

Submodularity

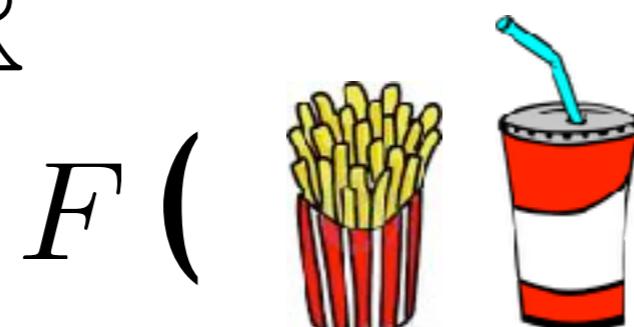
Stefanie Jegelka, MIT
MLSS London, 2019

Set functions

ground set

$$\mathcal{V} = \left\{ \begin{array}{c} \text{salad bowl} \\ \text{burrito} \\ \text{sub sandwich} \\ \text{stacked sandwich} \\ \text{apple} \\ \text{fries and drink} \end{array} \right\}$$

$$F : 2^{\mathcal{V}} \rightarrow \mathbb{R}$$



$F($ $)$ = **cost** of buying items
together, or
utility, or
probability, ...

Subset selection in Machine Learning

training examples



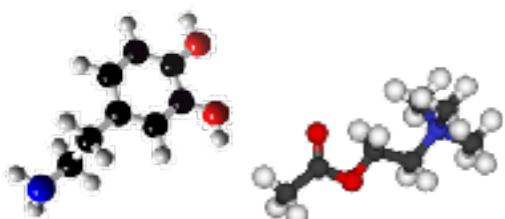
learn model

$$f(x, w)$$

prediction

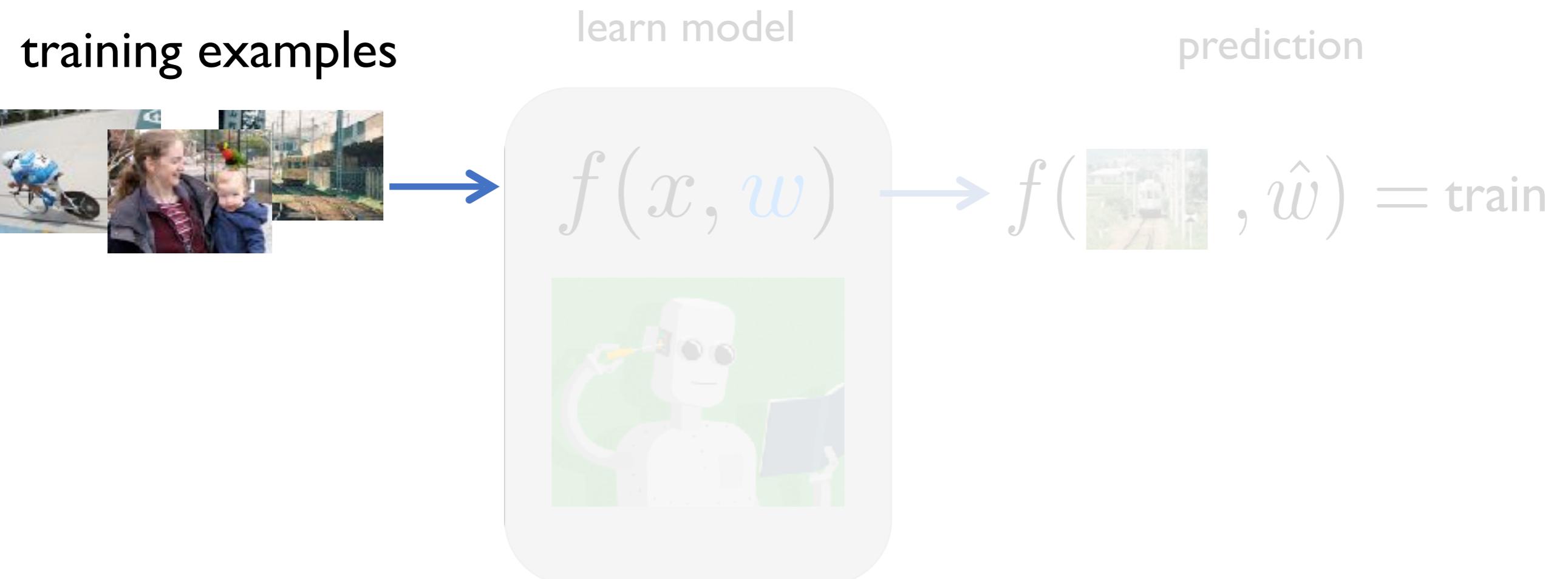


$$f(\text{train}, \hat{w}) = \text{train}$$



$$f(\text{awakening effect}, \hat{w}) = \text{awakening effect}$$

Subset selection in Machine Learning



Informative Subsets



- Compression
- Summarization

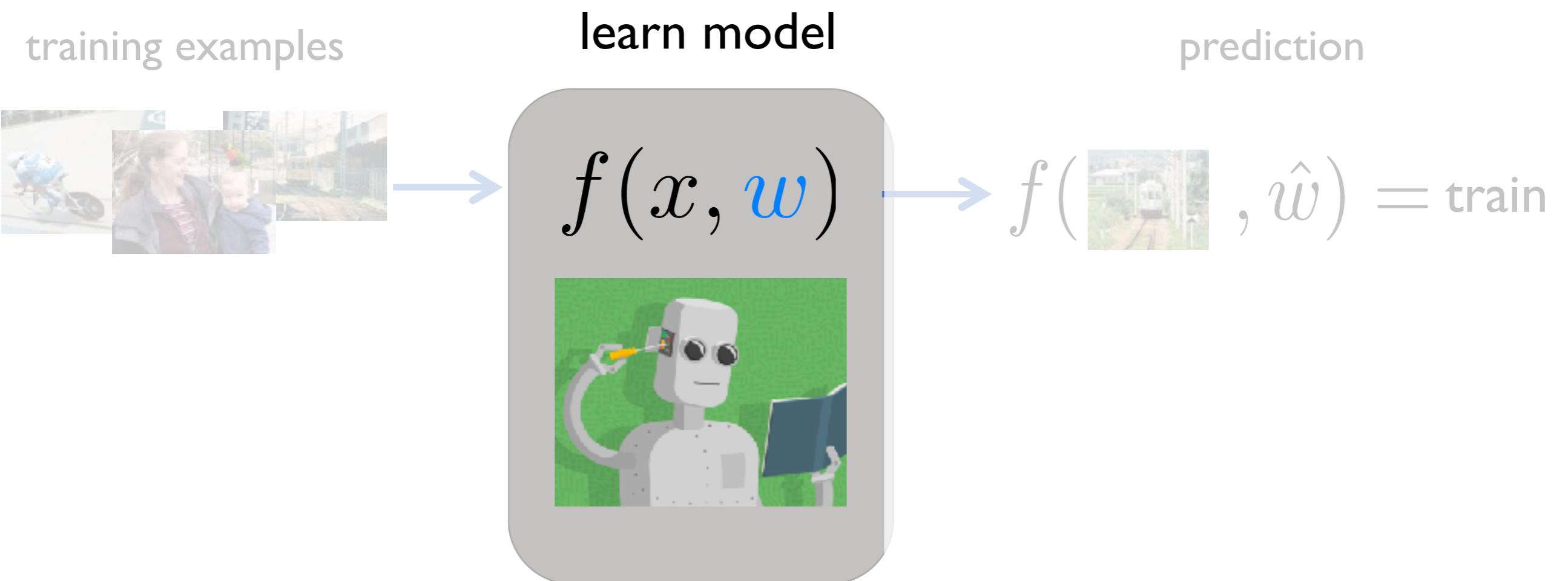


- Placing sensors
- Designing experiments



$$F(S) = \text{“information”}$$

Subset selection in Machine Learning



Variable (Coordinate) Selection

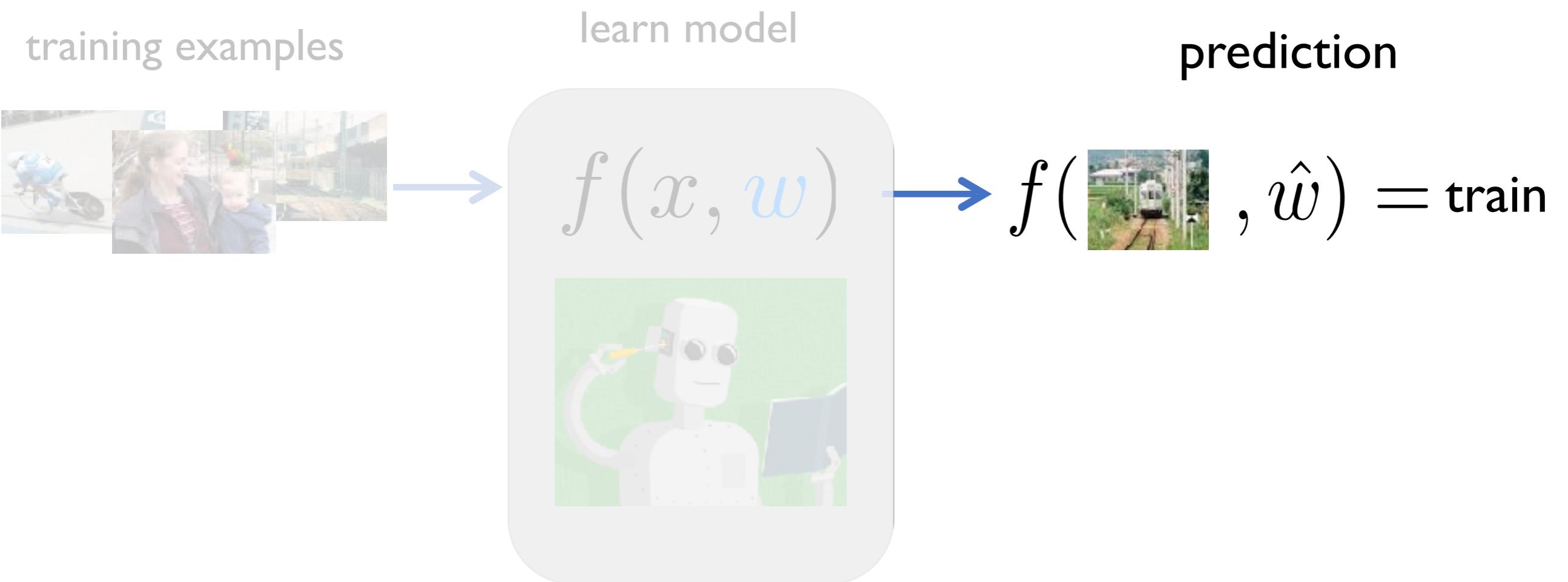


Only use few coordinates of x in $f(x, \textcolor{blue}{w})$

$$f(x, \textcolor{blue}{w}) = \sum_{i=1}^d w_i x_i$$

$F(S)$ = “coherence”

Subset selection in Machine Learning



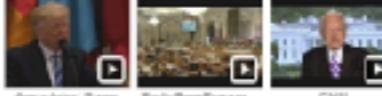
Summarization & Recommendation

News U.S. edition ▾ Modern ▾ Personalize

Top Stories

Trump Urges Muslim Leaders to Purge Their Societies of 'Foot Soldiers of Evil'
New York Times - 2 hours ago  [Share](#) [Email](#) [Print](#) [RSS](#) [PDF](#) [Twitter](#) [Facebook](#) [LinkedIn](#)
President Trump spoke about a renewed effort to stamp out extremism during a centerpiece speech to Muslim leaders in Saudi Arabia on Sunday.
Donald of Arabia: Politics
Borgen: The real reason Saudi rolled out the red carpet CNN
Live Updating: Trump calls on Muslim nations to unite in fight against terrorism - live updates CBS News

[SEE FULL COVERAGE](#)



Why Pence's frustration feels like déjà vu
CNN - 1 hour ago
Washington (CNN) A new, critical Trump administration vacancy. A vexed VP. And the domino effects of the special counsel investigation.

Dramatic Video Captures Sea Lion Dragging Little Girl Into Water From Pier
NBC News - 46 minutes ago
A cellphone video captured the nerve-wracking rescue of a young girl who was grabbed by her dress and dragged into a Canadian harbor by a sea lion on Saturday.

Sheriff David Clarke denies plagiarism, calls reporter a 'sleaze bag'
ABC News - 3 hours ago
Milwaukee County, Wis. Sheriff David Clarke speaks at the Republican National Convention in Cleveland, July 18, 2016. 0 Shares. Email

North Korea tests another missile; Seoul says dashes hopes for peace
Reuters - 4 hours ago
SEOUL North Korea fired a ballistic missile into waters off its east coast on Sunday, its second missile test in a week, which South Korea said dashed the hopes of the South's new liberal government for peace between the neighbors.

Republicans Watch Their Step in a Slow Retreat From Trump
New York Times - 1 hour ago
Senator Marco Rubio of Florida speaking to reporters on Capitol Hill this month. He and other Republicans have said they need more information about the firing of James B. Comey, the FBI.

Suggested for you ▾

Reese Witherspoon Visits Her Old Dorm Room at Stanford and Meets the Current Tenant
PEOPLE.com - 5 hours ago
Reese Witherspoon was on campus to speak at an event for the Stanford Graduate School of Business.

[Interested in Stanford University? Yes | No](#)

Personalize Google News

Suggested for you

World

U.S.

Business

Technology

Entertainment

Sports

Science

Health

Add any news topic

Examples: Astronomy, New England Patriots, White House

Adjust Sources

New York Times

Fox News

CNN

ESPN

Save Settings | Reset | Help

Recent

Spain's Socialists reelect hardliner Sanchez in leadership vote

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

<

Common formalization

Find a subset S that
maximizes / minimizes a set function $F(S)$

Can be difficult in general: exponentially many subsets!

Rescue: mathematical structure — **submodularity**

Why is submodularity interesting?

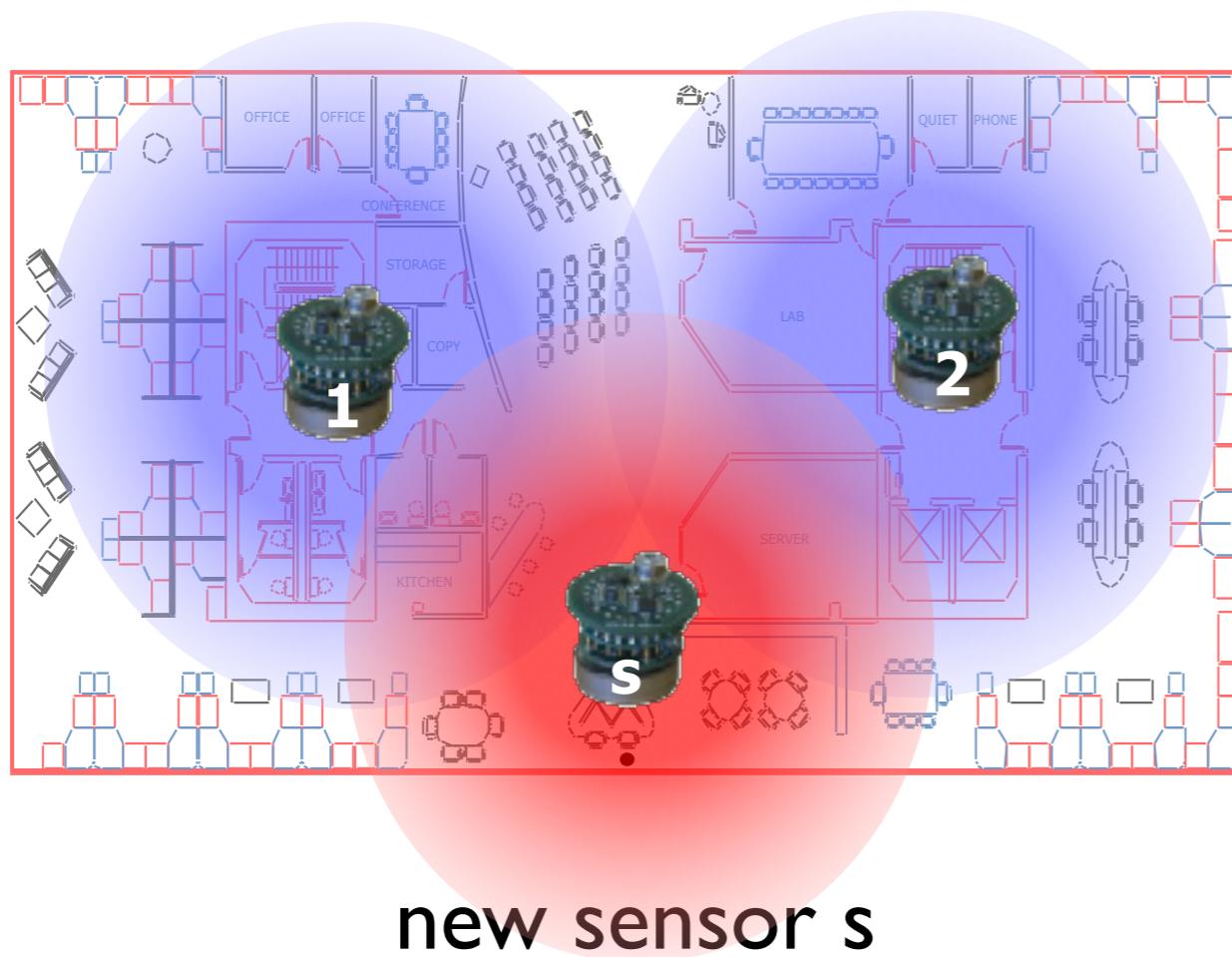
Importance of convex functions (Lovász, 1983):

- “*occur in many models* in economy, engineering and other sciences”, “often the only nontrivial property that can be stated in general”
- *preserved under many operations and transformations: larger effective range of results*
- sufficient structure for a “*mathematically beautiful and practically useful theory*”
- efficient minimization

“It is less apparent, but we claim and hope to prove to a certain extent, that a similar role is played in discrete optimization by submodular set-functions“ [...]”

