

UNIVERSITY of WISCONSIN-MADISON  
Computer Sciences Department

CS 202: Introduction to Computation Professor Andrea Arpaci-Dusseau

## How can computation... pick best data values? Or, ... turn math into searching?

## Single Dimensional Data

Most examples so far:  
Data of interest has only one dimension

List of "Valuable Numbers"

- Don't know anything else about numbers

## Multi-Dimensional Data

Most data of interest has multiple dimensions

	HW1	HW2	HW3	Exam 1
Bob	90	20	50	60
Alice	75	82	99	86
Evan	15	95	87	83

How would you represent in Scratch?

## How is this Multi-Dimensional Data?

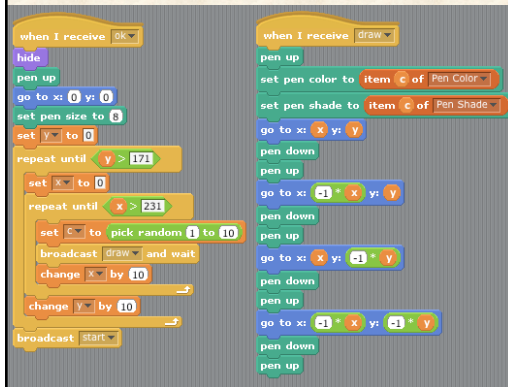
### Stage scripts

What is relationship between two lists?

Initializes two lists:  
Pen Color and Pen Shade

How long is each List?

## What will these scripts do?



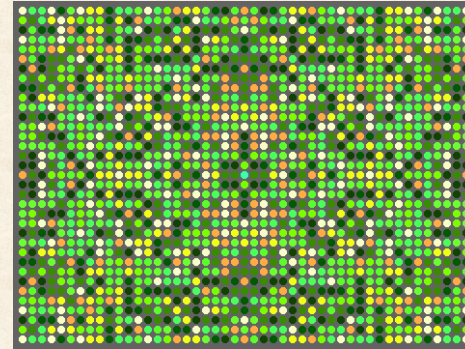
How are x, y varied?

How many pen points for each x,y?

What is relationship of two Lists?

What is purpose of c?

## Result of Running Scripts



Symmetric picture along both x and y axis (4 quarters)  
Ten different color/shade combinations (always paired!)

## More Multi-Dimensional Data: Mozart Musical Dice Game

Compose Minuet at random from 176 possible measures  
Each column – one of 16 measures to be played  
Each row – random roll of two dice summed together

Roll dice to pick specific measure to play (Specified in table)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	96	22	141	41	105	122	11	30	70	121	26	9	112	49	109	14
3	32	6	128	63	146	46	134	81	117	39	126	56	174	18	116	83
4	69	95	158	13	153	55	110	24	66	139	15	132	73	58	145	79
5	40	17	113	85	161	2	159	100	90	176	7	34	67	160	52	170
6	148	74	163	45	80	97	36	107	25	143	64	125	76	136	1	93
7	104	157	27	167	154	68	118	91	138	71	150	29	101	162	23	151
8	152	60	171	53	99	133	21	127	16	155	57	175	43	168	89	172
9	119	84	114	50	140	86	169	94	120	88	48	166	51	115	72	111
10	98	142	42	156	75	129	62	123	65	77	19	82	137	38	149	8
11	54	130	10	103	28	37	106	5	35	20	108	92	12	124	44	131
12	3	87	165	61	135	47	147	33	102	4	31	164	144	59	173	78

Representation in Scratch: 1 List for every measure



Implementation in Scratch

Play sound recording files; How to pick?

For each of 16 measures:

Generate two random numbers and sum together  
Use that sum as index into List for that measure  
Entry in that List tells you what recording to play

## Other Examples?

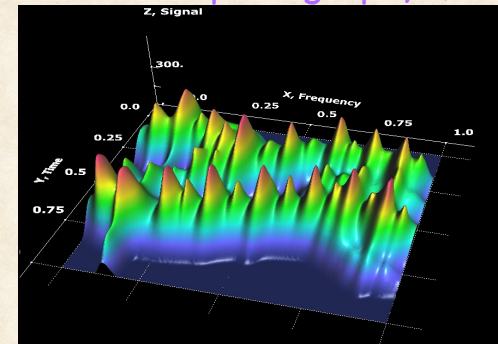
### Other multi-dimensional data

- Over time: pulse, bloodpressure, temp
  - At what time, did person have highest pulse?
- Medical procedures, cost, 5-year survival, 10-year
  - Which procedure has best survival?
- List of materials, strength at different temperatures
  - What material is strongest?
- Baseball players stats
  - Which player hits most triples?

