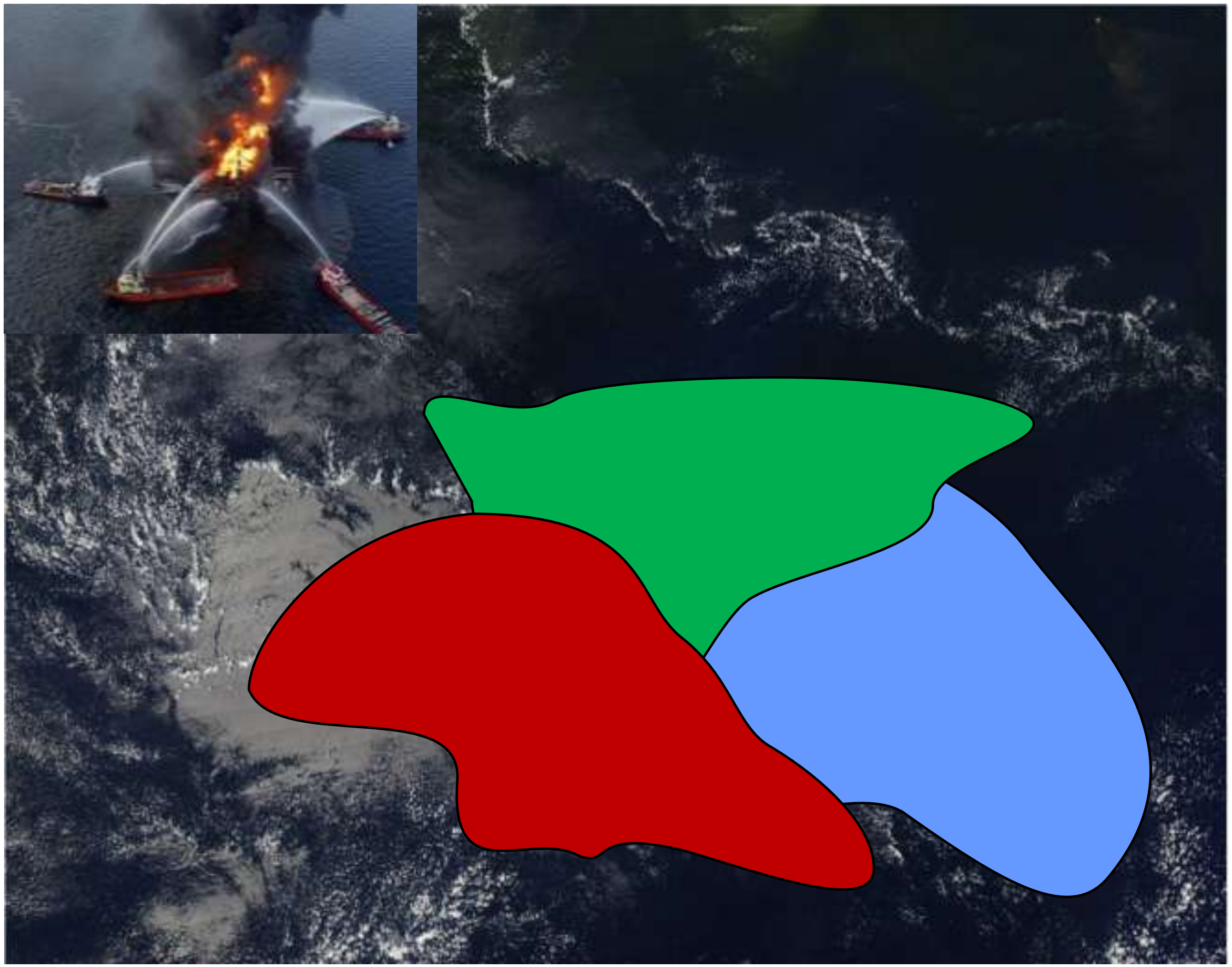
Adaptive Submodularity: A New Approach to Active Learning and Stochastic Optimization

Daniel Golovin and Andreas Krause

(FYI, a powerpoint version of these slides is available on Daniel Golovin's website.)