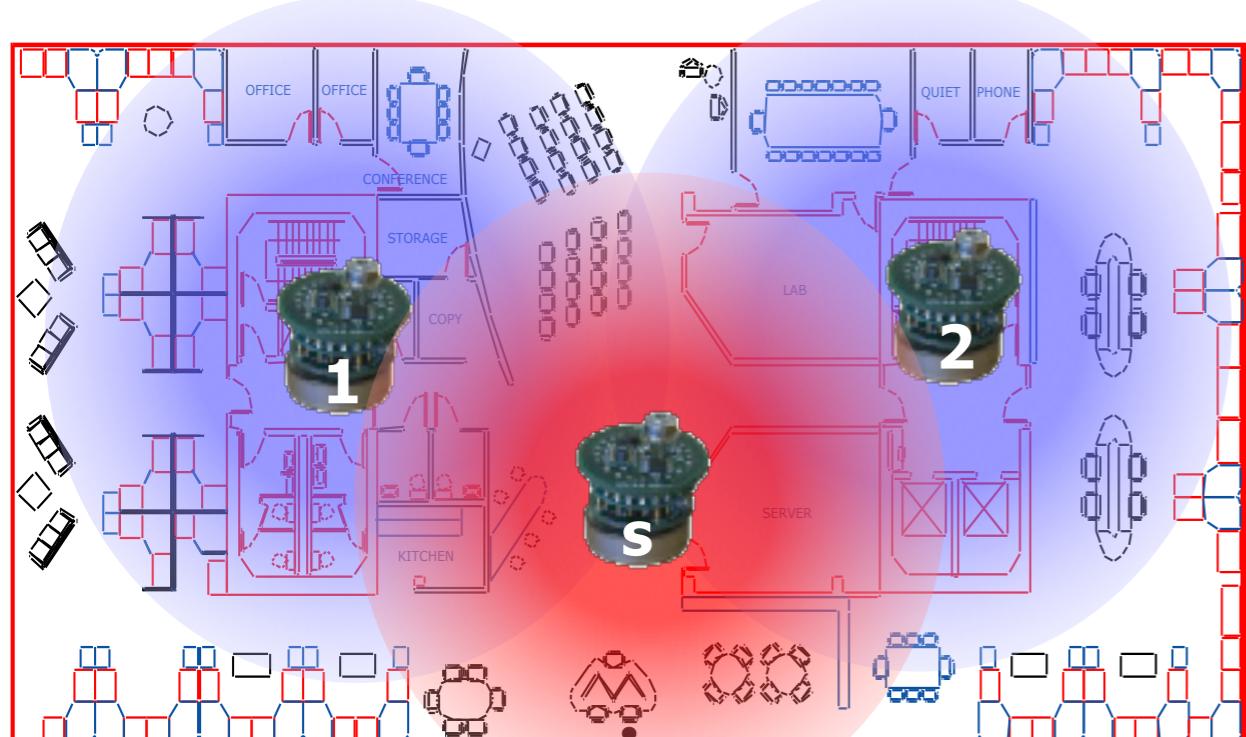
Marginal gain

- Given set function $F : 2^V \rightarrow \mathbb{R}$
- Marginal gain: $F(s|A) = F(A \cup \{s\}) - F(A)$



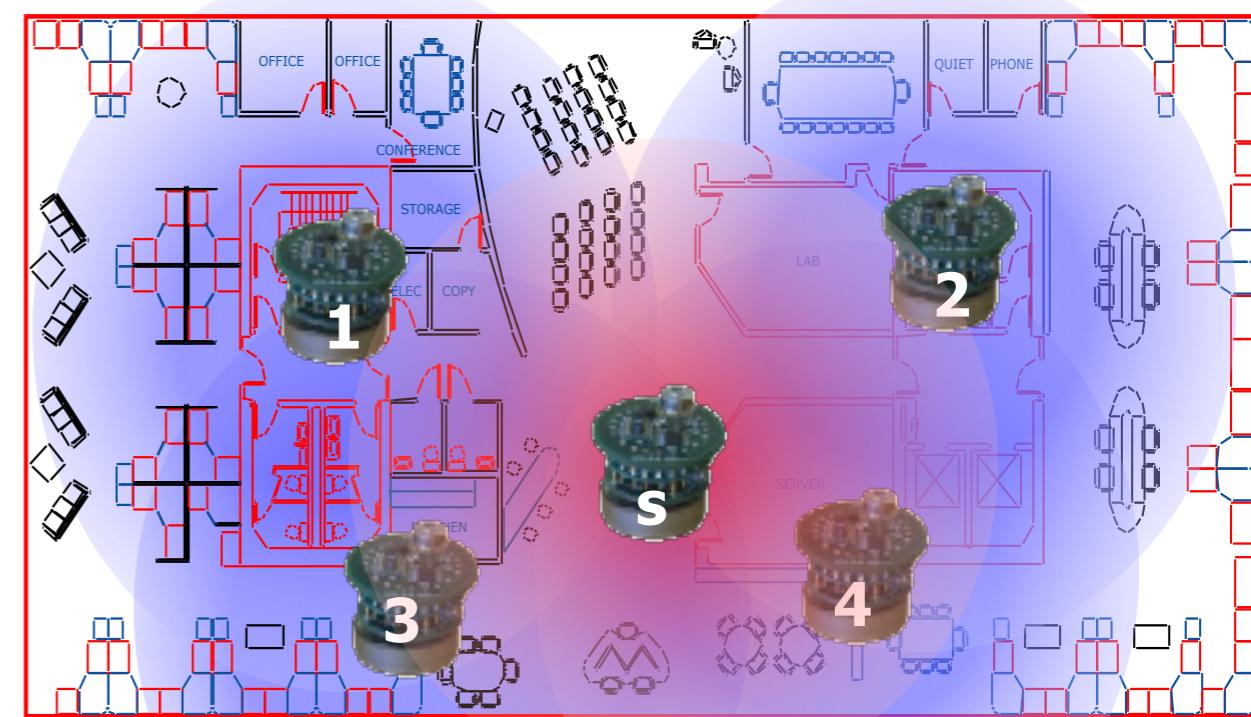
Diminishing gains

placement A = {1,2}



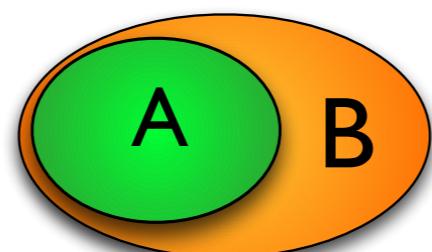
Big gain

placement B = {1,2,3,4}



small gain

for all $A \subseteq B$
and s not in B



$$F(A \cup s) - F(A) \geq F(B \cup s) - F(B)$$

Submodularity

set function: $F(S)$

- **diminishing returns:** $\forall S \subseteq T, a \notin T$

$$F(S \cup \{a\}) - F(S) \geq F(T \cup \{a\}) - F(T)$$

- **equivalent general definition:** $\forall A, B \subseteq V$

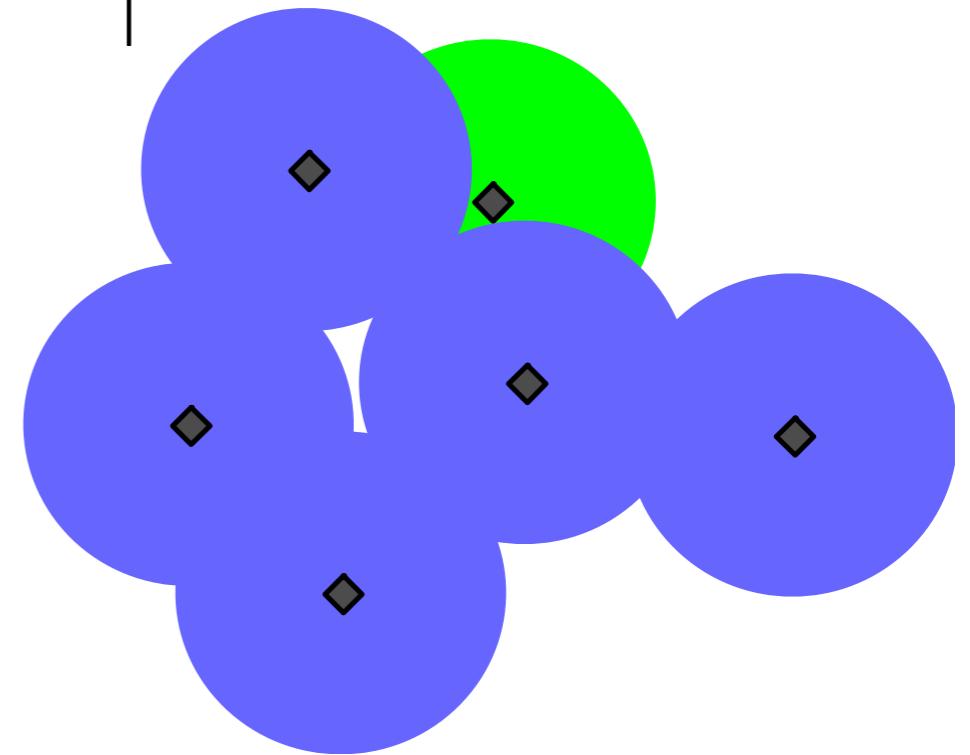
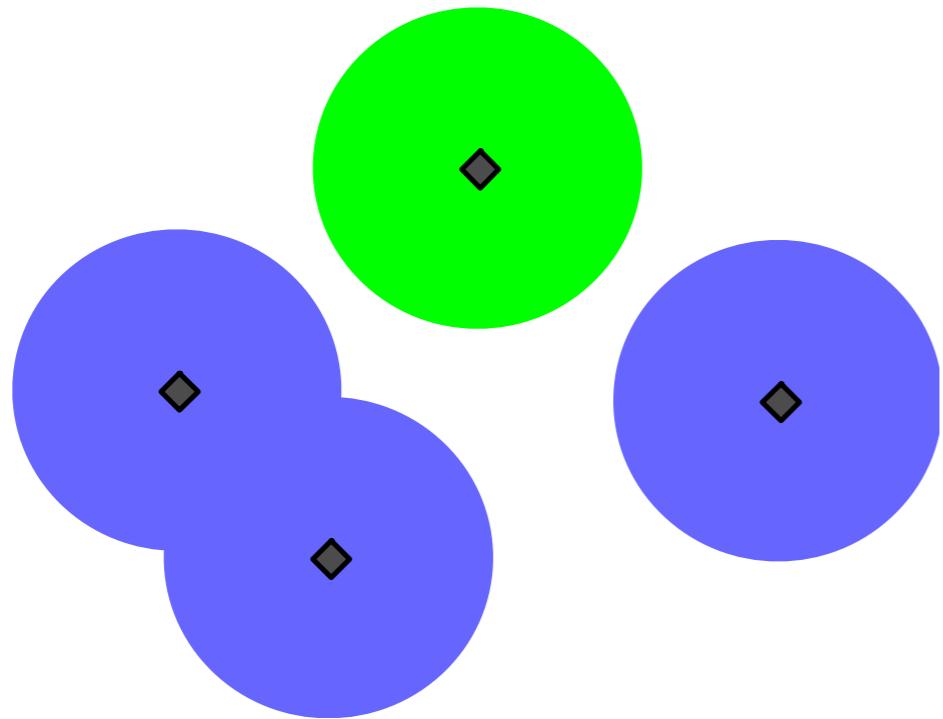
$$F(A) + F(B) \geq F(A \cup B) + F(A \cap B)$$

Roadmap

- Examples of submodular functions
- Maximizing submodular functions
- Minimizing submodular functions
- Further connections & perspectives

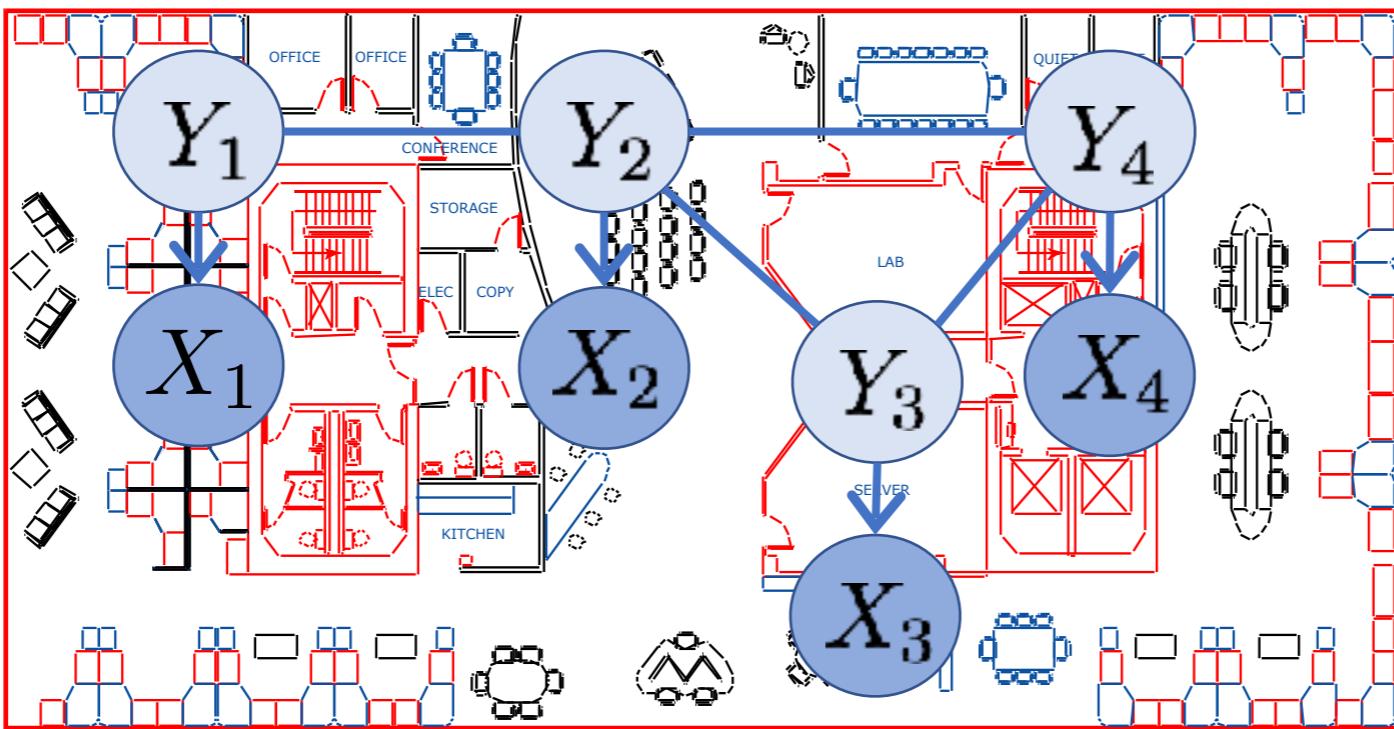
Example: cover

$$F(S) = \left| \bigcup_{v \in S} \text{area}(v) \right|$$



$$F(A \cup v) - F(A) \geq F(B \cup v) - F(B)$$

Example: sensing



- \mathcal{V} = random variables we can possibly observe
- Utility to have sensors in locations A :

$$F(A) = H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}_A) = I(\mathbf{Y}; \mathbf{X}_A)$$

*uncertainty about
temperature
before sensing*

*uncertainty about
temperature
after sensing*

Mutual
information

Recommendation & Summarization



We want:
relevance & coverage
diversity
personalization

If you bought this,
you may want
to add ...