Find max for one metric

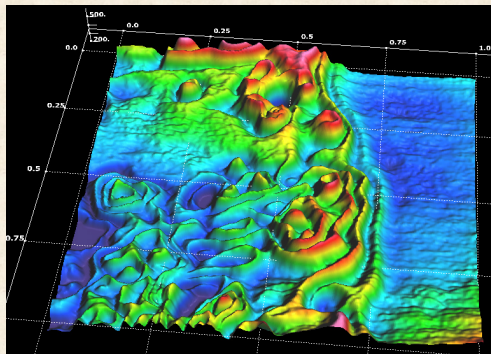
How if data set is very large?

## Data Example: Electroencephalography (EEG)



Assume two inputs: Time and frequency; Assume Output: Spectral signal  
How to find  $(x, y)$  that maximizes  $z$ ??

## Data Example: Quantum Hall effect



2D electrons: Topographic map of random potential by SPM microscopy  
How to find  $(x, y)$  that maximizes  $z$ ?

## How to find max value depending on data organization?

If completely random:

- Linear search (slow...)

Sorted by Key of interest

- Binary search for specific key
- In fact, maximum key is at the end!

Organized by different metric than one of interest

- Examples: Organized by  $(x, y)$
- Data has patterns, locality to it...
- How can we search through this space for max value?



## Much simplified Scenario: Small Data Set - Business Owner

Consider two-dimensions

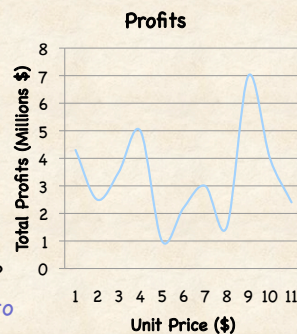
- Single input variable:  
– Unit Price
- Single output variable:  
– Profit

Implementation: Two Lists

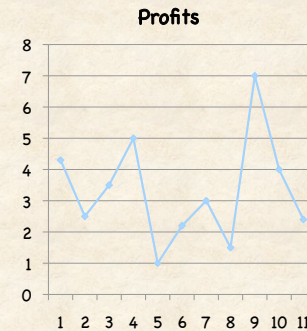
- Unit Price (1, 1.5, 2, 2.5, ... 11)
- Profits (4.2, 3.5, 2.6, ... 2.4)

Small data set: How would you set price to maximize profit?

How can you use computation to find the optimal price?



## Naive Optimization Approach



Evaluate all options and pick best

- Naïve approach:  
Search through entire Profit list for maximum  
– Linear or Binary search?
- Use Max Index to lookup value in Unit Price List (index=17)
- \$9

Why isn't this approach always a good idea?

## Why doesn't Naïve approach always work?

What if too many data points to check all?

- MB ( $10^6$ ), GB ( $10^9$ ), TB ( $10^{12}$ ) of data on disk

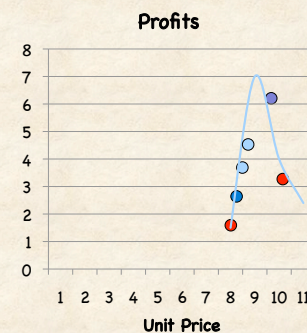
Conclusion: Can't explore entire data set

- Must explore only small number of points

Approach: Assume function has particular shape...

Try to find single maximum of a mysterious function...

## Gradient Ascent: Find Max



Pick 1<sup>st</sup> random point

Pick 2<sup>nd</sup> point in direction A

Worse! What should we do?

Switch direction

Pick 3<sup>rd</sup> point in direction B

Better! What should we do?

Keep going

Pick 4<sup>th</sup> point in dir B

Better! What should we do?

Keep going

Pick 5<sup>th</sup> point in dir B

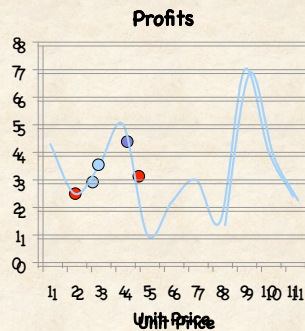
Worse! What should we do?

Switch directions

Pick 6<sup>th</sup> point

Better! Stop, good enough...

## What Can Go Wrong?



Can find local maxima instead of global!

How could you improve algorithm?

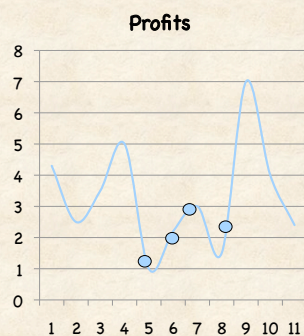
## What Improvements Could You Make?