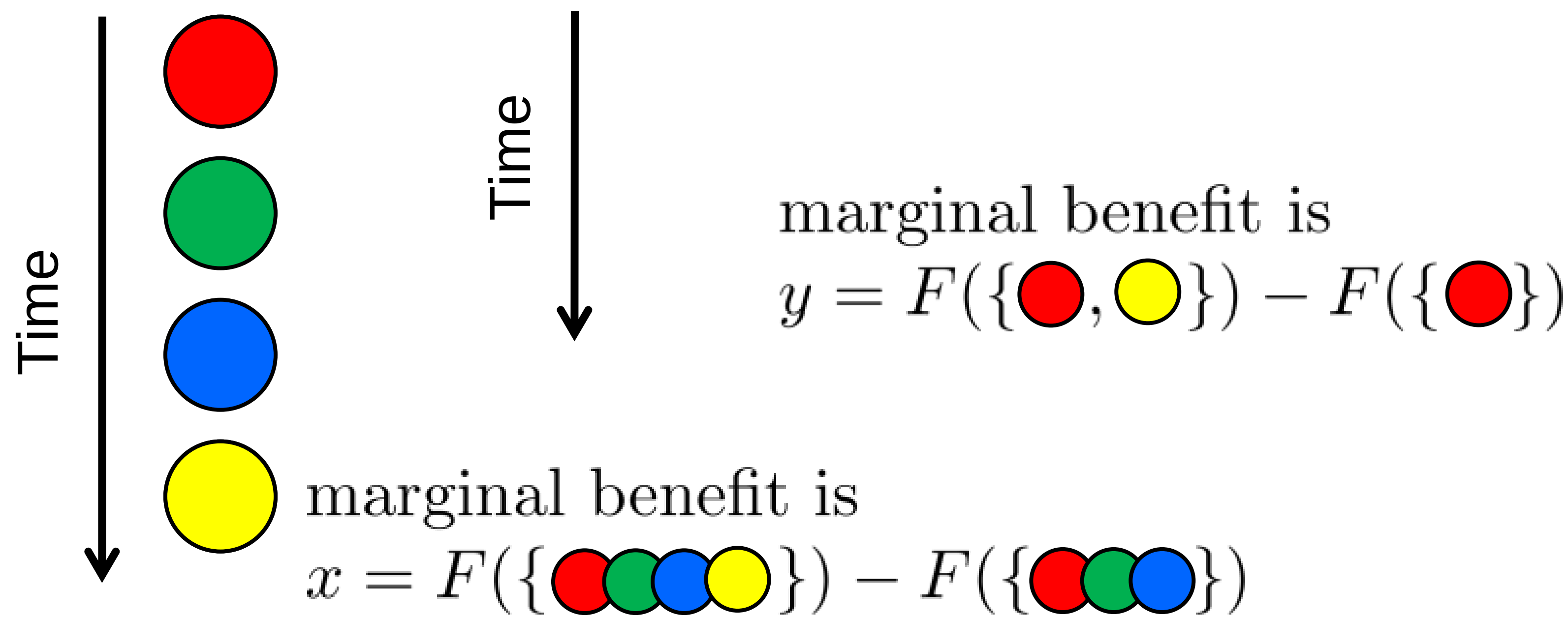
Max K-Cover (Oil Spill Edition)



Submodularity

Discrete diminishing returns property for set functions.

``Playing an action at an earlier stage
only increases its marginal benefit''



Submodularity implies $y \geq x$.

The Greedy Algorithm

Problem: Find $S^* = \arg\max\{f(S) : |S| \leq k\}$

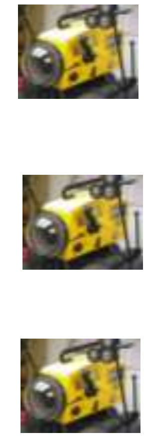
Initialize $S_0 = \emptyset$
For $i = 1, 2, \ldots, k$
$e_i = \arg\max_e f(S_{i-1} \cup \{e\})$
$S_i = S_{i-1} \cup \{e_i\}$
Select e_i

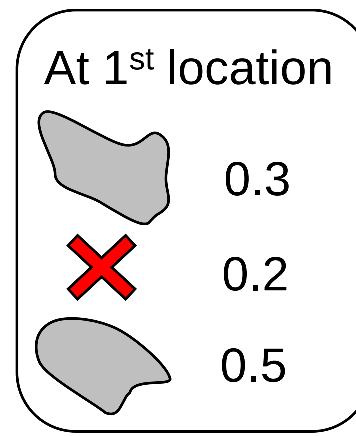
Theorem [Nemhauser *et al* '78]
Given a monotone submodular function f, $f(\emptyset) = 0$, the greedy algorithm selects a set S_{greedy} such that