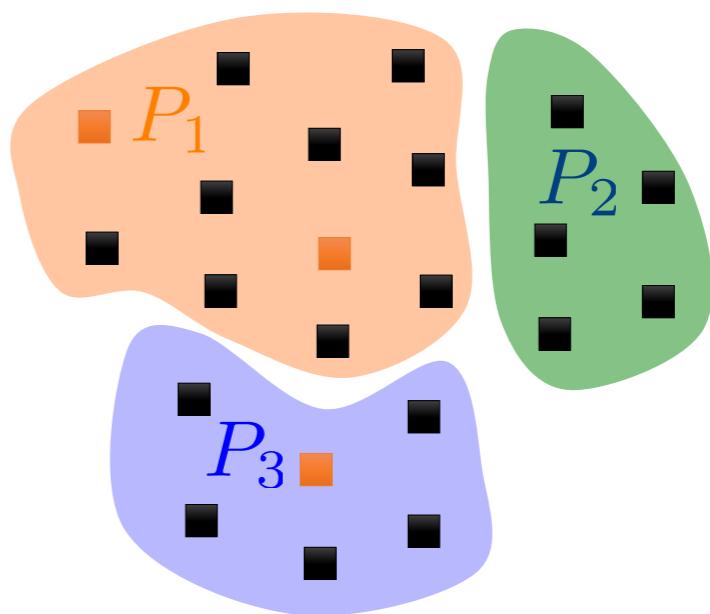
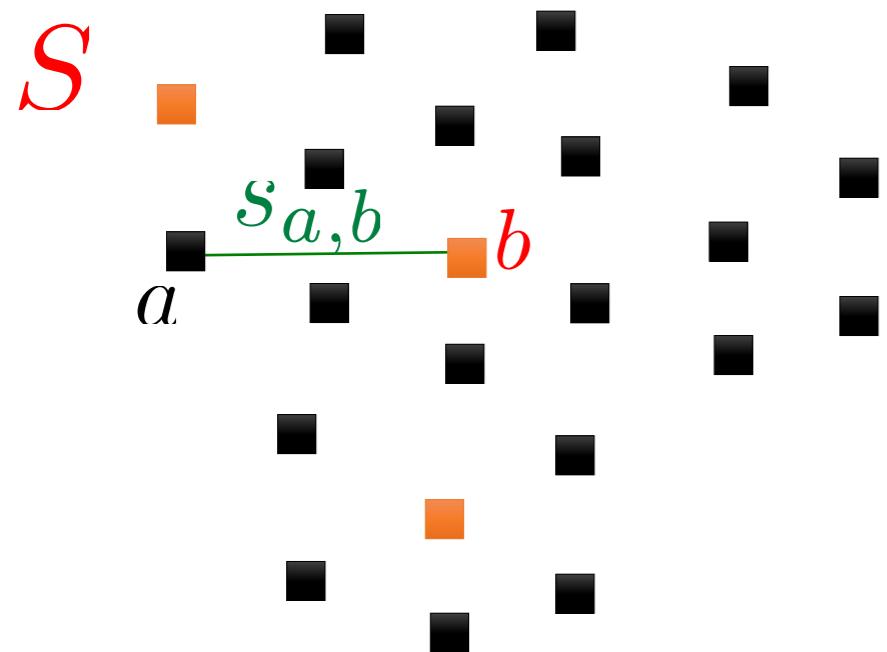
??

The image shows a screenshot of a news website's homepage. On the left, there is a sidebar with categories like "News", "Top Stories", "Suggested for you", and links for "World", "U.S.", "Business", "Technology", "Entertainment", "Sports", "Science", "Health", and "Spotlight". The main area features a grid of news stories with thumbnails and titles. Some visible titles include "Trump Urges Muslim Leaders to Purge Their S", "Why Pence's frustration feels like déjà vu", "Dramatic Video Captures Sea Lion Dragging Little Girl into Water", "Sheriff David Clarke denies plagiarism, calls reporter a 'sleaze bag'", "North Korea tests another missile; Seoul says dashes hopes for pe", "Republicans Watch Their Step in a Slow Retreat From Trump", and "Reese Witherspoon Visits Her Old Dorm Room at Stanford and Mee".

What could $F(S)$ be?

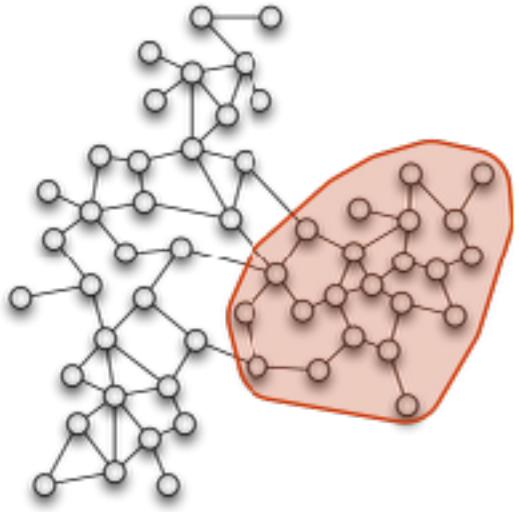
$$F(S) = \sum_{a \in \mathcal{V}} \max_{b \in S} s_{a,b}$$

$$F(S) = \sum_j \sqrt{|S \cap P_j|}$$



More in the practical!

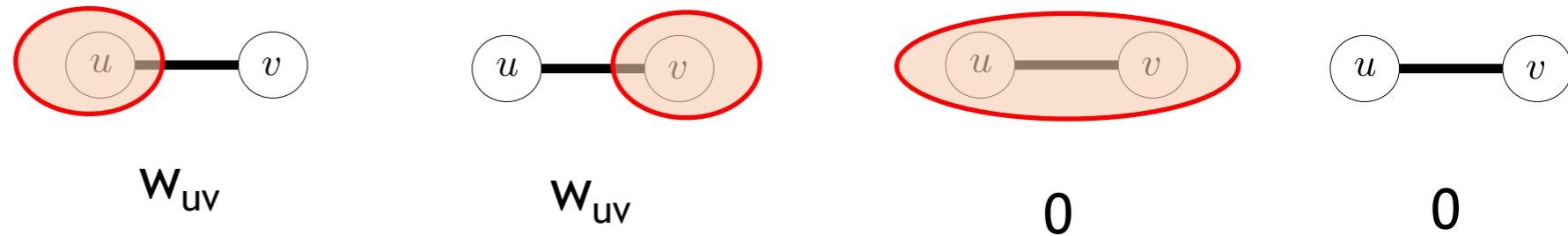
Example: graph cuts



cut for one edge:

$$F(S) = \sum_{u \in S, v \notin S} w_{uv}$$

$$\begin{aligned} F(S) + F(T) &= F(\{u\}) + F(\{v\}) \geq F(\{u, v\}) + F(\emptyset) \\ F(S \cup T) + F(S \cap T) &= F(\{u, v\}) + F(\emptyset) \end{aligned}$$



- cut of one edge is submodular!
- large graph: sum of edges

sum of submodular functions is submodular

Examples of submodular functions

- Entropy of discrete random variables
- Mutual information
- Matrix rank (as a function of columns)
- Coverage
- Spread in social networks
- Graph cuts
- ... many others!

Submodular functions (almost) everywhere!



graph
theory
(Frank 1993)

game
theory
(Shapley 1970)

matroid
theory
(Whitney, 1935)

submodular
functions

electrical
networks
*(Narayanan
1997)*

information
theory
(Fujishige 1978)

stochastic
processes
*(Macchi 1975,
Borodin 2003)*

machine
learning



Roadmap

- Examples of submodular functions
- Maximizing submodular functions
- Minimizing submodular functions
- Further connections & perspectives

Roadmap

- Maximizing submodular functions
- Minimizing submodular functions

Key Question:

Submodularity = Discrete Convexity or Discrete Concavity?

(Lovász, Fujishige, Murota, ...)

Roadmap

- Maximizing submodular functions: *discrete concavity*
- Minimizing submodular functions: *discrete convexity*

Key Question:

Submodularity = Discrete Convexity or Discrete Concavity?

(Lovász, Fujishige, Murota, ...)

Maximizing a submodular function?

$$\max_S F(S) \text{ s.t. } |S| \leq k$$

NP-hard 😞

Maximizing a submodular function?

$$\max_S F(S) \text{ s.t. } |S| \leq k$$

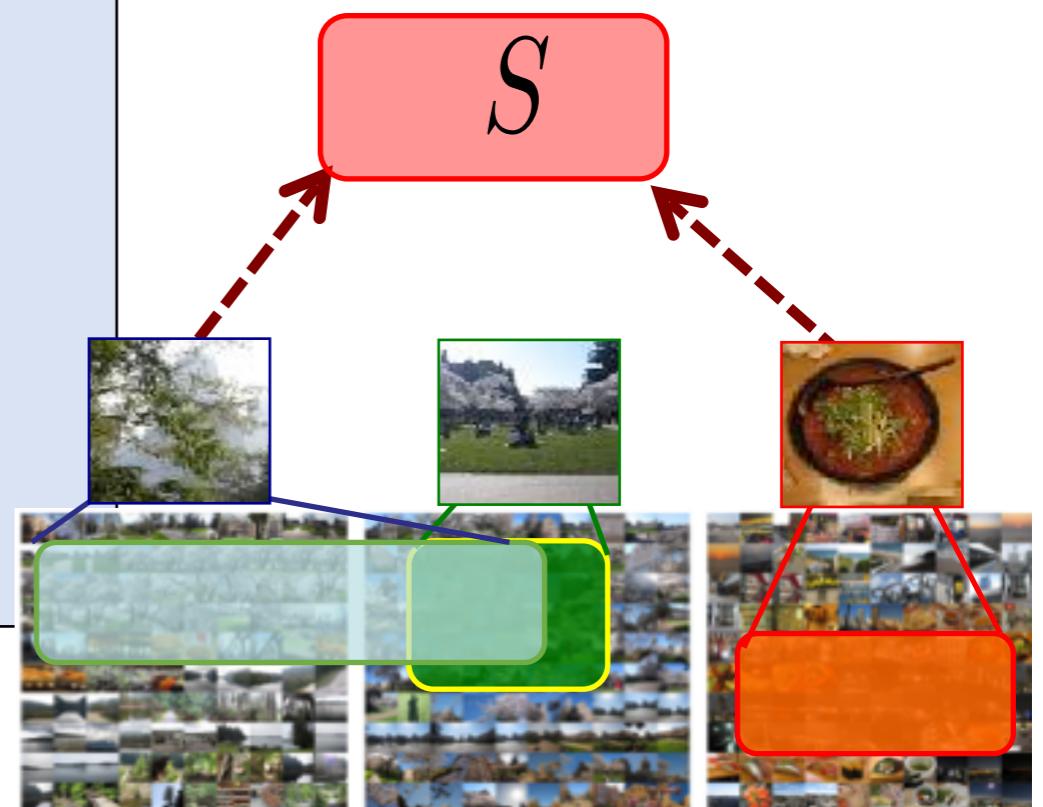
greedy algorithm:

$$S_0 = \emptyset$$

for $i = 0, \dots, k-1$

$$e^* = \arg \max_{e \in \mathcal{V} \setminus S_i} F(S_i \cup \{e\})$$

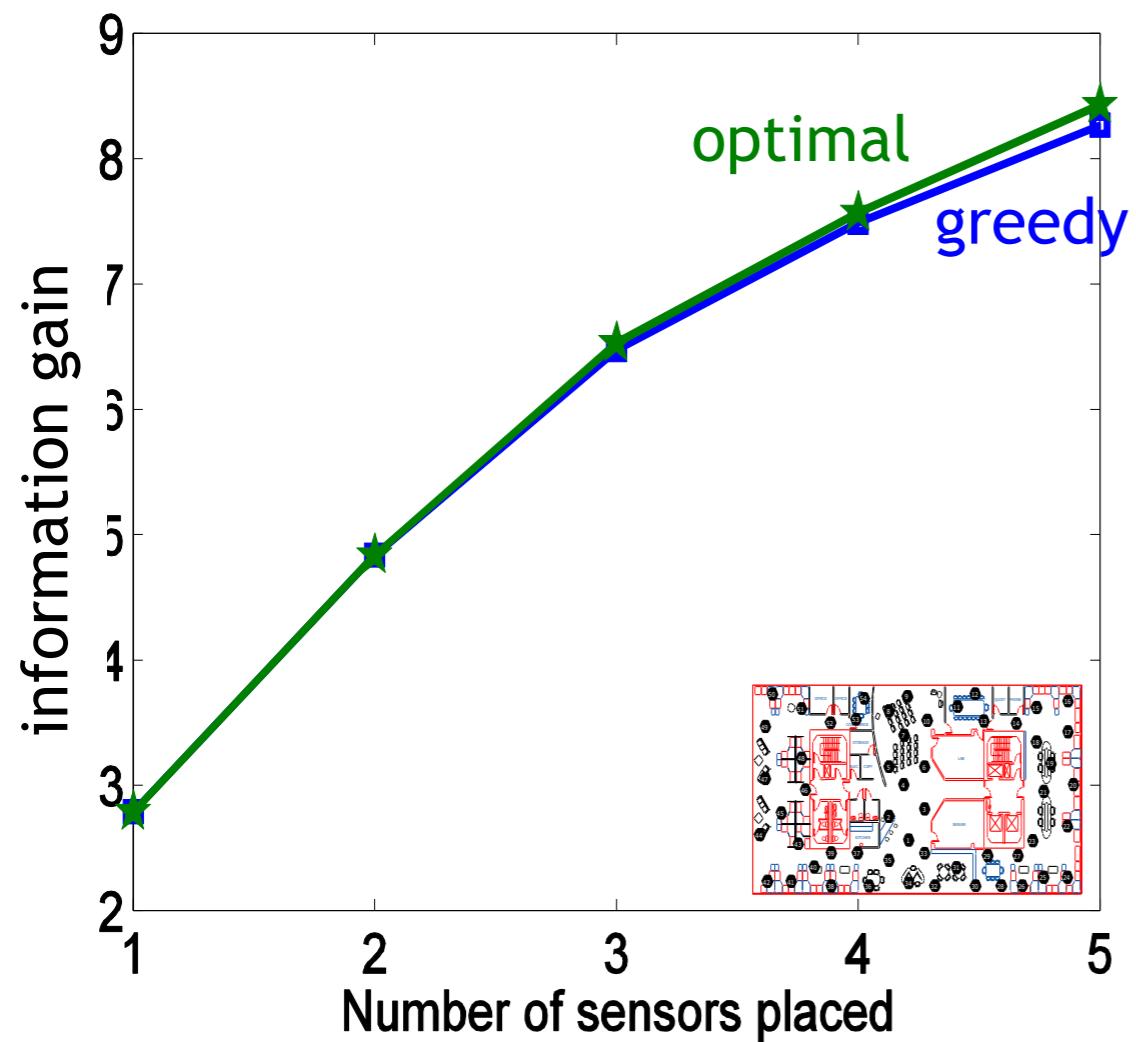
$$S_{i+1} = S_i \cup \{e^*\}$$



How “good” is S_k ?

How good is greedy?

