Teaching Is Technology

Steve Pittard wsp@emory.edu

Correlation is Not Causation

Teaching is Also Tough

Figure 1: Jeff Berlin - Virtuoso Bassist and Educator

Teaching is Also Tough

"I feel a large part of music education is a disaster today. A lot of it is coming from people who don't know how to teach"

Wild Claims !

PLAY GUITAR IN 7 DAYS OR GET YOUR MONEY BACK

ED SALE, TOP RADIO GUITARIST, TEACHER OF HUNDREDS OF GUITARISTS, PROFESSIONALS, WILL POSITIVELY TEACH YOU TO PLAY A BEAUTIFUL SONG THE FIRST DAY! AND ANY SONG BY EAR OR NOTE IN 7 DAYS!

Have to be able to play this

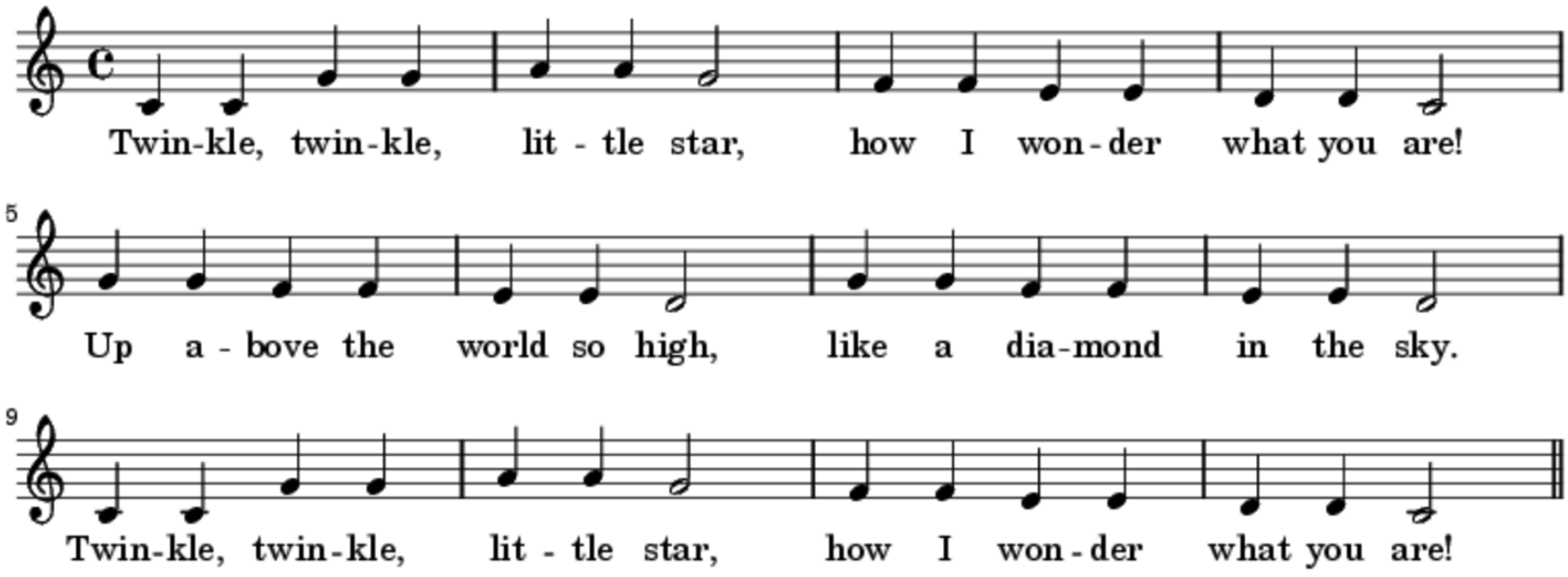


Figure 2: Easy

Before playing this

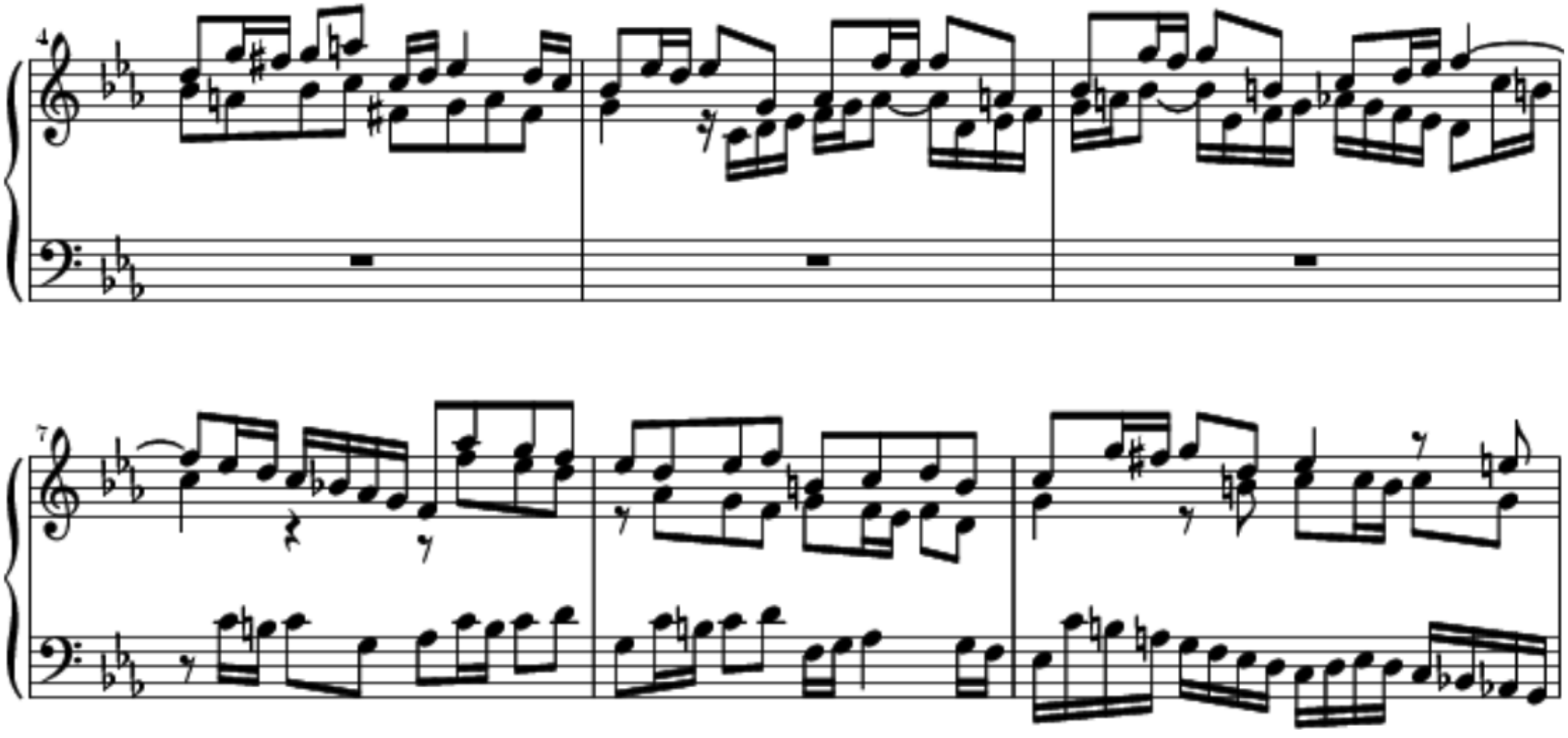


Figure 3: Not Easy

Wild Claims Continued ?

Figure 4: Quickie Route to Data Scientist

Specialization Trend

- R for Business Intelligence
- Data Science for Epidemiologists
- Data Science for Nursing Research
- Intro to Statistics using R
- Data Visualization for Biologists
- Statistics for Clinicians
- Machine Learning for City Planners
- R for SAS Programmers

Evolution of Data Science Roles

- Data Scientist
- Data Analyst
- Data Engineer
- Data Architect
- Chief Data Officer

I (Co)Teach Several Classes and Do Frequent Guest Lectures

- BIOS 545 Introduction to R Programming
  - For BIOS students only (for the most part)
  - Ground up introduction to R
  - Fast-paced
  - Assumes some previous programming

- BIOS 534 Introduction to Machine Learning
  - Drafted at the last minute to help out
  - Blend of theoretical and practical
  - Students don't like the theoretical yet it's important
  - The practical draws together lots if approaches
  - Everyone thinks Data Science is all Deep Learning

Guest Lectures

- Nursing 741 Big Data Analytics
  - Analyze ethical issues related to the use of big data in health related research
  - Understand the principles of reproducible research
  - Analyze ethical issues related to big data analytics in health related research

- LAW 890 Fundamentals of Innovation

  TI:GER, brings together graduate students in law, business, science, and engineering to work on start-up projects to transform highly promising research into economically viable projects.