$$f(S_{\text{greedy}}) \geq (1 - 1/e) \max_{|S| \leq k} f(S)$$

Stochastic Max K-Cover

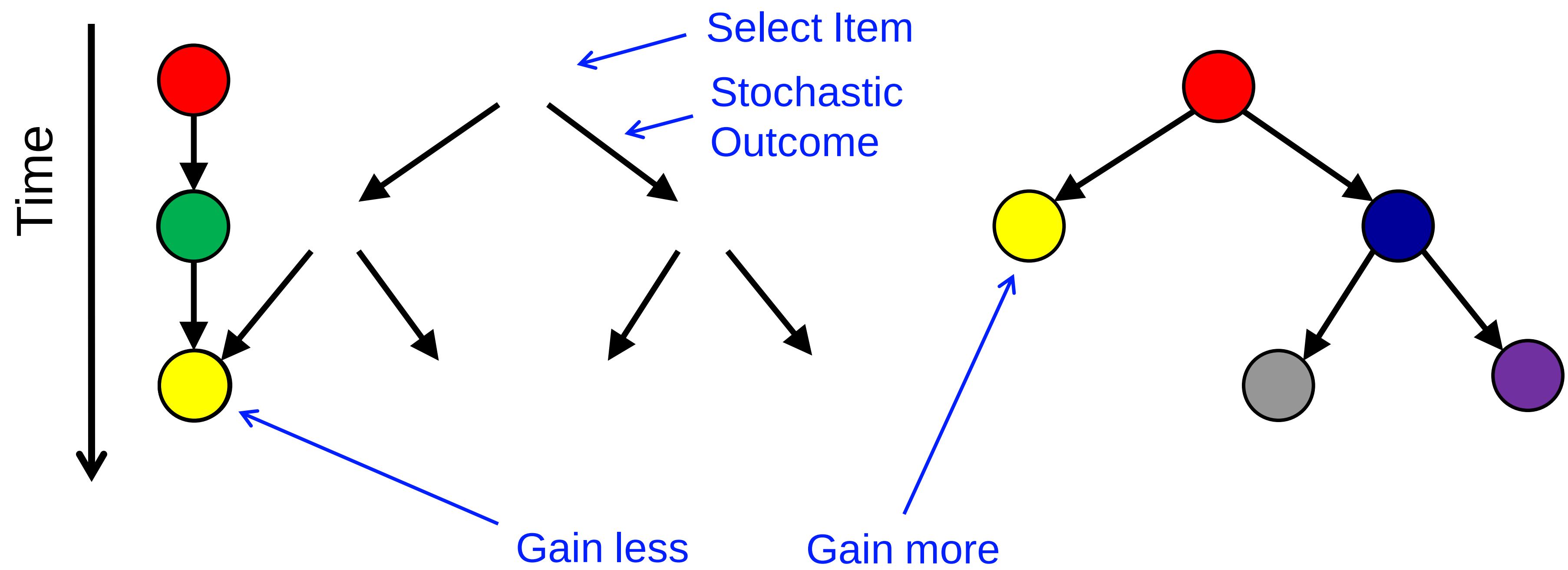
Bayesian: Known failure distribution.
Adaptive: Deploy a sensor and see what you get. Repeat K times.





Asadpour & Saberi (`08): (1-1/e)-approx if sensors (independently) either work perfectly or fail completely.

Adaptive Submodularity



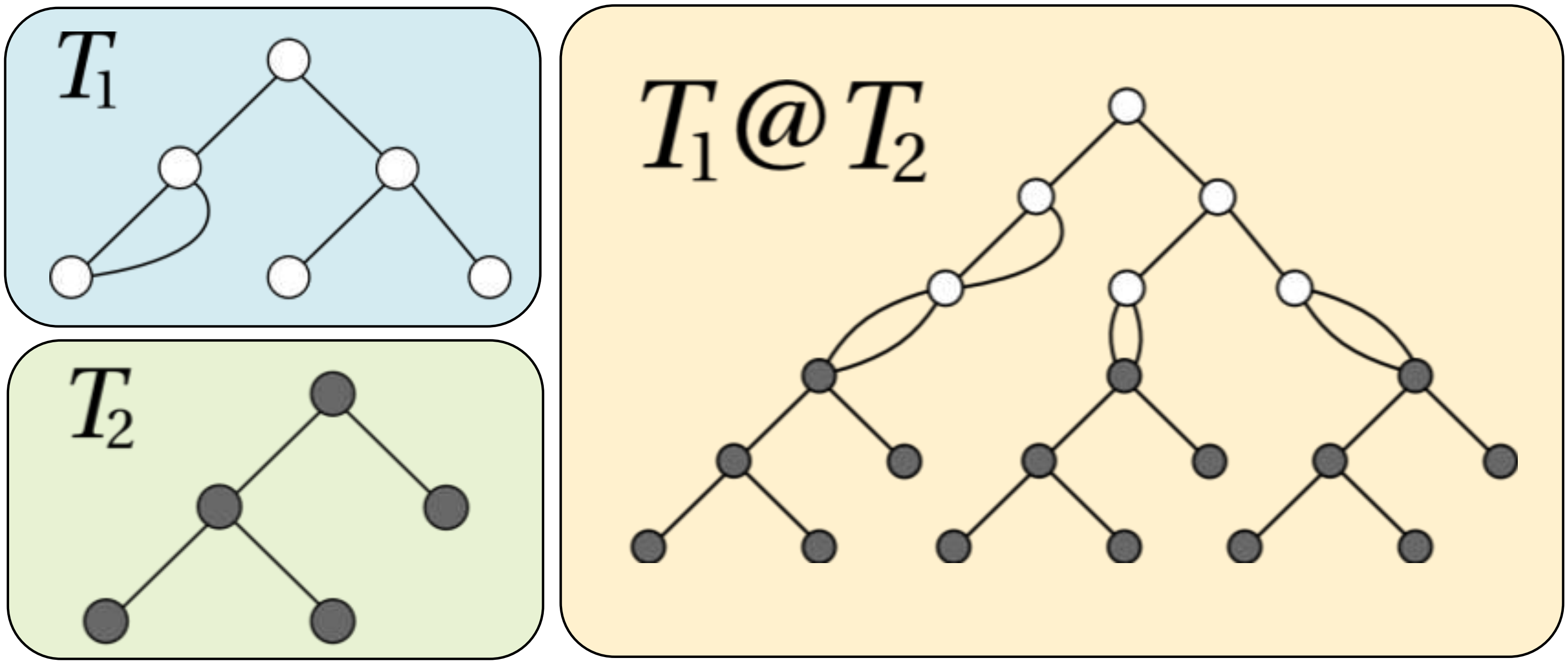
Playing an action at an earlier stage (i.e., at an ancestor) only increases its marginal benefit

expected

(taken over its outcome)

Adaptive Monotonicity

Definition: For all policies T_1 and T_2, the benefit of $\boxed{T_1 @ T_2}$ is at least that of $\boxed{T_2}$



What is it good for?