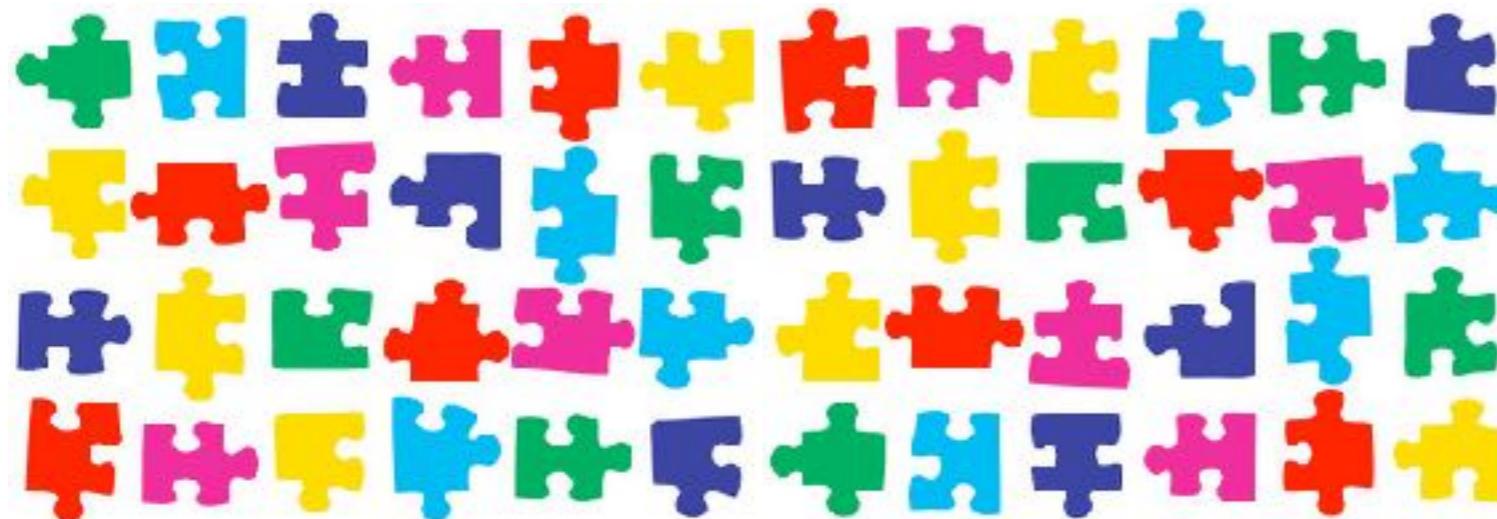
empirically:



Theorem (Nemhauser, Wolsey, Fisher 1978):
If F is monotone submodular, then
Greedy is **guaranteed** to achieve at least
63% of optimum:

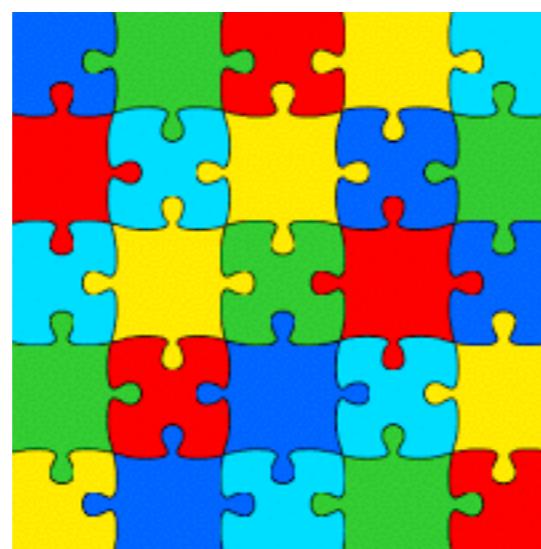
$$F(S_k) \geq \left(1 - \frac{1}{e}\right) F(S^*)$$

Greedy can fail ... without submodularity



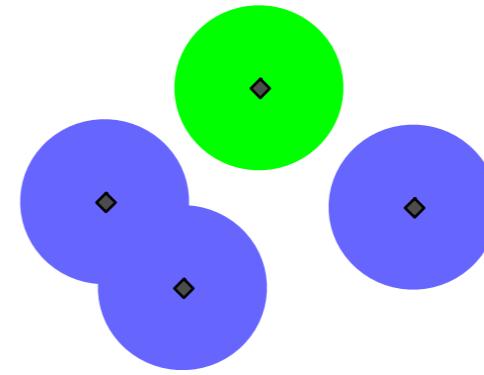
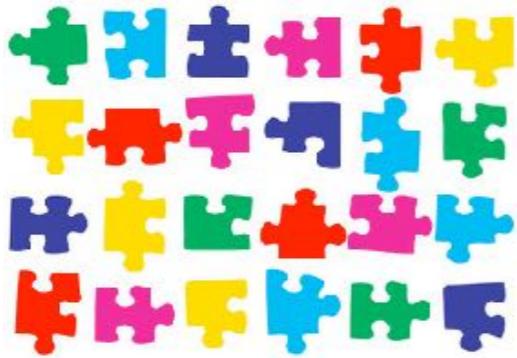
But: this **never** happens with diminishing returns! 😊

If $S =$



then $F(S) = 100$.
Otherwise, $F(S) = 0$

Recap: why does greedy work?

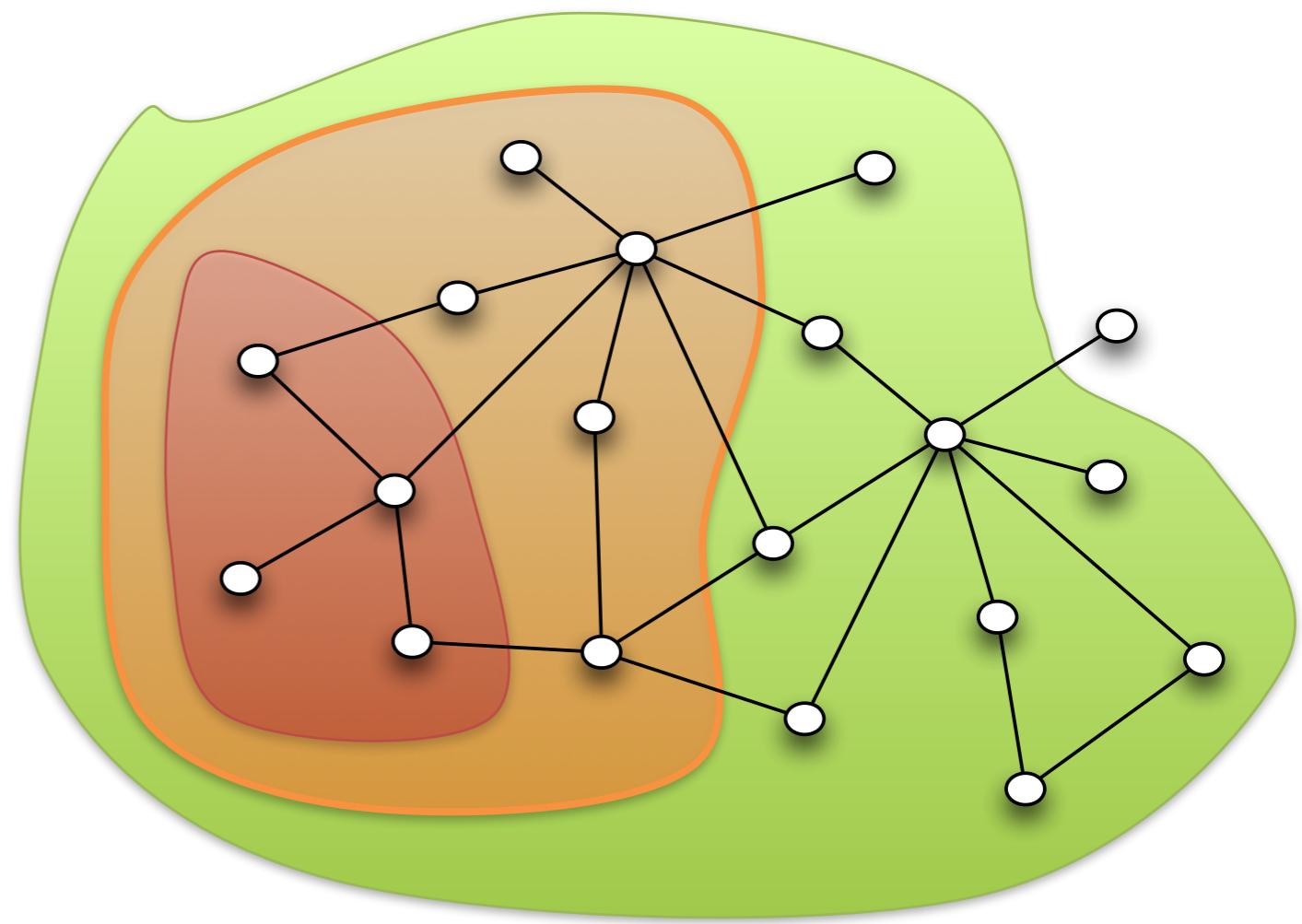
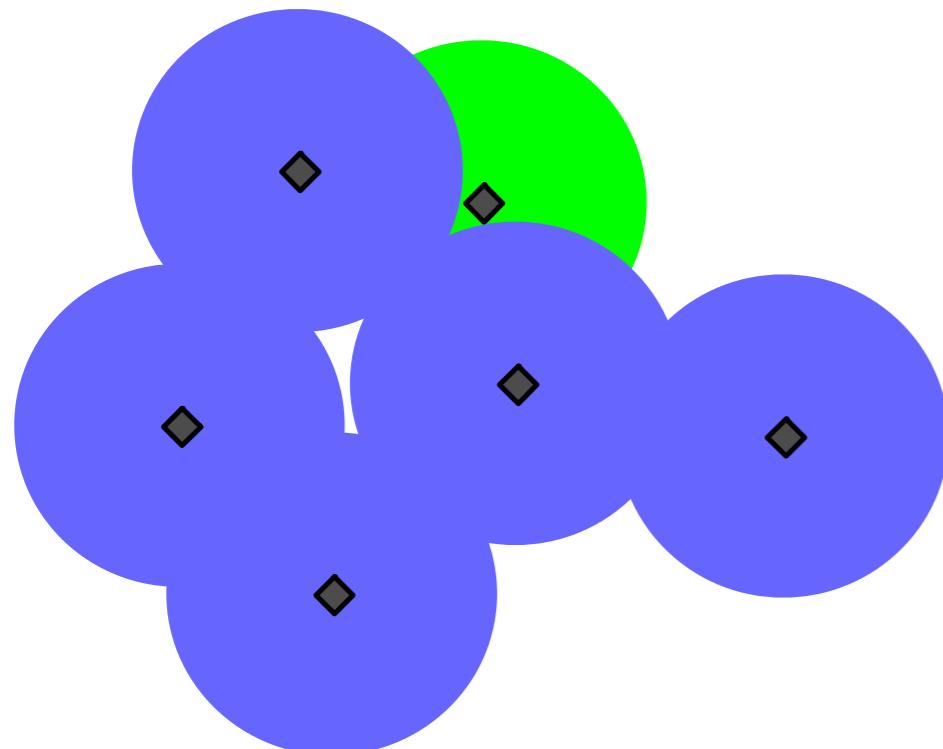


- I. Submodularity:** global information from local information
Marginal gain of single item gives information about global value

- 2. Monotonicity:** items can never harm (= reduce F)

Monotonicity

if $S \subseteq T$ then $F(S) \leq F(T)$



3

5

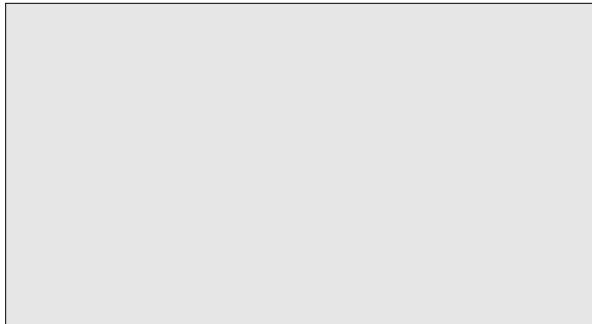
1

Beyond greedy?

- Non-monotone functions?
- Large-scale greedy?
- Other constraints?

Greedy++

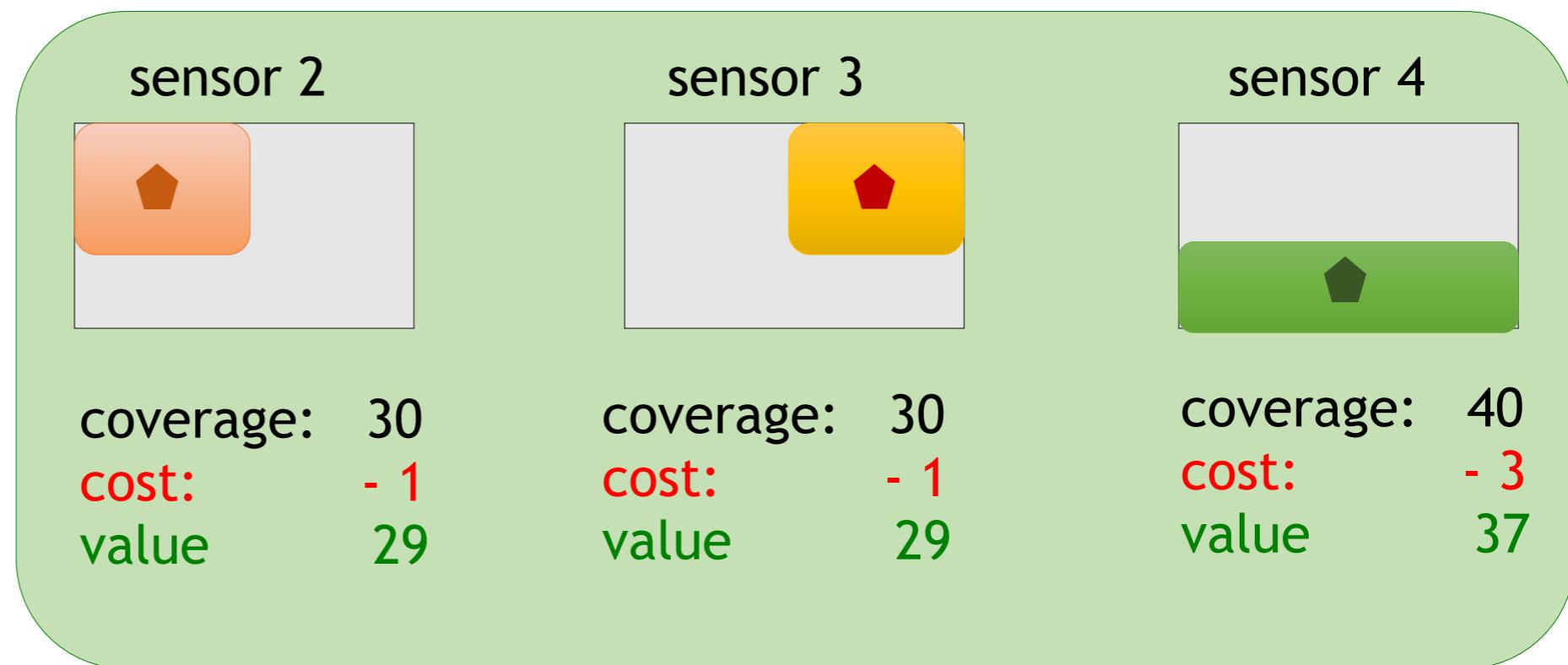
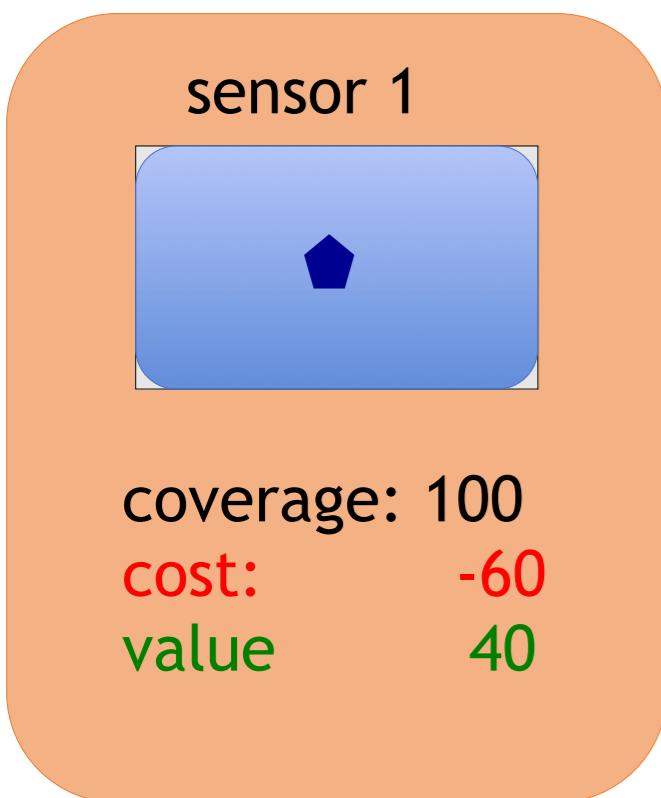
Non-monotone functions



$$F(A) = \left| \bigcup_{a \in A} \text{area}(a) \right| - \sum_{a \in A} c(a)$$

greedy solution:

$$F(A) = 40$$



Random Sampling

- Flip a coin for each item. Select with probability 1/2.
- Approximation Guarantee:

$$\mathbb{E}[F(S)] \geq \frac{1}{4} F(S^*)$$

Can we do better?