Approach #1: Try multiple times

- Different random starts
- Keep best result

## What Improvements Could You Make?



Approach #2:  
"Simulated annealing"

- Gradient ascent, but sometimes continue in same direction when see worse results
- Pick worse point with some probability  $p$
- Gradually decrease  $p$  over time

## What about other constraints?

Other multi-dimensional data

- Over time: pulse, bloodpressure, temp
  - At what time, did person have highest pulse and temp < 98?
- Medical procedures, cost, 5-year survival, 10-year
  - Which procedure has lowest cost with reasonable survival?
- List of materials, strength at different temperatures
  - What material is strongest at cold temperature
- Baseball players stats
  - Which player hits most triples, rarely strikes out?

Find max for one metric and satisfy constraints on others

## Optimization: How to Set up Function?

Scenario: You own a business upholstering cushions  
Customer wants you to cover a bolster  
(cylindrical cushion) for \$30

- Must contain 10m<sup>3</sup> of stuffing (Volume)
- Don't care about the dimensions (radius or height)

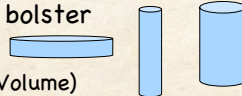
Your costs include material

- Less material you use, the more profit you'll make...
- Goal: Minimize Surface Area

Useful formulas:

$$V = \pi \cdot r^2 \cdot h$$

$$SA = 2\pi r^2 + 2\pi rh$$



## How to Solve Bolster Problem?

$$V = \pi \cdot r^2 \cdot h$$

$$SA = 2\pi r^2 + 2\pi rh$$

Need to find values of r and h that lead to V = 10 and minimize SA

How can you solve with computation? By searching?

Vary one variable

Compute SA as function of that variable

Pick value that gives minimal SA

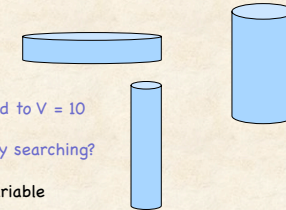
What is our one variable?

Specify h in terms of r (given fixed V)

Substitute for h in SA equation

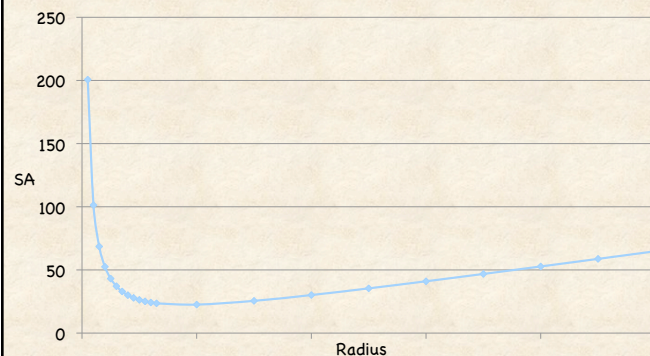
$$h = \frac{V}{\pi r^2}$$

$$SA = 2\pi r^2 + \frac{2V}{r}$$



## What does function look like?

How should you search for radius leading to minimal Surface Area?



## Questions to Ask and Answer

What variable are we varying?

- Radius (input to function)

What variable are we trying to minimize?

- Surface area (output of function)

How do we know the value of the surface area?

$$SA = 2\pi r^2 + \frac{2V}{r}$$

How should we initialize radius?

- Don't know
- Ask the user!

How much should we increase it on each iteration?

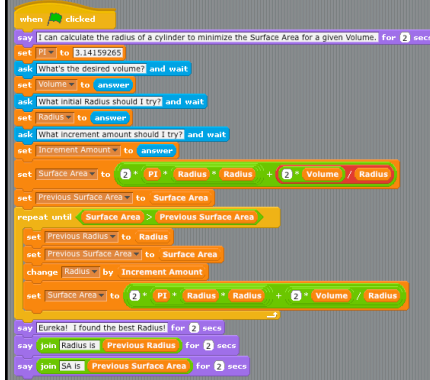
- Don't know, so ask the user!

How do we know we've found the minimum?

- SA for new radius > SA for prev radius
- Previous radius is the minimum



## Simple Implementation $SA = 2\pi r^2 + \frac{2V}{r}$



Keep trying larger values  
of Radius  
until  $SA > \text{prev SA}$

For  $V = 10\text{m}^3$   
Best radius = 1.2m  
Leads to SA of  $25.71\text{m}^2$   
Height = 2.21m

Code structure similar for  
any optimization  
problem

- Replace  $f(\text{radius})$

## Announcements

### Optimization

- Many engineers and scientists use computation for optimization
- How to use computation find parameters leading to best (max, or min) result
- Simplest: Search through parameter space linearly, stop at max/min; more sophisticated techniques to find global optimum

### Announcements

- HW 5: Due today at 5pm
  - Victor lab hours after class from 11-1pm
  - Scratch website down recently, so no voting yet...
- Monday: Laptop Day
  - Work together on constructing HISTOGRAMS for wordle in HW 6 (List intensive!)
  - XO laptops for those without