Allows us to generalize various results to the adaptive realm, including:

- (1-1/e)-approximation for Max K-Cover, submodular maximization subject to a cardinality constraint
- (ln(n)+1)-approximation for Set Cover

Recall the Greedy Algorithm

Problem: Find $S^* = \arg\max\{f(S) : |S| \leq k\}$

Initialize $S_0 = \emptyset$
For $i = 1, 2, \ldots, k$
    $e_i = \arg\max_e f(S_{i-1} \cup \{e\})$
    $S_i = S_{i-1} \cup \{e_i\}$
    Select e_i

Theorem [Nemhauser *et al* '78]
Given a monotone submodular function f, $f(\emptyset) = 0$, the greedy algorithm selects a set S_{greedy} such that

$$f(S_{\text{greedy}}) \geq (1 - 1/e) \max_{|S| \leq k} f(S)$$

The Adaptive-Greedy Algorithm

Problem: Find $T^* = \arg\max\{F(T) : \text{depth}(T) \leq k\}$

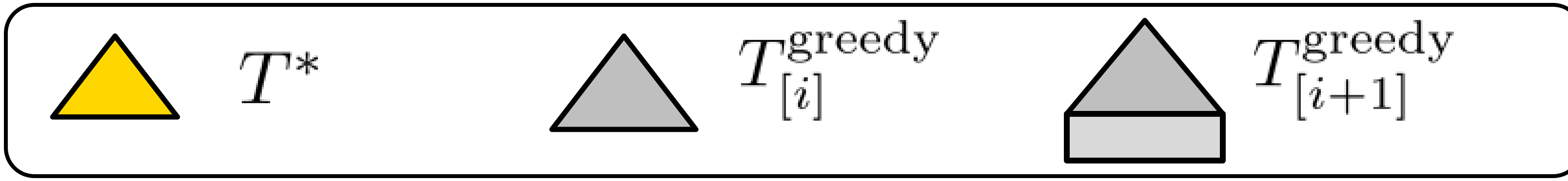
Initialize $S_0 = \emptyset$
For $i = 1, 2, \ldots, k$
$e_i = \arg\max_e \mathbb{E}[f(S_{i-1} \cup \{e\}) \mid \text{outcomes } o_1, \ldots, o_{i-1}]$
$S_i = S_{i-1} \cup \{e_i\}$
Select e_i and observe outcome o_i for it.

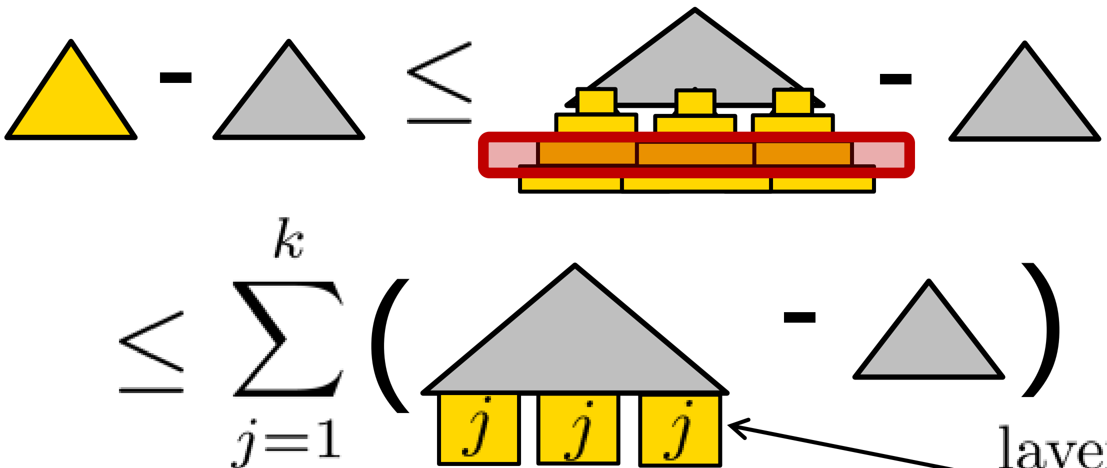
Theorem

Given an adaptive monotone submodular function f with $f(\emptyset) = 0$, the adaptive greedy algorithm returns T^{greedy} such that

$$F(T^{\text{greedy}}) \geq (1 - 1/e) \max_{T : \text{depth}(T) \leq k} F(T)$$

Lemma: $OPT - F(T^{\text{greedy}}_{[i]}) \leq k \cdot \left(F(T^{\text{greedy}}_{[i+1]}) - F(T^{\text{greedy}}_{[i]}) \right)$





[Adapt-monotonicity]