Double (bidirectional) greedy

Start: $A = \emptyset, B = \mathcal{V}$

\mathcal{V}

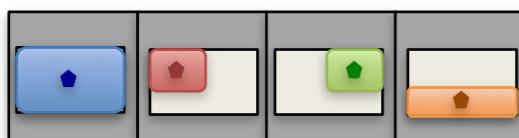


A



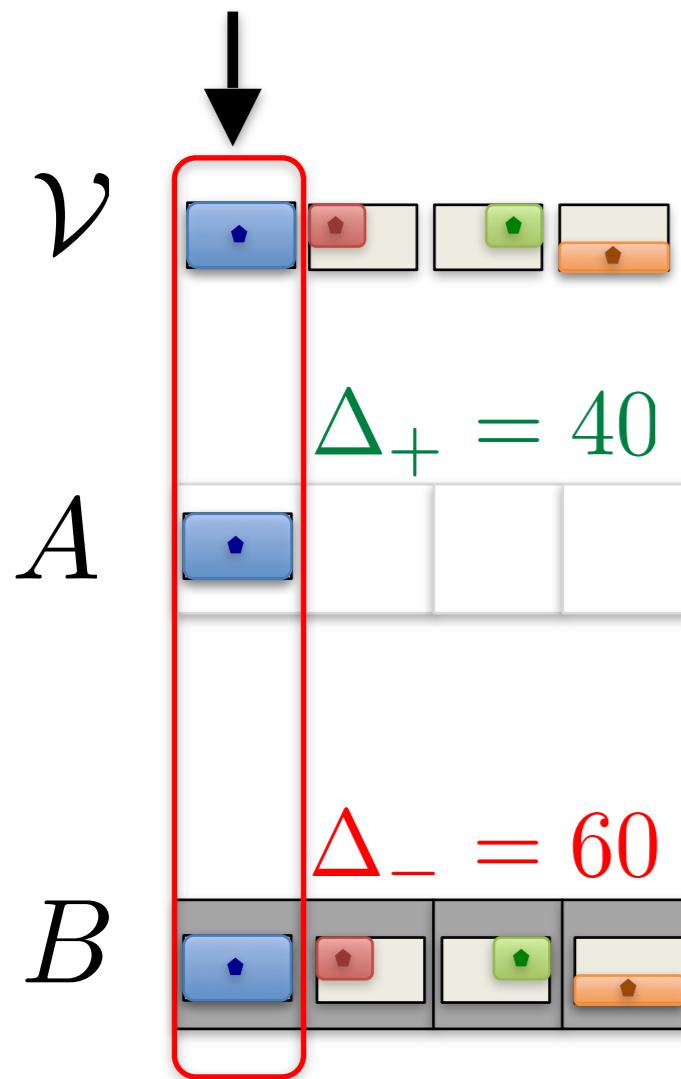
Definitely keep

B



Not (yet) decided to remove

Double (bidirectional) greedy



Start: $A = \emptyset, B = \mathcal{V}$

for $i=1, \dots, n$ //add or remove?

- gain of adding (to A):

$$\Delta_+ = [F(A \cup a_i) - F(A)]_+$$

- gain of removing (from B):

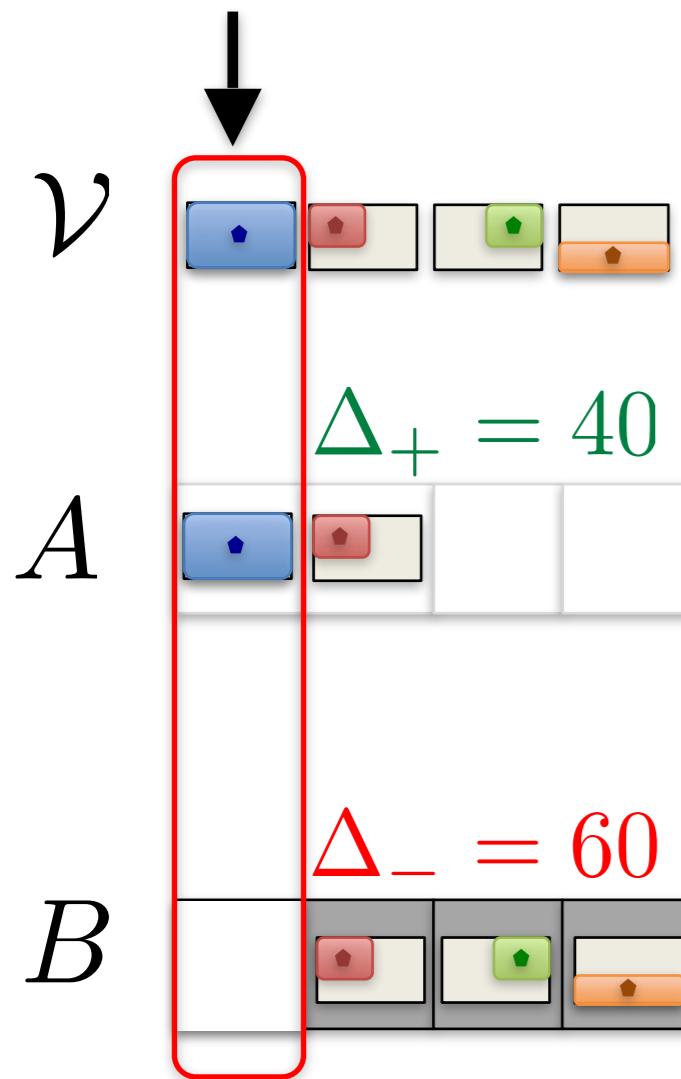
$$\Delta_- = [F(B \setminus a) - F(B)]_+$$

add with probability

coverage: 100
cost: -60

$$\mathbb{P}(\text{add}) = \frac{\Delta_+}{\Delta_+ + \Delta_-} = 40\%$$

Double (bidirectional) greedy



Start: $A = \emptyset, B = \mathcal{V}$

for $i=1, \dots, n$ //add or remove?

add with probability

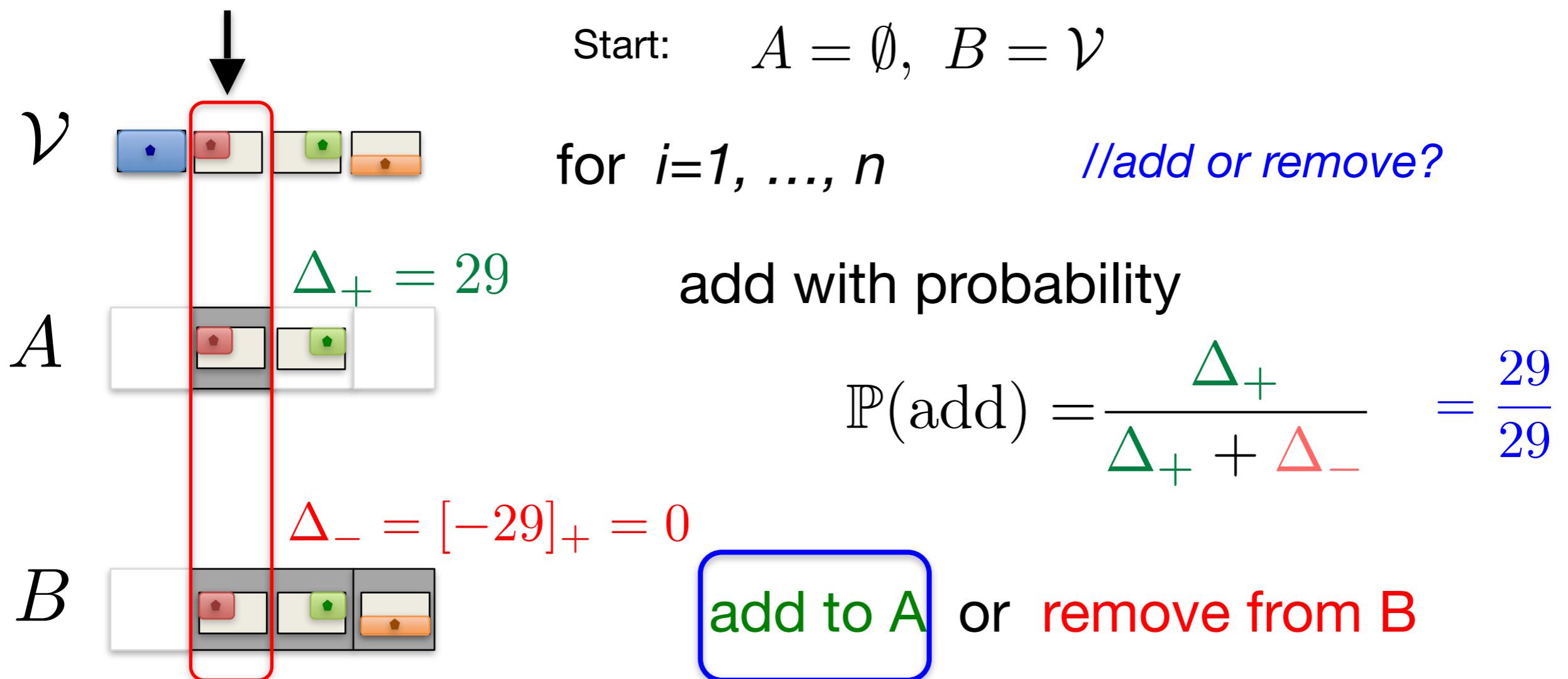
$$\mathbb{P}(\text{add}) = \frac{\Delta_+}{\Delta_+ + \Delta_-}$$

add to A or **remove from B**

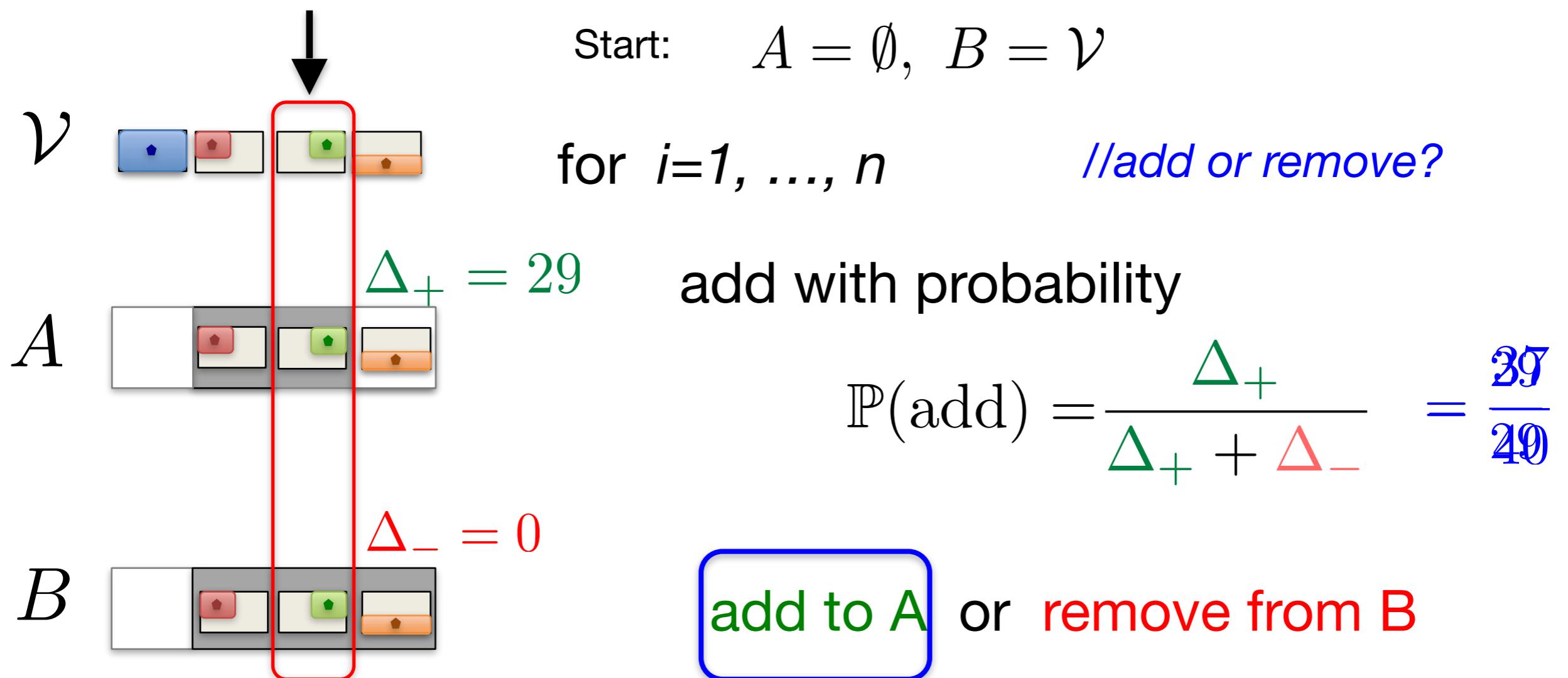


coverage: 100
cost: -60

Double (bidirectional) greedy



Double (bidirectional) greedy



Double greedy

$$\max_{S \subseteq \mathcal{V}} F(S)$$

Theorem (Buchbinder, Feldman, Naor, Schwartz '12)

F submodular, S_g solution of double greedy. Then

$$\mathbb{E}[F(S_g)] \geq \frac{1}{2} F(S^*)$$

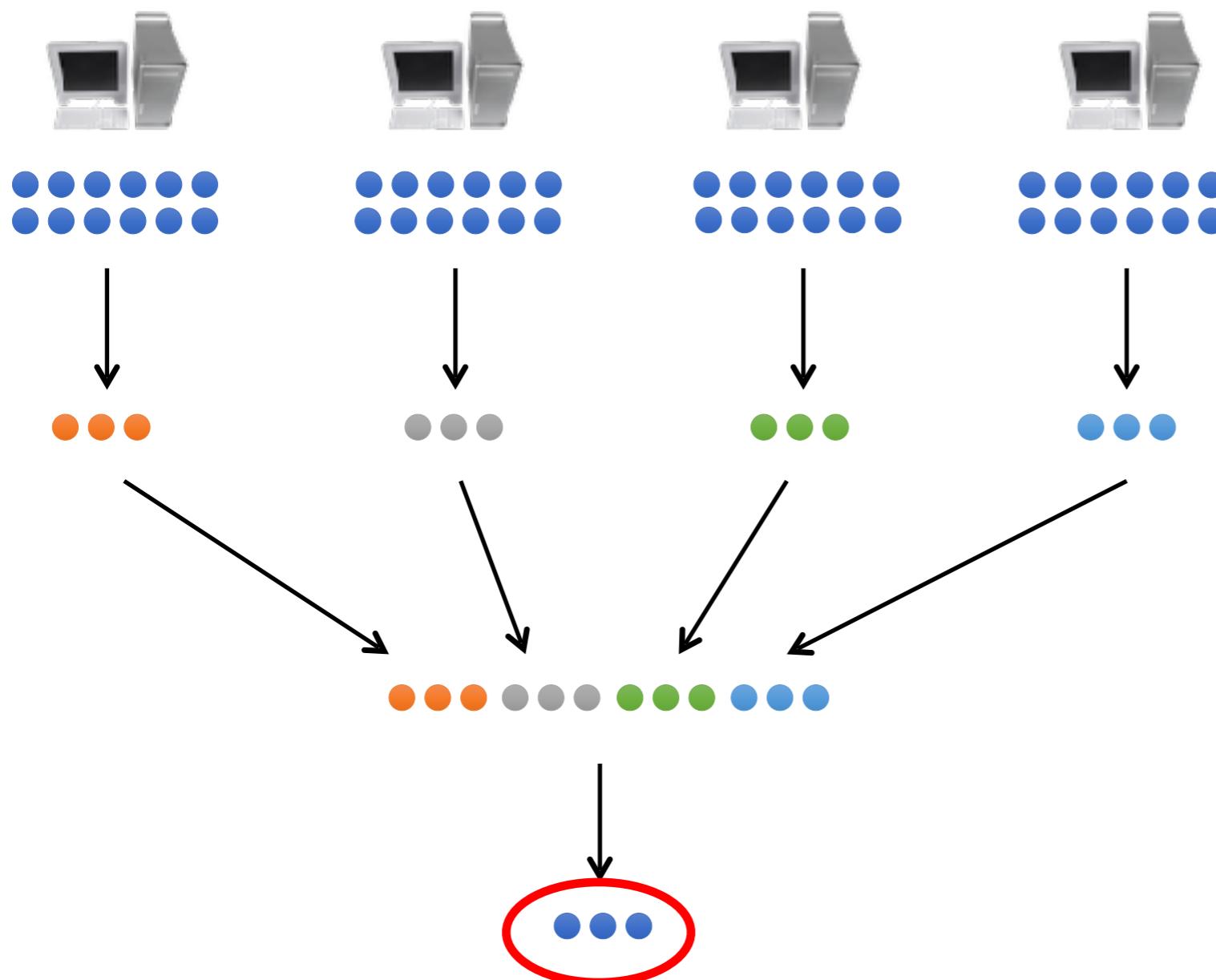
optimal solution

Beyond greedy?

- Non-monotone functions?
Double greedy (or repeated greedy)
- Large-scale greedy?
- Other constraints?

