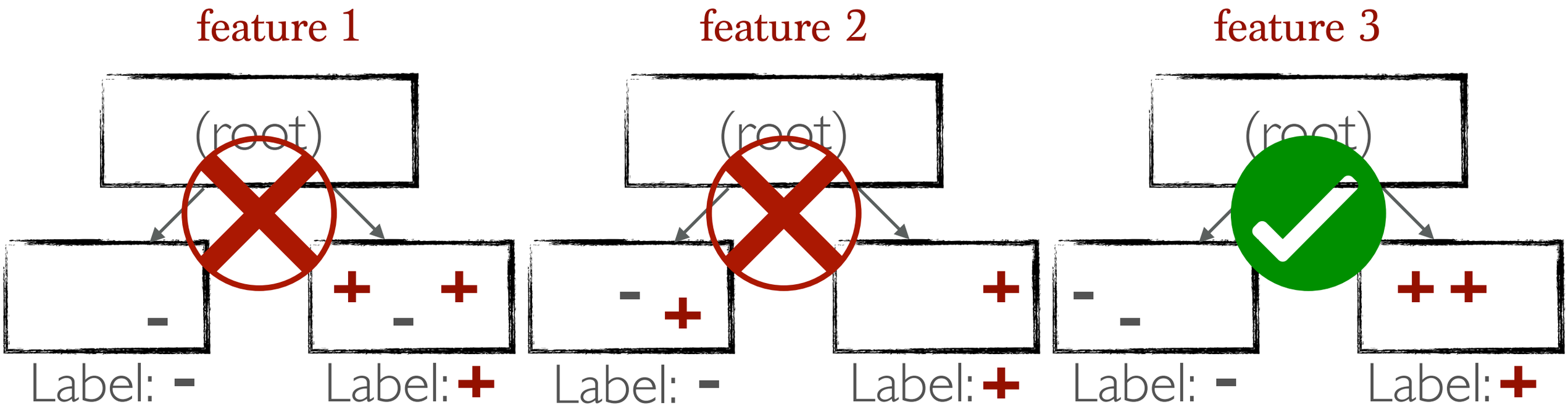
Tree	Correct	Incorrect	Min Acc	**Max Acc**
1	0.40	0.20	0.40	0.80
~~2~~	~~0.20~~	~~0.40~~	~~0.20~~	~~0.60~~
3	0.60	0.00	**0.60**	1.00



AP-BOOST INTUITION

Add document 4, weight 0.20, rule out tree 1, pick tree 3.

Tree	Correct	Incorrect	Min Acc	Max Acc
~~1~~	~~0.60~~	~~0.20~~	~~0.60~~	~~0.80~~
~~2~~	~~0.20~~	~~0.40~~	~~0.20~~	~~0.60~~
3	0.80	0.00	**0.80**	1.00



