Lecture 21: How can computation... pick best data values? Or, ... turn math into searching?

"On the left, this year's reading scores. On the right, the same data digitally enhanced."
Data is Important

Every field values their data
  • Science, engineering, medical, business, sports
  • Data source: Measurements, simulations, mathematical functions

What does one often want to do with data?
  • Pick set of input parameters that lead to best result
    - Which materials at which temps are strongest?
    - Which medical procedure saves most lives at least cost?
    - What stock portfolio gives best profits while minimizing risk?
    - Which pitchers most likely to get next batters out?

General Question: Which parameter values optimize (maximize or minimize) desired function while meeting some constraints?
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 \(z\)?
Much simplified Scenario: Business Owner

Consider single input variable:
- Unit Price

Imagine: Profit = f(unit price)
- Someone magically knows f()
- f() might be very complex...

How would you set the price to maximize profit?

How can you use computation to find the optimal price?

![Graph showing total profits vs. unit price]
Naive Optimization Approach

Evaluate set of options and pick best
- Pick some prices (x values) (every $1 thru $11)
- Compute profit = f(price) and store in profit List
- Naïve approach: Search through entire list for maximum
  - Linear or Binary search?

Why isn’t this approach always a good idea?
Why doesn’t Naïve approach always work?

What if too many data points to test all?
  • MB (10^6), GB (10^9), TB (10^{12}) of data on disk

What if very costly to compute f(x)?
  • Computing f(x) could be running a day-long simulation on 1000 processors

Can’t explore entire data set
  • Must explore only small number of points

Approach: Assume function has particular shape...
Gradient Ascent: Find Max

Pick 1st random point
Pick 2nd point in direction A
  Worse! What should we do?
  Switch direction
Pick 3rd point in direction B
  Better! What should we do?
  Keep going
Pick 4th point in dir B
  Better! What should we do?
  Keep going
Pick 5th point in dir B
  • Worse! What should we do?
  • Switch directions
Pick 6th 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
Implementing Optimization in Scratch

Scenario: You own a business upholstering cushions
Customer wants you to cover a bolster (cylindrical cushion) for $30
• Must contain 10m$^3$ 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

What variable are we trying to minimize?
• Surface area

How do we know the value of the surface area?

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

\[ SA = 2\pi r^2 + \frac{2V}{r} \]
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 25.71m²
Height = 2.21m

Code structure similar for any optimization problem
- Replace \( f(r) \)
Today’s Summary

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

• Programming Project 1: Draft due today at 5pm in Scratch Gallery