Greedy++

Distributed greedy algorithms



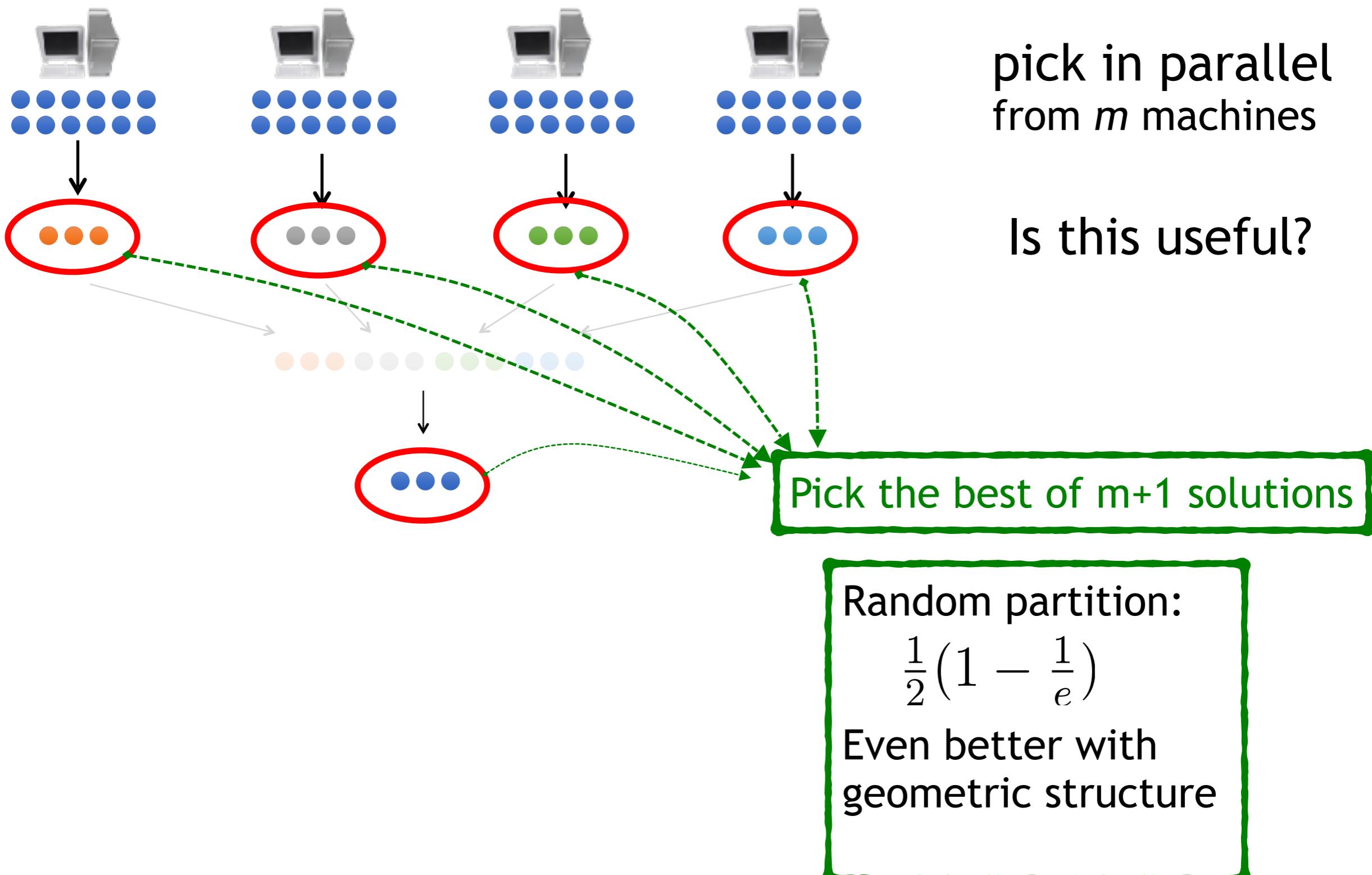
greedy is **sequential**.
pick in parallel??

pick k elements
on each machine.

combine and run
greedy again.

Is this useful?

Distributed greedy algorithms



More scalability

- Streaming algorithms
- Stochastic greedy
- **Adaptivity** (*Balkanski-Singer 2017, Balkanski-Rubinstein-Singer 2018, ...*)
 - # rounds of parallel queries
 - needs **adaptive sampling**: e.g.
up-sample growing set or down-sample increasing set,
distribution evolves over rounds

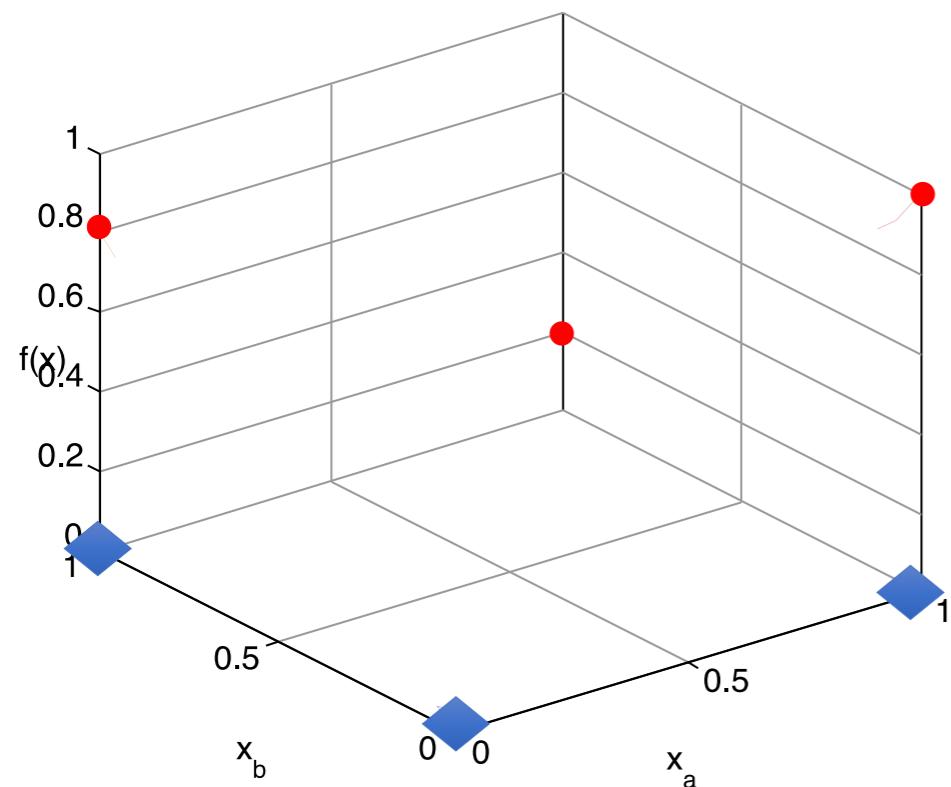
Beyond greedy?

- Non-monotone functions?
Double greedy (or repeated greedy)
- Large-scale greedy?
Distributed, streaming, stochastic,...
- Other constraints?

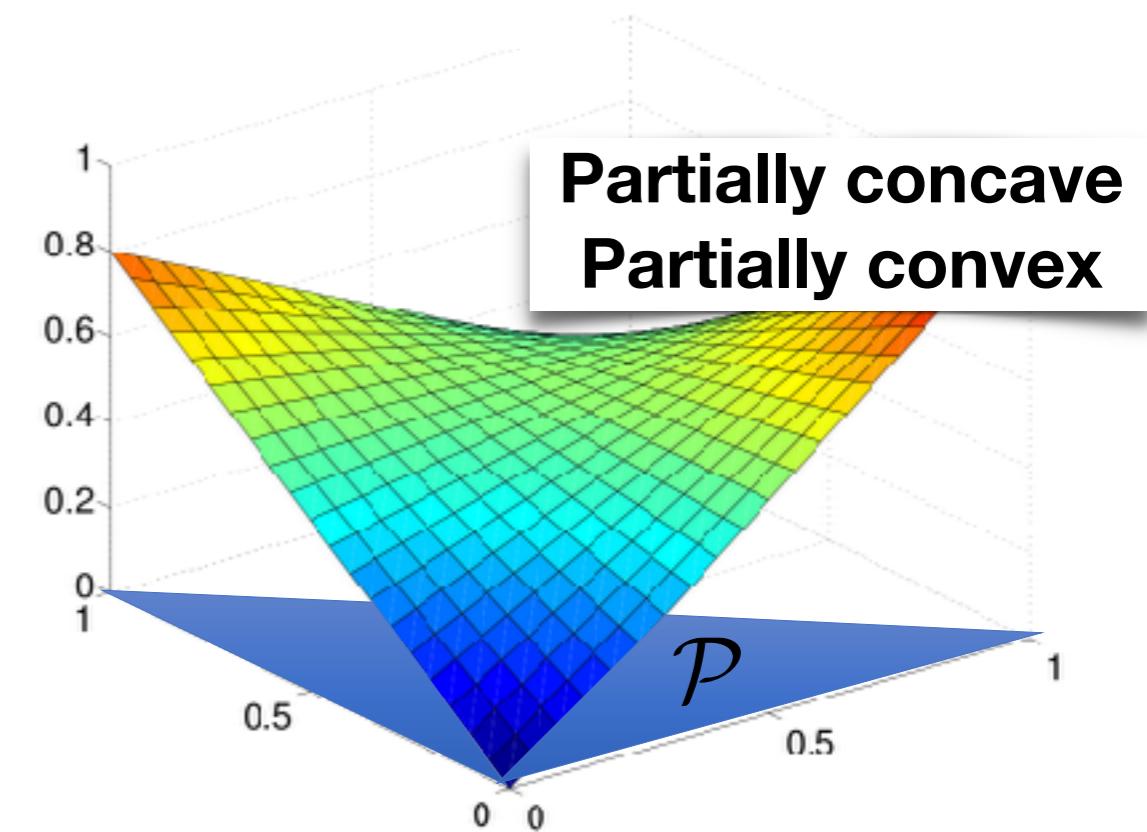
Continuous Optimization!

Relax: Discrete to continuous

$$\max F(S)$$



$$\max f_M(x)$$



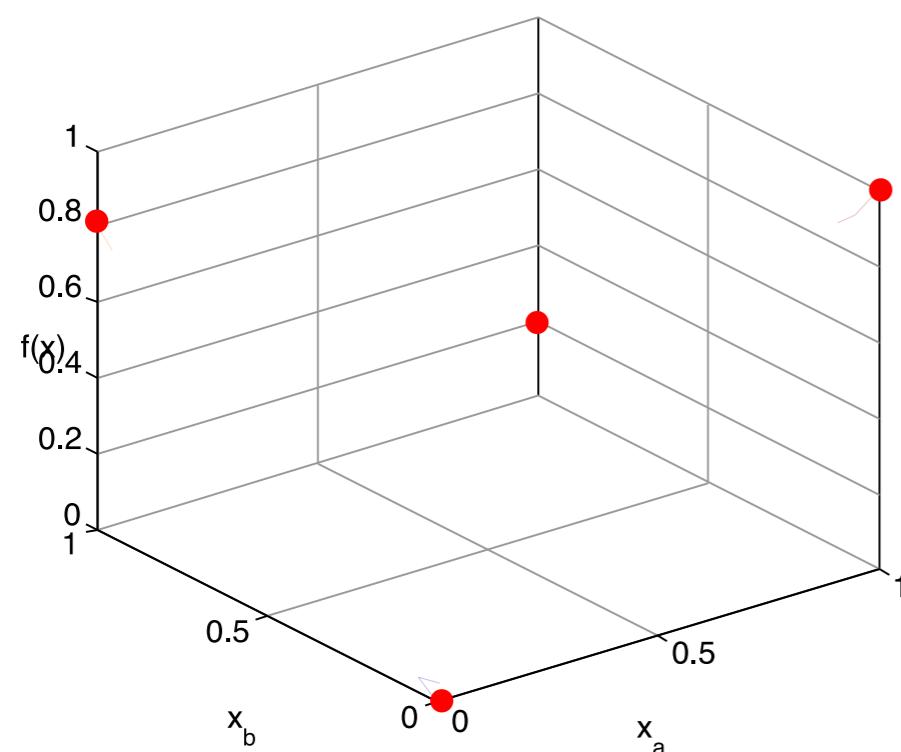
Algorithm: “continuous greedy”

1. approximately maximize f_M over $\mathcal{P} = \text{conv}(\mathcal{I})$
2. round to discrete set

Between discrete & continuous optimization

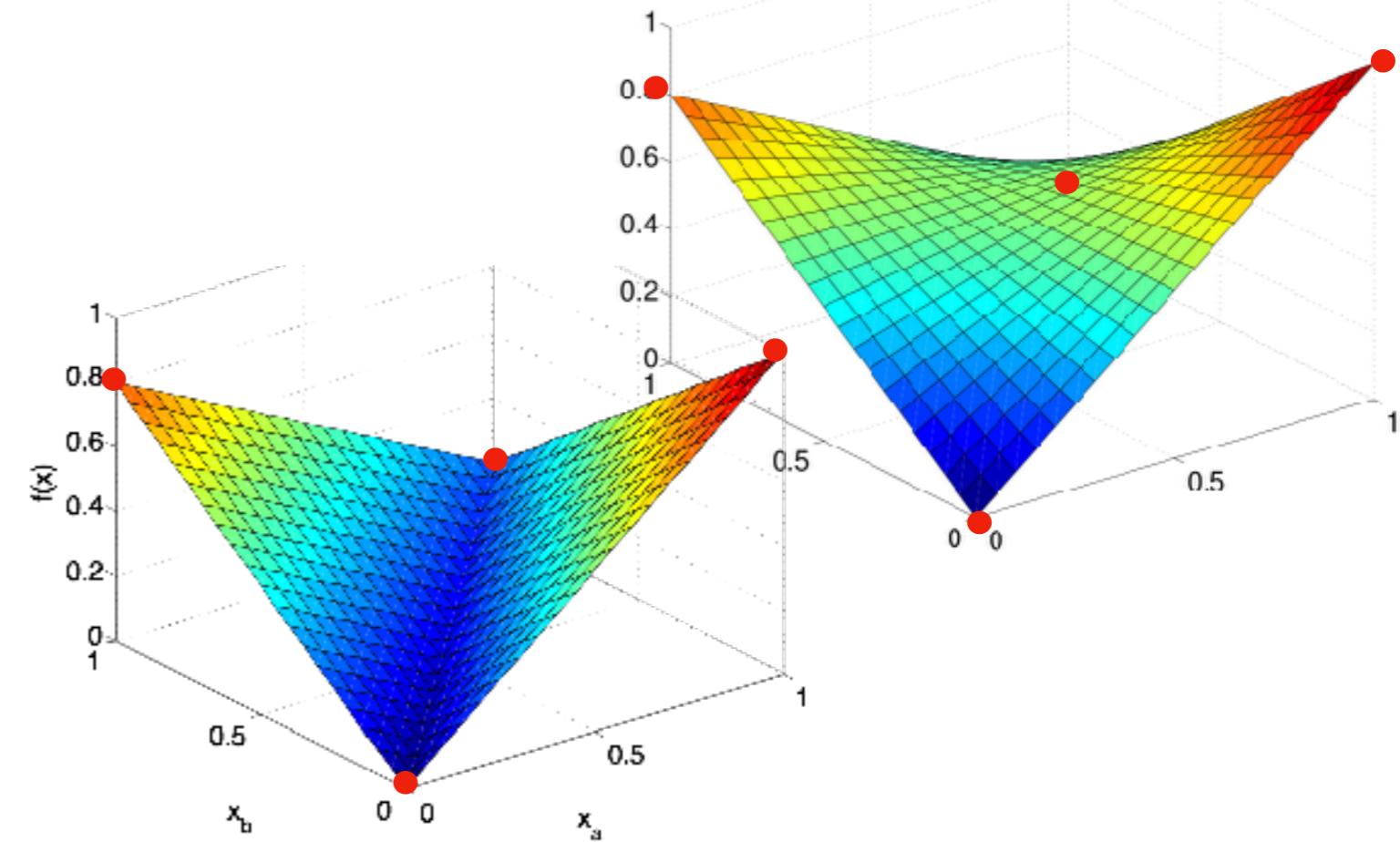
nonlinear extension/optimization

$$F : \{0, 1\}^n \rightarrow \mathbb{R}$$



$$\min_{x \in \mathcal{C} \subseteq \{0, 1\}^n} F(x)$$

$$f : [0, 1]^n \rightarrow \mathbb{R}$$



$$\min_{z \in \text{conv}(\mathcal{C}) \subseteq [0, 1]^n} f(z)$$

Generic construction

$$F : \{0, 1\}^n \rightarrow \mathbb{R} \quad \longrightarrow \quad f : [0, 1]^n \rightarrow \mathbb{R}$$

discrete set:
 $T = \{a, d\}$

1	a
0	b
0	c
1	d

a	.5
b	.5
c	0
d	.8

continuous z

- Define **probability measure** over subsets (joint over coordinates) such that **marginals agree with z**:

$$\mathbb{P}(i \in S) = z_i$$

$$f(z) = \mathbb{E}[F(S)]$$

- Extension:

$$f(z) = F(z)$$

- for discrete z:

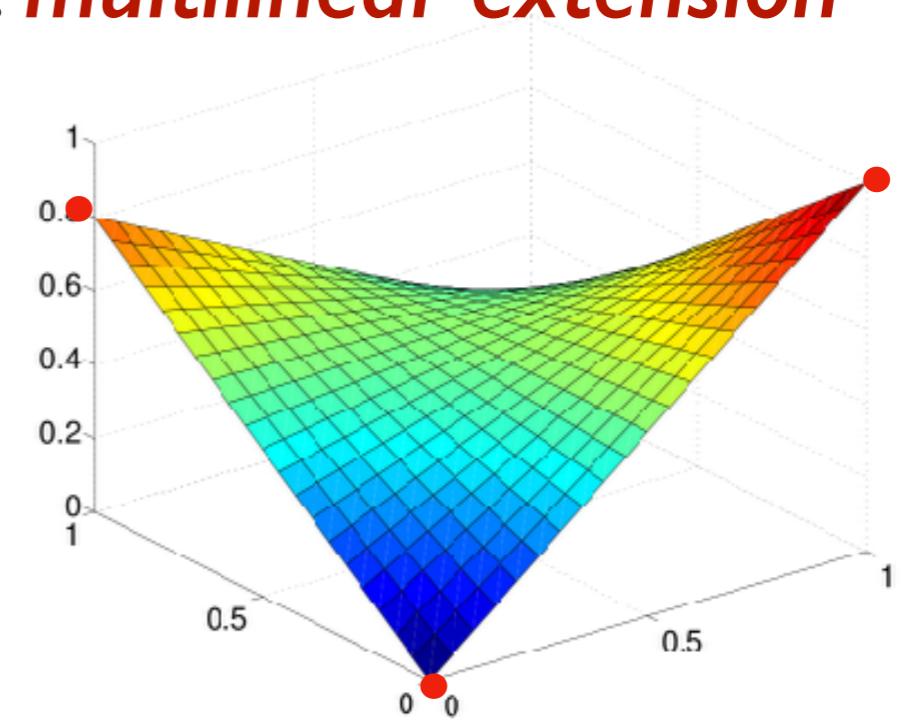
Independent coordinates

$$f(z) = \mathbb{E}[F(S)]$$

$$P(S) = \prod_{i \in S} z_i \cdot \prod_{j \notin S} (1 - z_j)$$

a	.5
b	.5
c	0
d	.8

- $f(z)$ is a multilinear polynomial: **multilinear extension**
- neither convex nor concave...
but enough for optimization



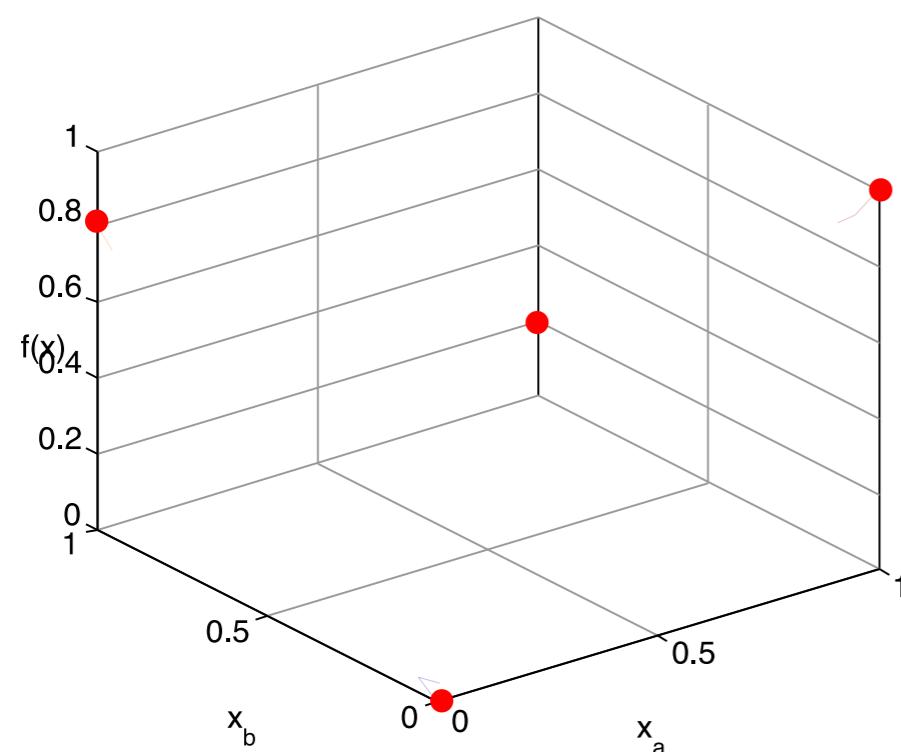
Roadmap

- Examples of submodular functions
- Maximizing submodular functions
 - greedy
 - double greedy
 - continuous greedy
- Minimizing submodular functions
- Further connections & perspectives

Between discrete & continuous optimization

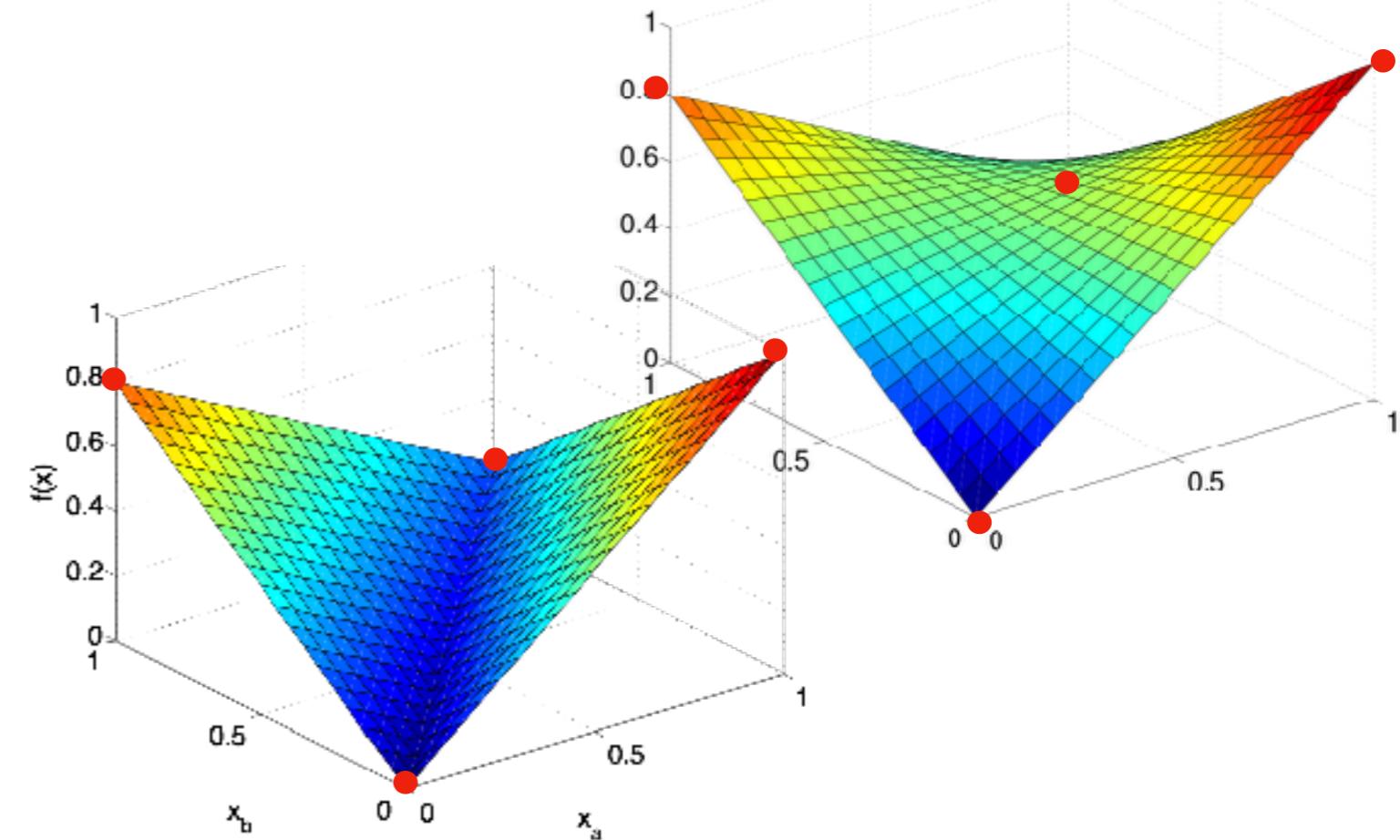
nonlinear extension/optimization

$$F : \{0, 1\}^n \rightarrow \mathbb{R}$$



$$\min_{x \in \mathcal{C} \subseteq \{0,1\}^n} F(x)$$

$$f : [0, 1]^n \rightarrow \mathbb{R}$$



$$\min_{z \in \text{conv}(\mathcal{C}) \subseteq [0,1]^n} f(z)$$

Lovász extension

$$f(z) = \mathbb{E}[F(S)]$$

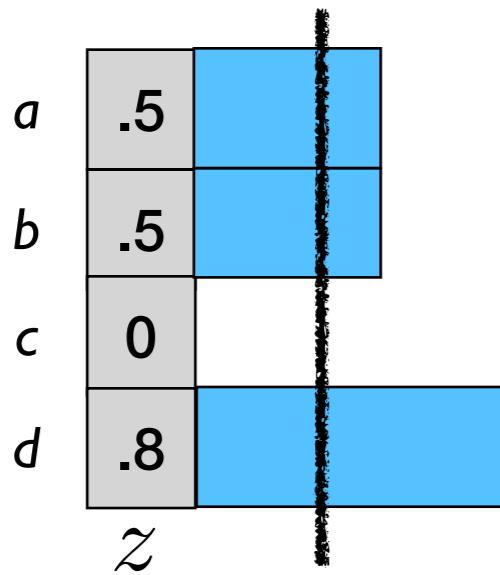
$$\mathbb{P}(i \in S) = z_i$$

Lovász extension

$$f(z) = \mathbb{E}[F(S)]$$

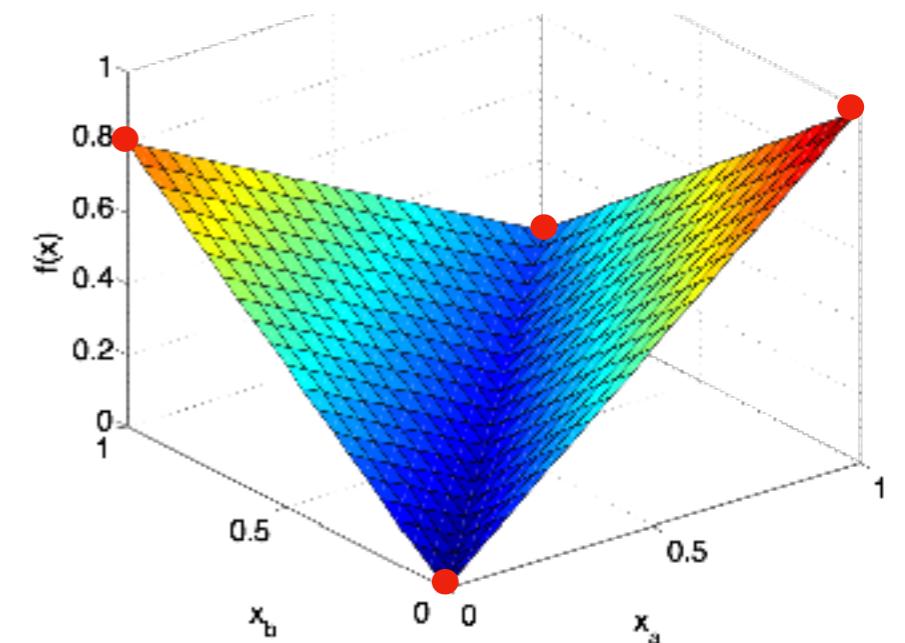
$$\mathbb{P}(i \in S) = z_i$$

- “coupled” distribution defined by **level sets**



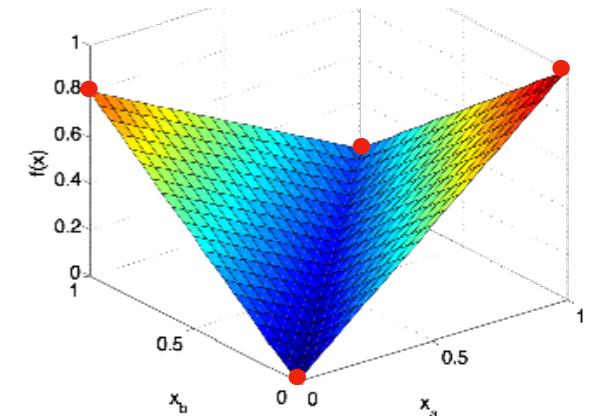
$$S_0 = \{\}, S_1 = \{d\}, S_2 = \{a, b, d\}, \\ S_3 = \{a, b, c, d\}$$

Theorem (Lovász 1983)
 $f(z)$ is convex iff $F(S)$ is submodular.



Implications of the Lovász extension

$$f(z) = \mathbb{E}[F(S)]$$



- piecewise linear & convex
- can compute **subgradient** of $f(z)$ in $O(n \log n)$
→ *convex optimization*
- **rounding**: use one of the level sets of z^*

a	.5
b	.5
c	0
d	.8

z

exact convex relaxation!

$$\min_{S \subseteq V} F(S) = \min_{z \in [0,1]^n} f(z)$$

Submodular minimization: a brief overview

convex optimization

- **ellipsoid method** (Grötschel-Lovász-Schrijver 81)
- **subgradient method** (improved: Chakrabarty-Lee-Sidford-Wong 16)

$$\min_{z \in [0,1]^n} f(z)$$

combinatorial optimization

- **network flow based** (Schrijver 00, Iwata-Fleischer-Fujishige-01)
 $O(n^4T + n^5 \log M)$ (Iwata 03), $O(n^6 + n^5T)$ (Orlin 09)

convex + combinatorial

- **cutting planes** (Lee-Sidford-Wong 15)

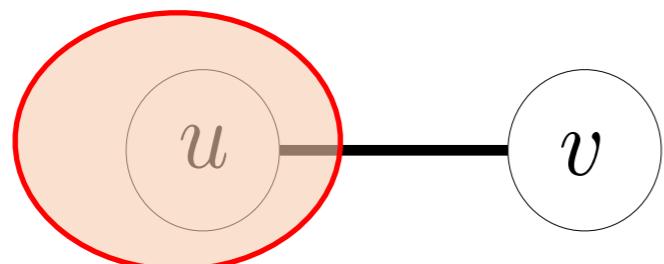
$$O(n^2T \log nM + n^3 \log^c nM) \quad O(n^3T \log^2 n + n^4 \log^c n)$$

Examples of Lovasz extensions

1. $F(S) = \min\{|S|, 1\}$

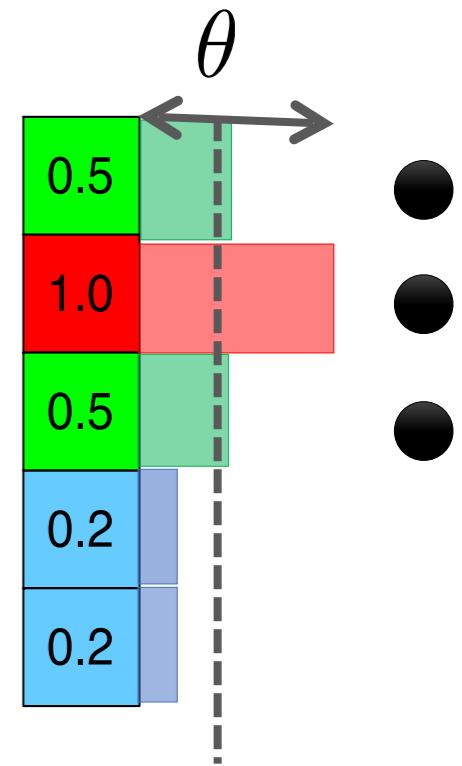
$$f(x) = \max_i x_i$$

2. Cut function: 2 items (nodes)



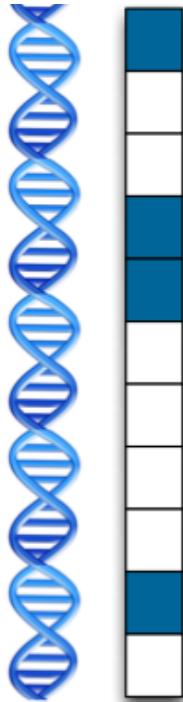
$$F(S) = \begin{cases} 1 & \text{if } |S| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$f(x) = |x_u - x_v|$$



Applications

- structured sparsity (*Bach 10*)



Only use few coordinates of x in $f(x, w)$

Convex regularizer from submodular function!

- decomposition & parallel algorithms
(Komodakis-Paragios-Tziritas 11, Stobbe-Krause 10, Jegelka-Bach-Sra 13,
Nishihara-Jegelka-Jordan 14, Ene-Nguyen 15)
- variational inference (*Djolonga-Krause 14*)

How far does relaxation go?