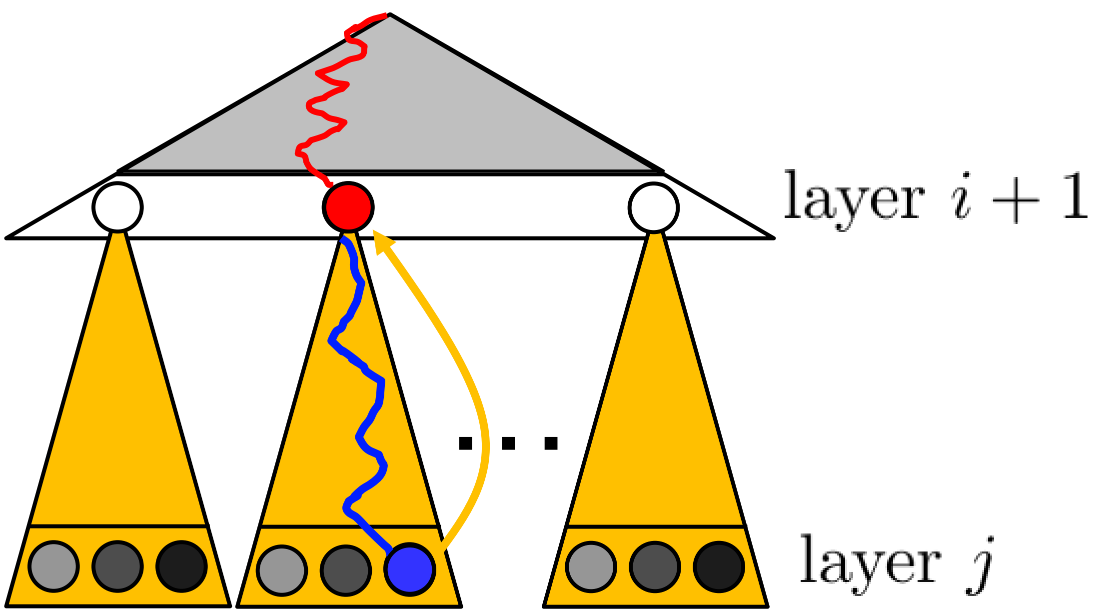
[Adapt-submodularity]

layer j of T^* played as
layer $(i+1)$ of T^{greedy}

How to play layer *j* at layer *i+1*

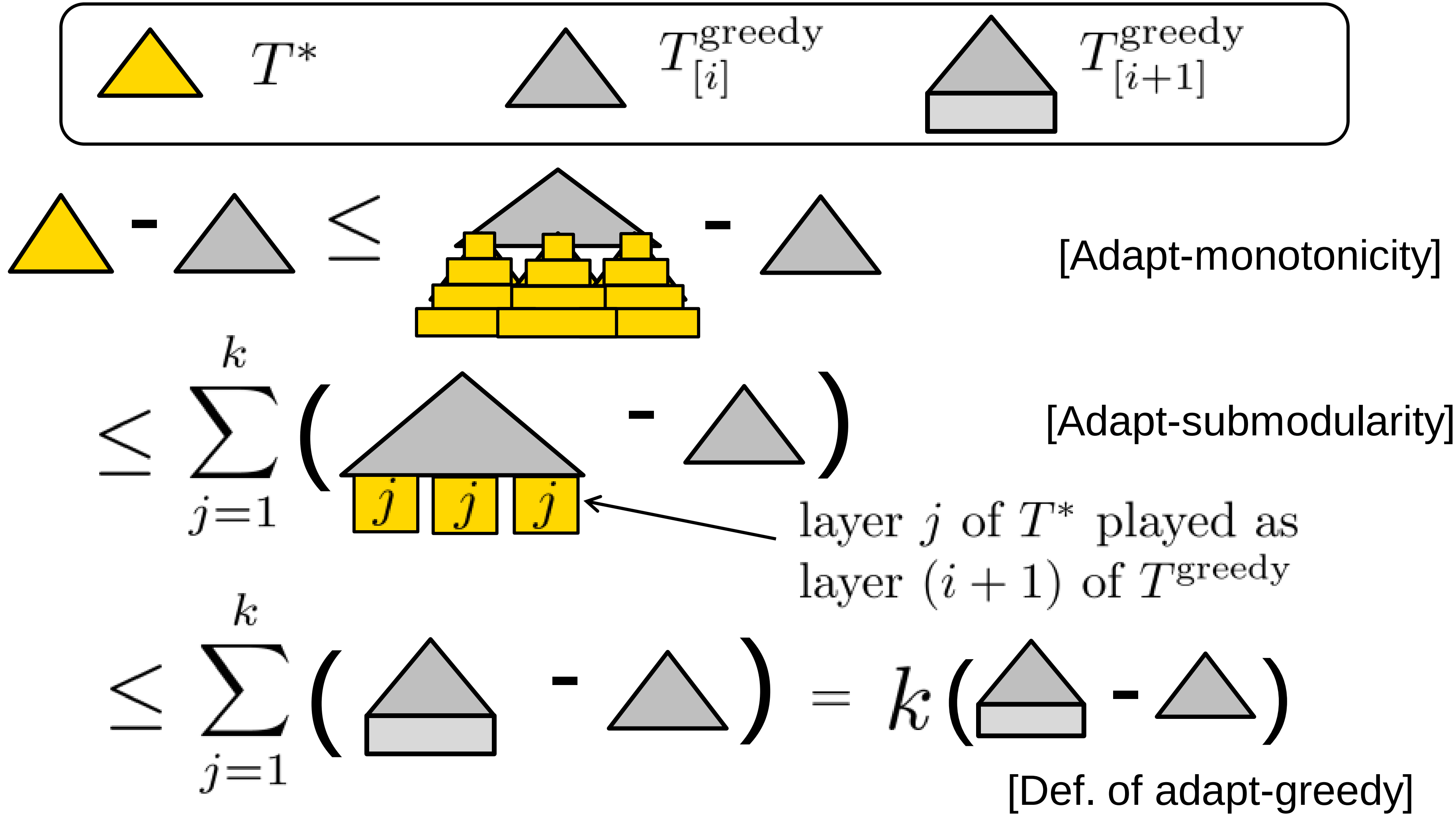
The world-state dictates which path in the tree we'll take.