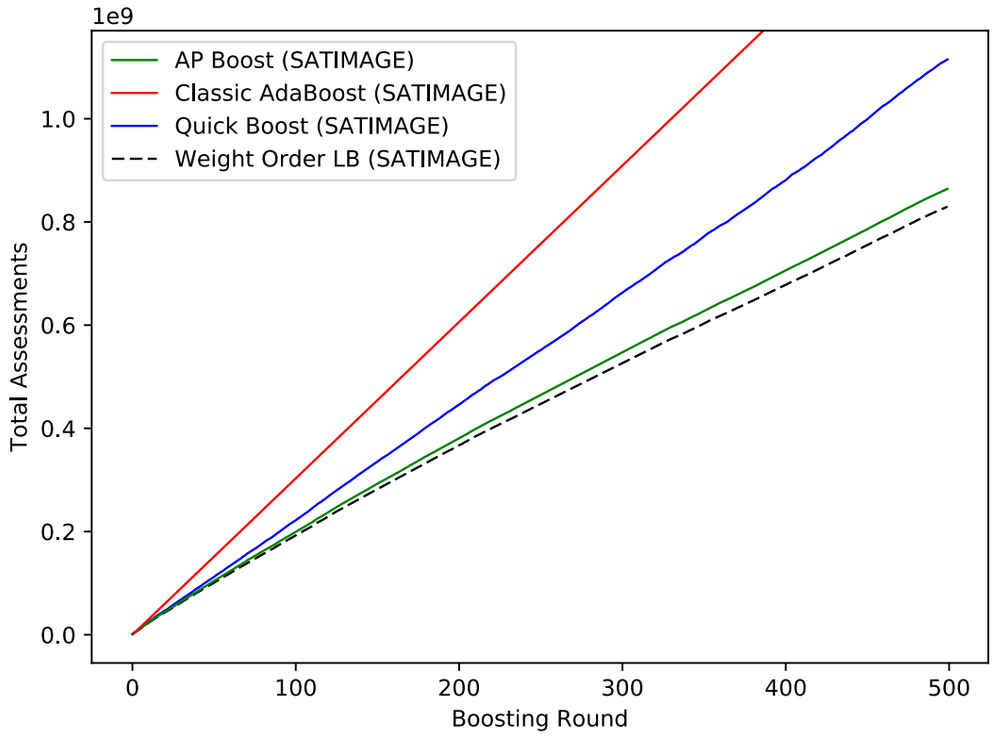
AP BOOST SUMMARY

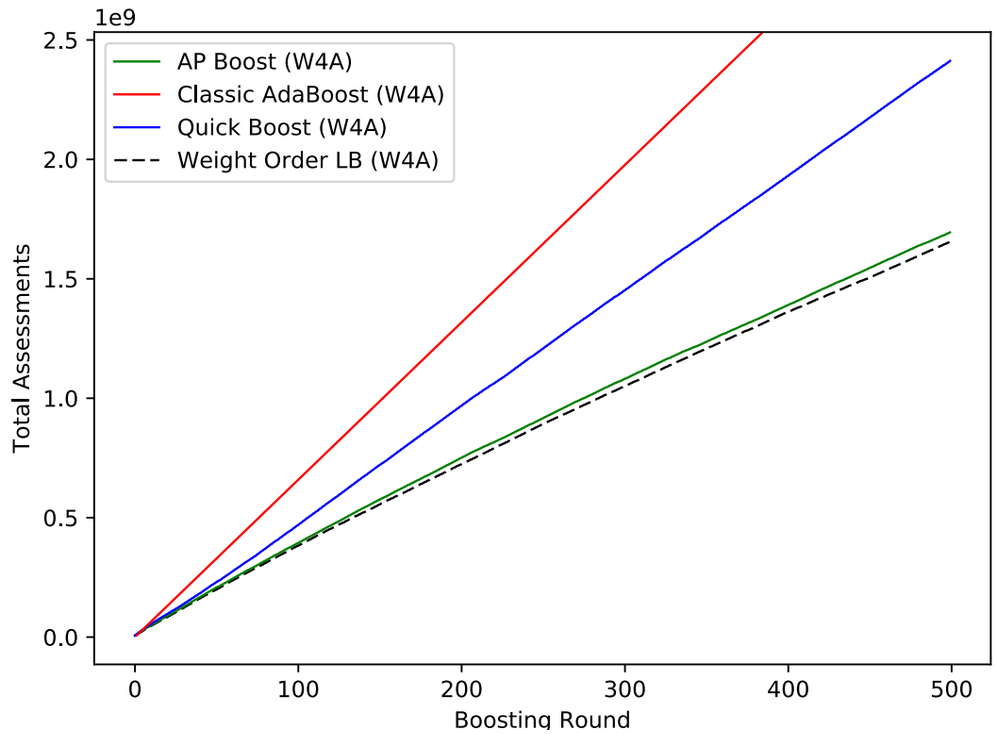
- Use intervals to rule out bad trees with few examples.