- **strongly convex** version:

$$\min_{z \in [0,1]^n} f(z) \quad \dashrightarrow \quad \min_{z \in \mathbb{R}^n} f(z) + \frac{1}{2} \|z\|^2$$

\updownarrow

dual: $\min_{s \in \mathcal{B}_F} \frac{1}{2} \|s\|^2$

- Fujishige-Wolfe / minimum-norm point algorithm
- actually solves *parametric submodular minimization*
- **But:** no relaxation is tight for **constrained minimization**
typically hard to approximate

Roadmap

- Examples of submodular functions
 - Maximizing submodular functions: *discrete concavity*
 - Minimizing submodular functions: *discrete convexity*
- Further connections & perspectives
 - nonconvex optimization
 - discrete probability

Submodularity beyond sets

- sets: for all subsets $A, B \subseteq V$

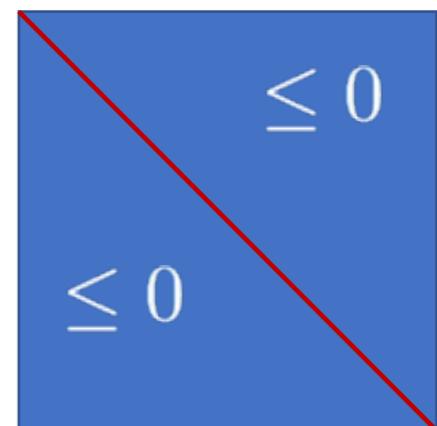
$$F(A) + F(B) \geq F(A \cup B) + F(A \cap B)$$

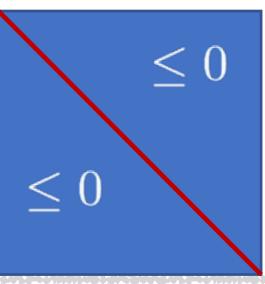
- replace sets by vectors:

$$F(x) + F(y) \geq F(x \vee y) + F(x \wedge y)$$

- or: Hessian has all off-diagonals ≤ 0 . (*Topkis 1978*)

$$\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0 \quad \forall i \neq j$$

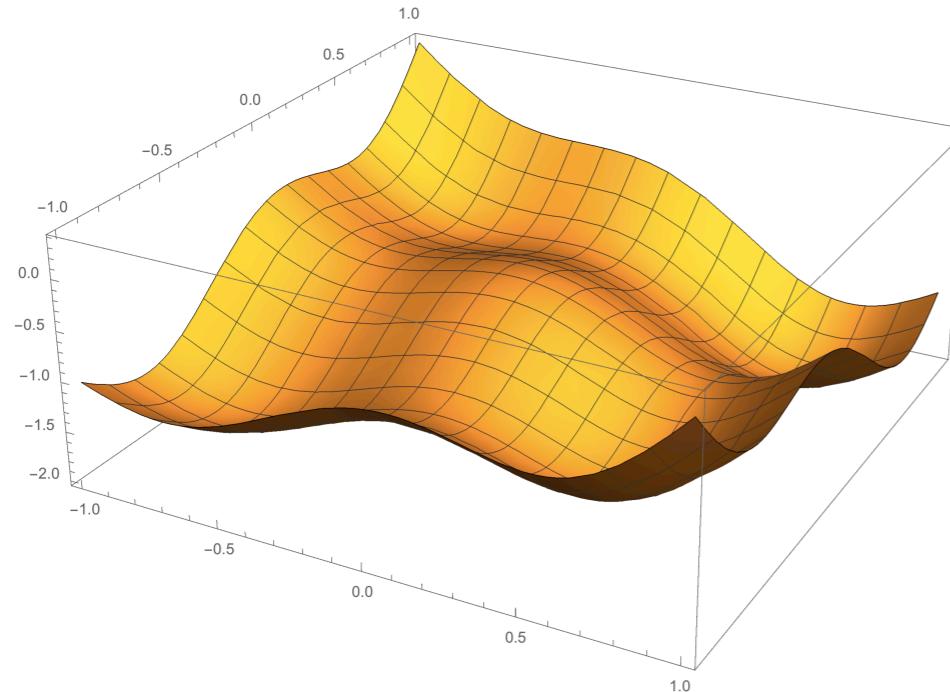




Examples

$$F(x) + F(y) \geq F(x \vee y) + F(x \wedge y)$$

$$\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0 \quad \forall i \neq j$$



submodular function can be
convex, concave or neither!

- any separable function $F(x) = \sum_{i=1}^n F_i(x_i)$
- $F(x) = g(x_i - x_j)$ for concave g
- $F(x) = h(\sum_i x_i)$ for convex h

Optimization

Maximization

- “diminishing returns” stronger than submodularity
- with DR, many results generalize
(incl. “continuous greedy”)

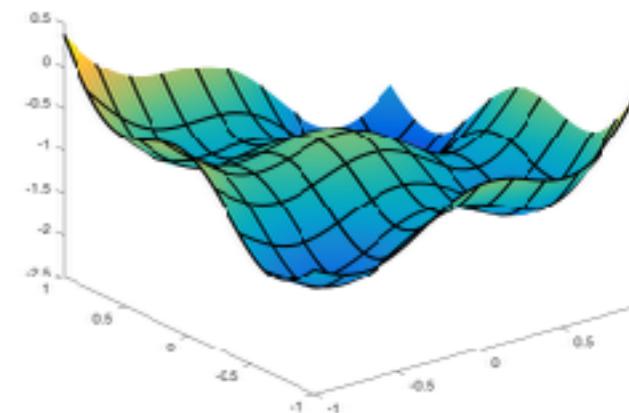
(Kapralov-Post-Vondrák 2010, Soma et al 2014-15, Ene & Nguyen 2016, Bian et al 2016, Gottschalk & Peis 2016)

Minimization

- discretize continuous functions
- then same approaches as for set functions

(Schrijver 2000, Orlin 2007, Bach 2015)

→ *nonconvex optimization via discrete optimization!*



Discrete Probability

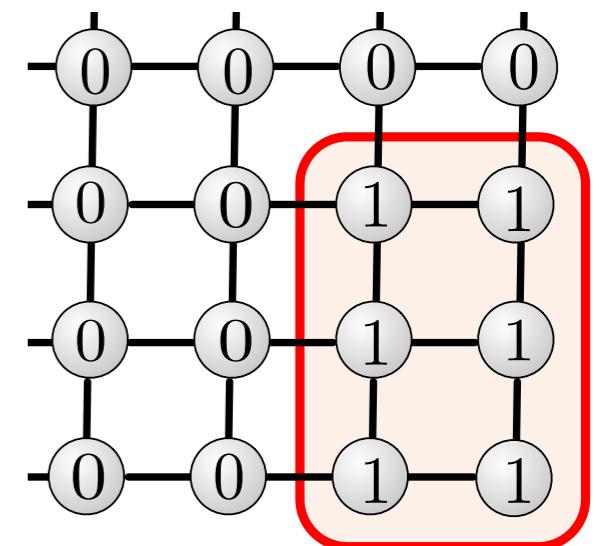
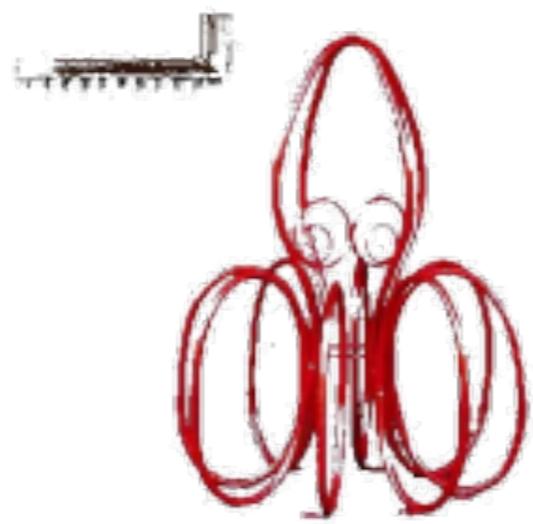
- Distribution over subsets
- Log-submodularity / log-supermodularity

$$P(S) \propto \exp(F(S))$$

- equivalently: n binary random variables

Log-supermodularity

E.g. ferromagnetic Ising model / Conditional Random Field



strong positive correlation:

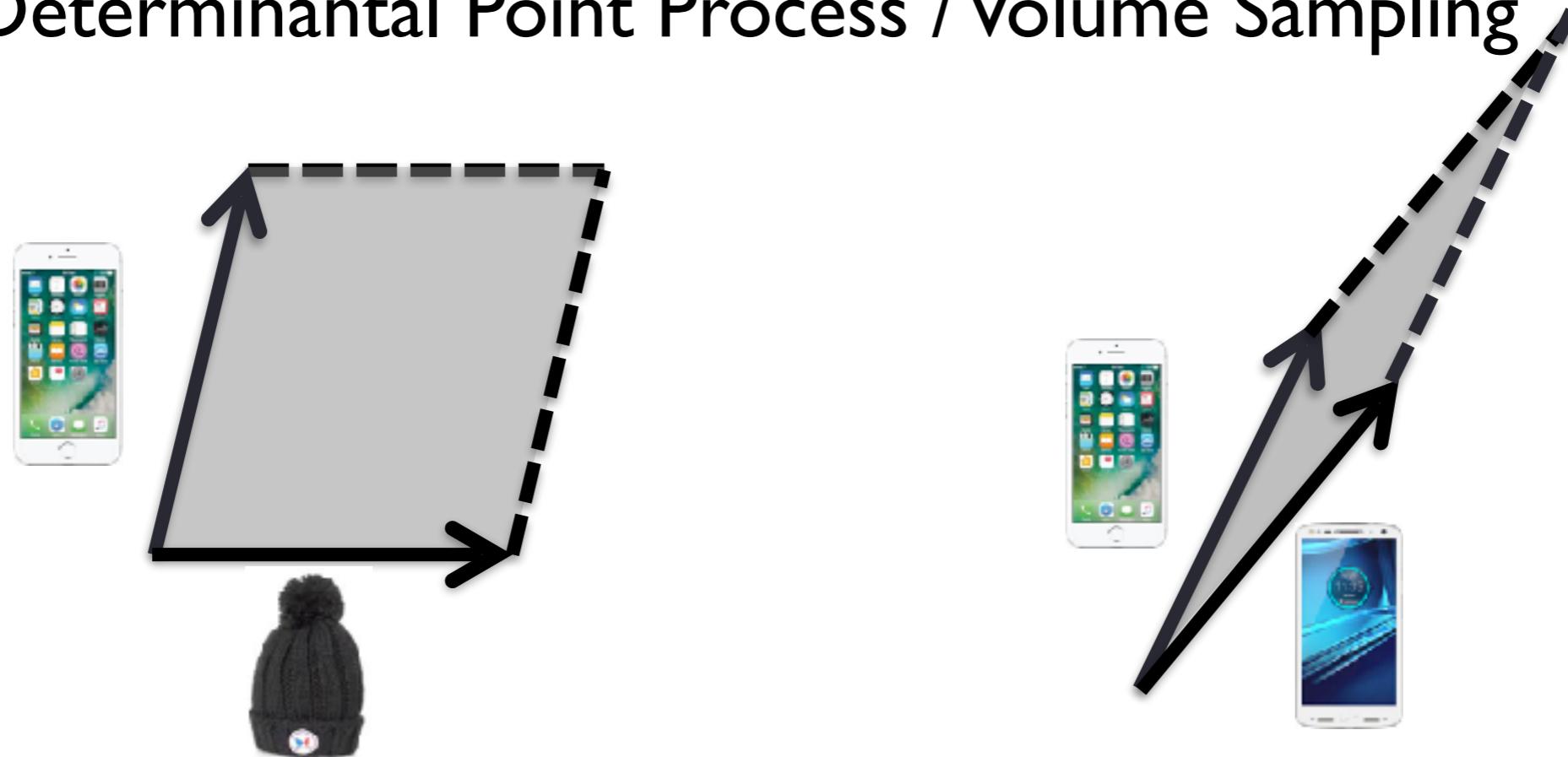
“multivariate totally positive of order 2”, “positive association”

Benefits:

finding the mode = minimizing a submodular function
approximating partition function & marginals ...

Log-Submodularity

E.g. Determinantal Point Process / Volume Sampling



Negative Correlation / Repulsion

Sub-Family has strong negative dependence

Concentration

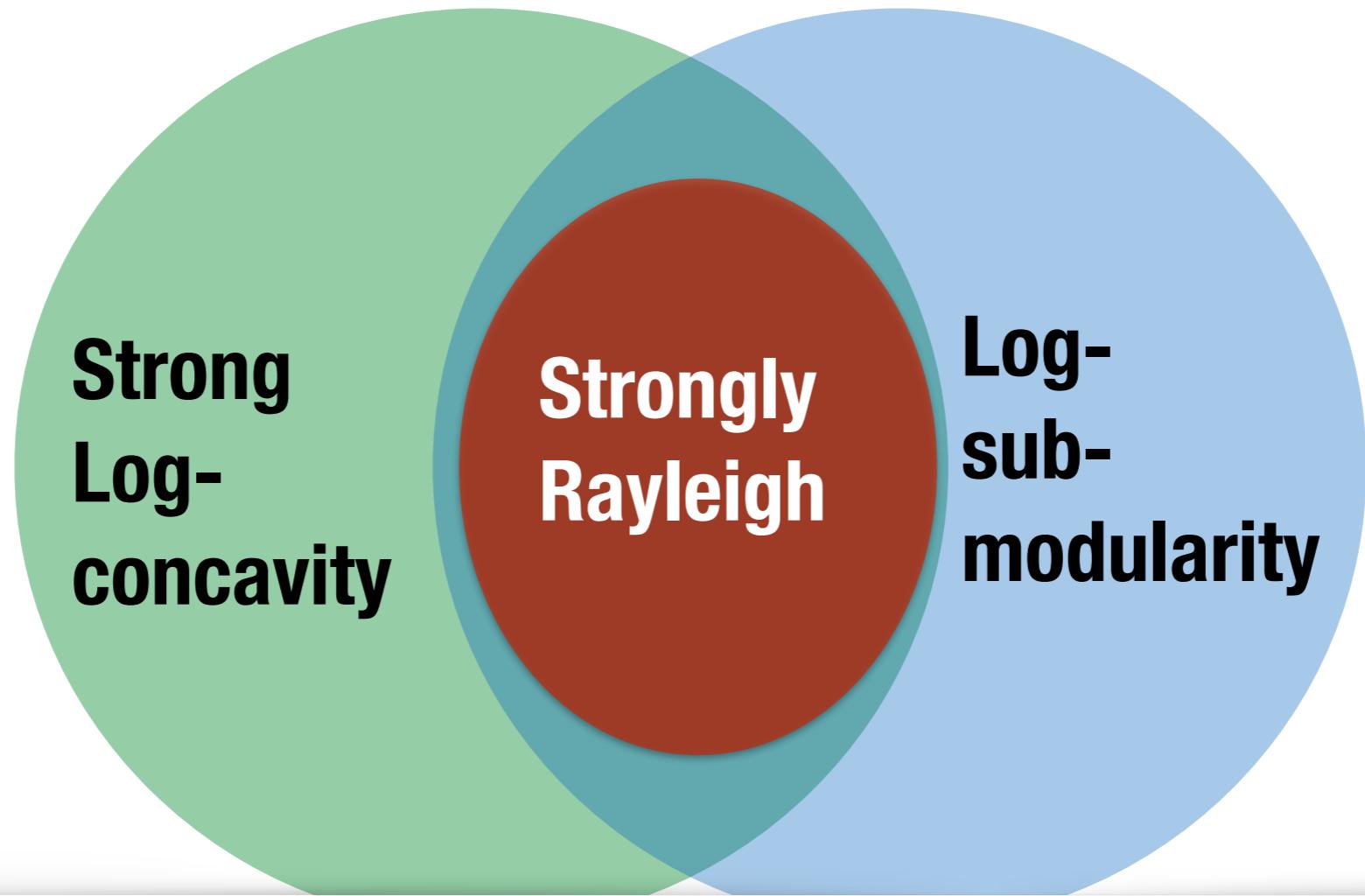
Partition
Functions,
Marginals

**Strong
notions of
Negative
Dependence**

Fast
Sampling

Randomized
Algorithms,
Approximations

Notions of Negative Dependence



NeurIPS | 2018

Thirty-second Conference on Neural Information
Processing Systems

Year (2018) ▾

Help ▾

My Registrations

Dates

Calls ▾

Schedule ▾

Committees ▾

Mon Dec 3rd 11:00 AM -- 01:00 PM @ Room 517 CD

Negative Dependence, Stable Polynomials, and All That

Suvrit Sra · Stefanie Jegelka

Slides Part 1 »

Slides Part 2 »

Indexed Video »

Other directions

- **Approximate Submodularity**
Optimization results hold with additional constants
- **Stochastic Optimization**
Submodular function is an expectation
- **Robust Optimization**
Adversarial Perturbations of Inputs
- **Sequential Decisions / Planning**
- **Learning Submodular Functions**

nonconvex optimization
lattice / continuous submodularity
*many optimization results
generalize*

probability measures
log-supermodular
log-submodular
*sampling, mode,
approx. partition function*

submodular set functions

convexity:
minimization
maximize coherence

dim. returns (concavity):
maximization
maximize diversity

many examples:

- linear/modular functions
- entropy
- mutual information
- rank functions
- coverage
- diffusion in networks
- volume
- graph cut ...