1. For each node at layer *i+1*,
2. Sample path to layer *j*,
3. Play the resulting layer *j* action at layer *i+1*.



By adapt. submod., playing a layer earlier only increases it's marginal benefit

$$\text{Lemma: } OPT - F(T^{\text{greedy}}_{[i]}) \leq k \cdot \left(F(T^{\text{greedy}}_{[i+1]}) - F(T^{\text{greedy}}_{[i]}) \right)$$



$$\text{Lemma: } OPT - F(T_{[i]}^{\text{greedy}}) \leq k \cdot \left(F(T_{[i+1]}^{\text{greedy}}) - F(T_{[i]}^{\text{greedy}}) \right)$$

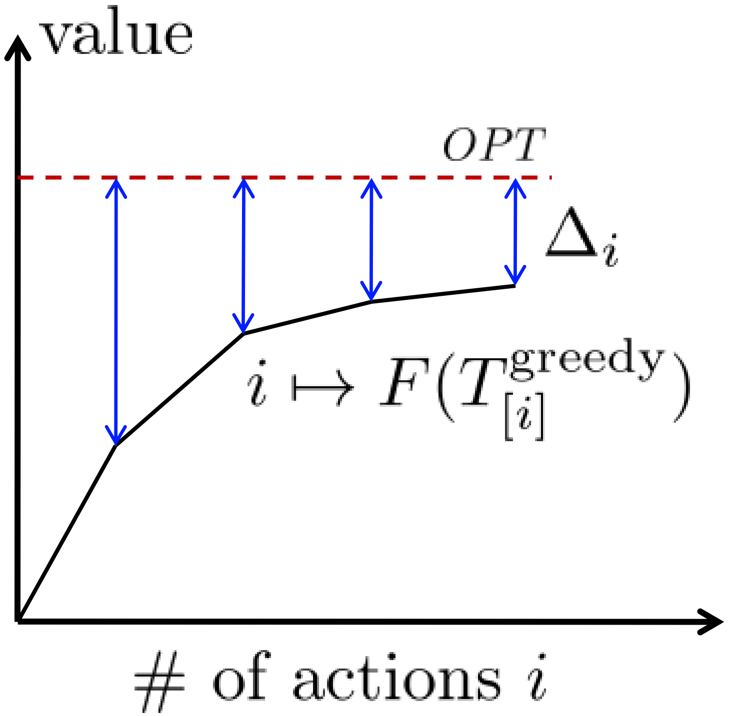
Let $\Delta_i := OPT - F(T_{[i]}^{\text{greedy}})$

$$\Delta_i \leq k\,(\Delta_i - \Delta_{i+1})$$

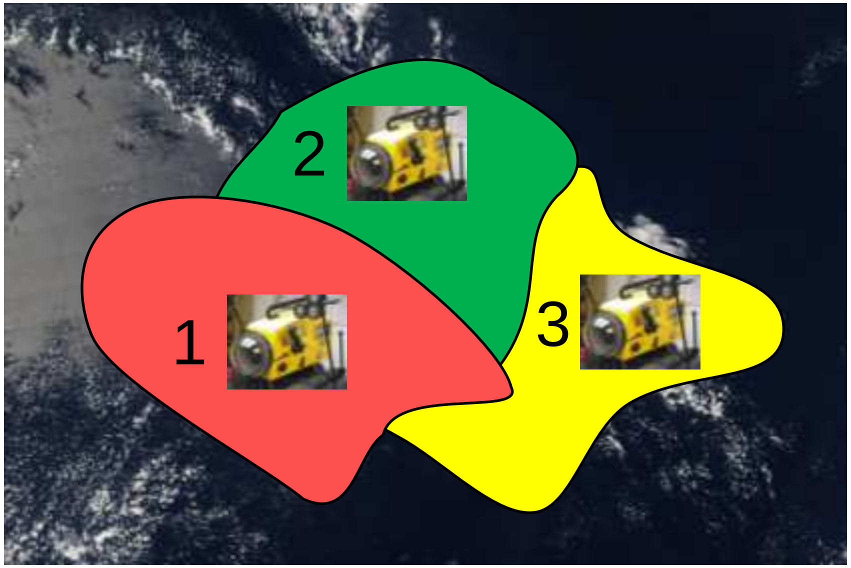
$$\Delta_{i+1} \leq \left(1 - \frac{1}{k}\right) \Delta_i$$