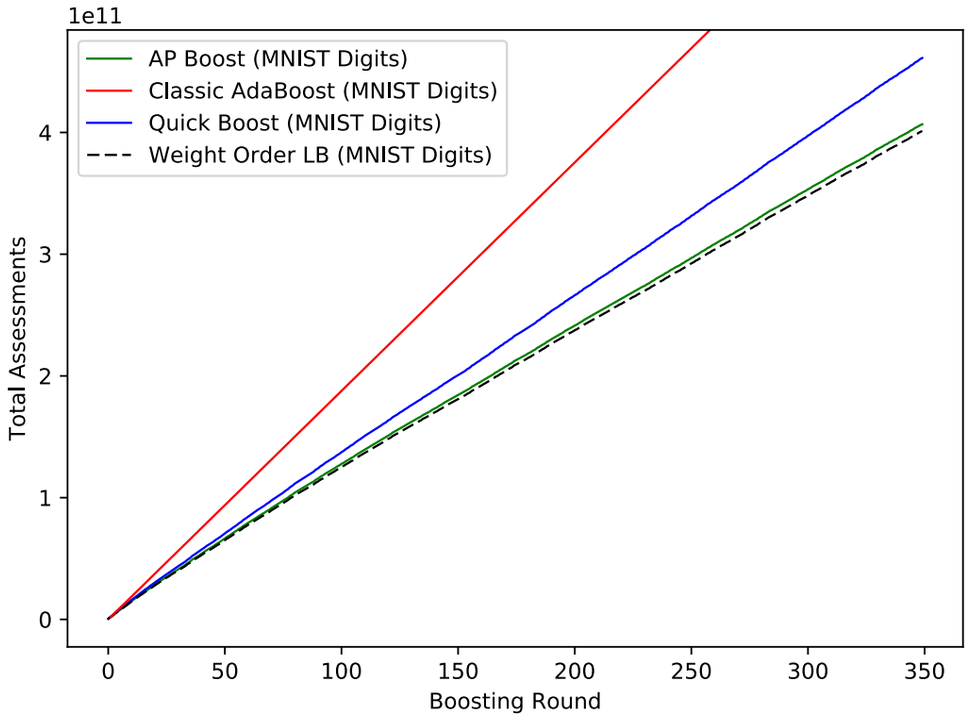
- Use LUCB-style bandit algorithm to strategically pick next tree to assess examples for.

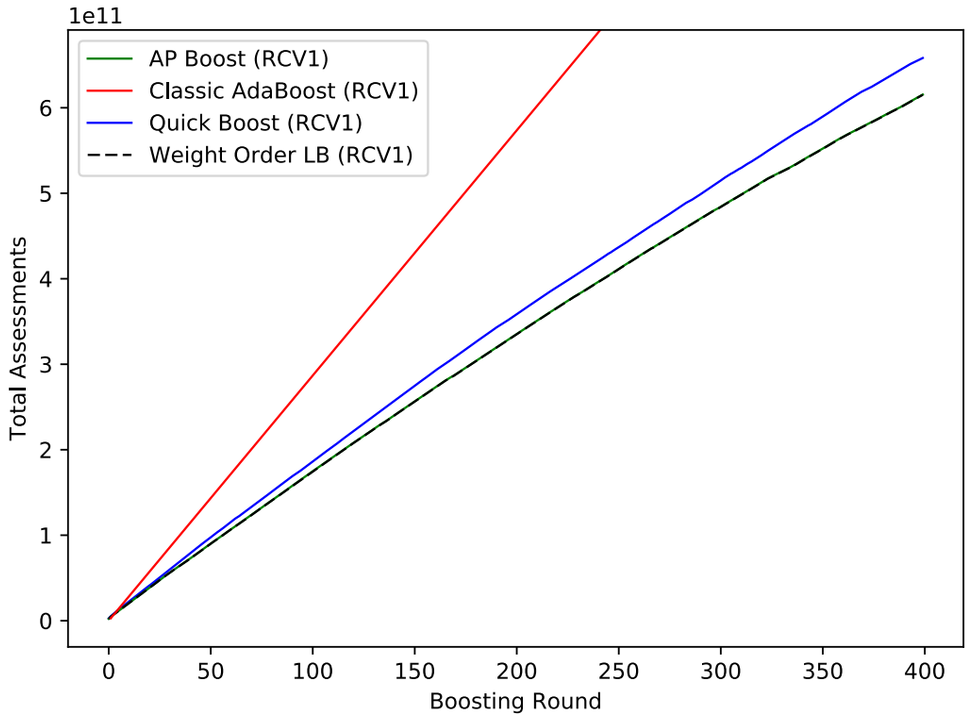
- Add examples in decreasing weight order, because that changes intervals the most.

AP BOOST – A BANDIT-INSPIRED ALGORITHM









CONCLUSION

Theory:

Pure Exploration for the Infinitely Armed Bandit Models		
	Lower Bound	**Algorithm/Upper Bound**
Fixed Confidence	Completed Work	Completed Work
Fixed Budget	Completed Work	Completed Work

Applications:

1) Dose Finding,
2) Boosted Decision Tree Training