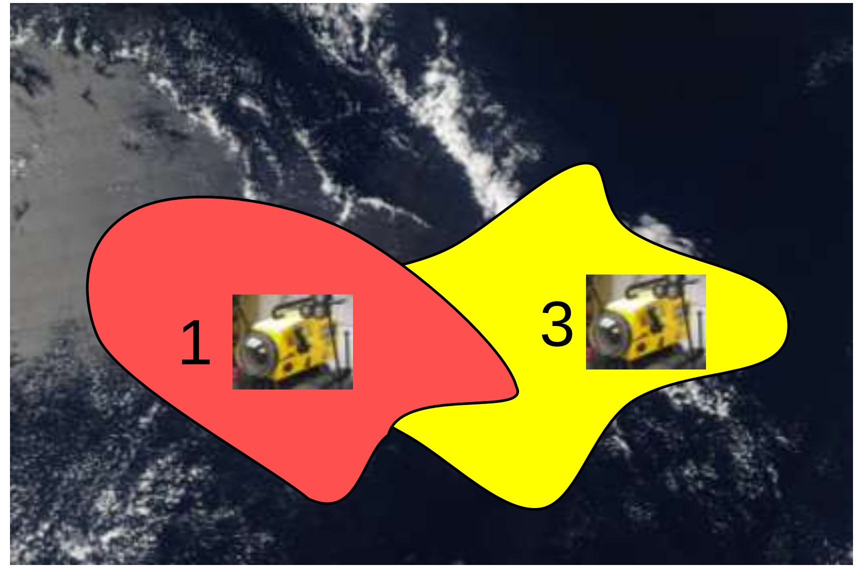
$$\Delta_k \leq \left(1 - \frac{1}{k}\right)^k \Delta_0 \leq \frac{OPT}{e}$$

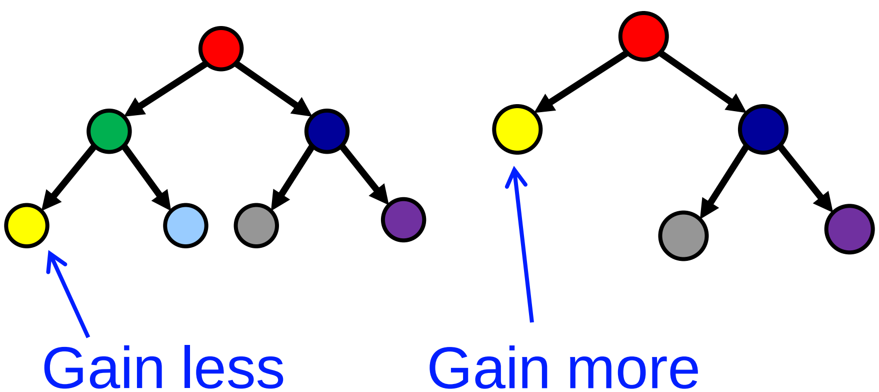
$$F(T_{[k]}^{\text{greedy}}) \geq \left(1 - \tfrac{1}{e}\right) OPT$$



Stochastic Max Cover is Adapt-Submod







Random sets distributed independently.

adapt-greedy is a (1-1/e) ≈ 63% approximation to the adaptive optimal solution.

Stochastic Min Cost Cover

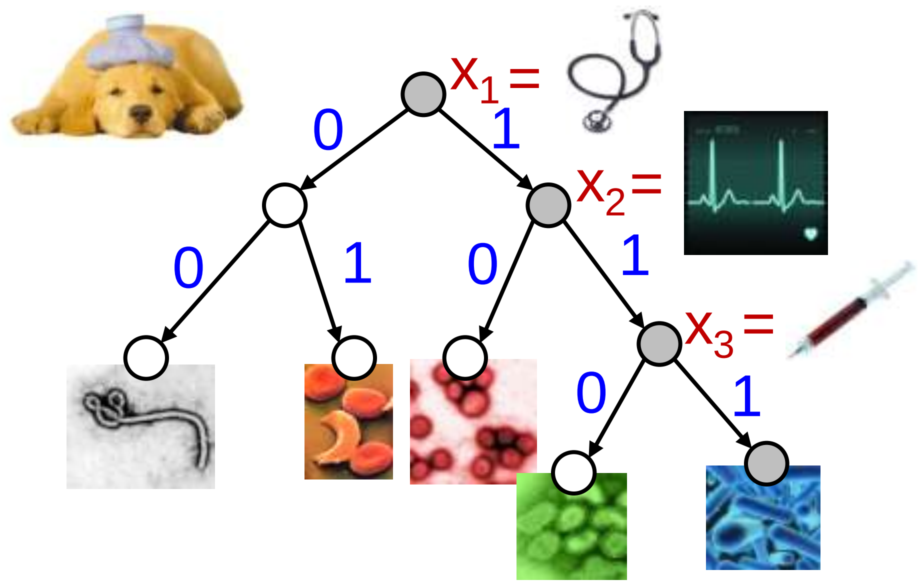
- Adaptively get a threshold amount of value.
- Minimize expected number of actions.
- If objective is adapt-submod and monotone, we get a logarithmic approximation.

$\ln(n) + 1$ for Stochastic Set Cover, [Goemans & Vondrak, LATIN '06] [Liu et al., SIGMOD '08]
matching the Set Cover lower bound
assuming $\mathrm{NP} \not\subseteq \mathrm{DTIME}(n^{O(\log \log n)})$ [Feige, JACM '98]

c.f., Interactive Submodular Set Cover [Guillory & Bilmes, ICML '10]

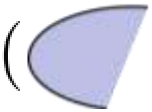
Optimal Decision Trees

"Diagnose the patient as cheaply as possible (w.r.t. expected cost)"



Garey & Graham, 1974;
Loveland, 1985;
Arkin et al., 1993;
Kosaraju et al., 1999;
Dasgupta, 2004;
Guillory & Bilmes, 2009;
Nowak, 2009;
Gupta et al., 2010

Adaptive-Greedy is a $\boxed{(\ln(1/p_{\min}) + 1)}$ approximation.

$b_0 := \mathbb{P}($  $)$

$g_0 := \mathbb{P}($  $)$

$$\Delta_{\text{init}}(x) = \frac{2 g_0 b_0}{g_0 + b_0}$$

$b_1 := \mathbb{P}($  $)$

$g_1 := \mathbb{P}($  $)$

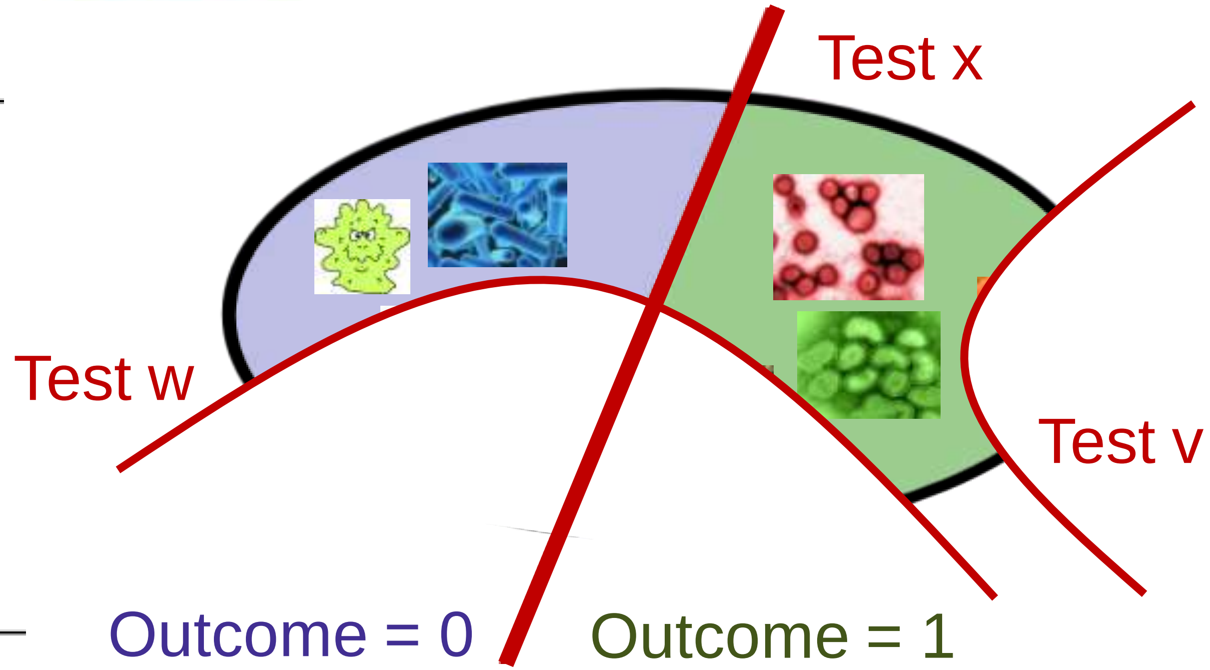
$$\Delta_{\text{final}}(x) = \frac{2 g_1 b_1}{g_1 + b_1}$$

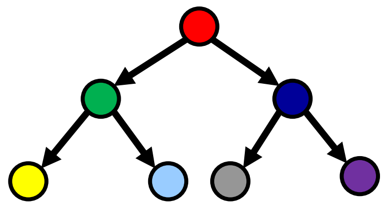
$$b_0 \geq b_1, \quad g_0 \geq g_1$$

Not hard to show $\Delta_{\text{final}}(x) \leq \Delta_{\text{init}}(x)$

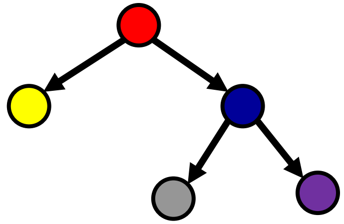
Objective = probability mass of hypotheses you have ruled out.

It's Adaptive Submodular.





Conclusions



- New structural property useful for design & analysis of adaptive algorithms
- Powerful enough to recover and generalize many known results in a unified manner. (We can also handle costs)
- Tight analyses and optimal approximation factors in many cases.