(https://tigerinnovation.org/)

Lean Start-up

Most of what I do (or try to) is based on the Lean Start-up Concept which relies heavily on "customer discovery".

- Seventy five percent of ALL business start-ups fail
- Cool technology and ideas are frequently NOT ready for market
- Large Business Plans are conceived using stale market research
- No one "really" wants to talk to customers

Lean Start-up Continued

Lean Start-up avoids this by:

- Frequent hypothesis testing over elaborate planning
- Customer feedback over intuition or "we know what's best"
- Iterative Design (e.g. Fail towards success) over Big Design
- Frequent "pivots" into new directions

https://leanlaunchpad.stanford.edu/
https://steveblank.com/slides/lean-startup/

Lean Start-Up Continued

Using this approach I have:

- Obtained a patent
- Gotten funding from Georgia Research Alliance
- Active Intellectual Property declarations with Emory OTT
- Function as Entrepreneurial Lead for a Natural Language Processing recruitment tool

Motivating Learning

Why Do You Take a Class ?

- The course is required
- It will be useful for later work
- Someone advised me to take this class
- I couldn't find a more interesting course to take

(Better) Motivations for Learning ?

- Less reliance upon others
- Enhanced productivity
- Eliminate barriers to research
- Improve employment prospects

You Are Really Facilitating Independence

- As a teacher or student you SHOULD have this goal
- You Need to know the material
- Better Yet, you need to know how to explain it

Feynman Approach

1. Identify the topic you want to learn.
2. Explain the topic as if teaching a sixth-grader.
3. Identify areas of improvement in your explanation.
4. Improve and optimize your explanation.
5. Repeat the steps until you've achieved mastery of the topic.

Relative to step 2, replace the phrase "sixth grader" with a term corresponding to a beginner relative to the topic of interest (e.g. "first year")

Thelonious Monk Approach

DON'T PLAY EVERYTHING (OR EVERYTIME); LET SOME THINGS GO BY. SOME MUSIC JUST IMAGINED. WHAT YOU DON'T PLAY CAN BE MORE IMPORTANT THAN WHAT YOU DO PLAY.

A NOTE CAN BE SMALL AS A PIN OR AS BIG AS THE WORLD, IT DEPENDS ON YOUR IMAGINATION.

STAY IN SHAPE! SOMETIMES A MUSICIAN WAITS FOR A GIG, & WHEN IT COMES, HE'S OUT OF SHAPE & CAN'T MAKE IT.

WHEN YOU'RE SWINGING, SWING SOME MORE! (WHAT SHOULD WE WORK TONIGHT? SHARP AS POSSIBLE!)

ALWAYS LEAVE THEM WANTING MORE.

DON'T SOUND ANYBODY FOR A GIG, JUST BE ON THE SCENE. THOSE PIECES WERE WRITTEN SO AS TO HAVE SOMETHING TO PLAY, & TO GET CATS INTERESTED ENOUGH TO COME TO REHEARSAL.

YOU'VE GOT IT! IF YOU DON'T WANT TO PLAY, TELL A JOKE OR DANCE, BUT IN ANY CASE, YOU GOT IT! (TO A DRUMMER WHO DIDN'T WANT TO SOLO).

WHATEVER YOU THINK CAN'T BE DONE, SOMEBODY WILL COME ALONG & DO IT'. A GENIUS IS THE ONE MOST LIKE HIMSELF.

THEY TRIED TO GET ME TO HATE WHITE PEOPLE, BUT SOMEONE WOULD ALWAYS COME ALONG & SPOIL IT.

Thelonious Monk Approach

"Stop playing all those weird notes, play the melody"

- Focus on the basics and build into the "edge" cases.
- Don't start with oblique or uncommon topics
- Explain exceptions AFTER having laid the foundation

"Just because you are not a drummer, doesn't mean you don't have to keep time"

- The phrase "keeping time" means establishing a pace and reliable pulse
- Some class sessions will be harder than others but the "pulse" should be same across all sessions.

Thelonious Monk Approach

"Stay in shape. Sometimes a musician waits for a gig and when it comes, he is out of shape and can't make it"

Being "in shape" refers to readiness for playing (not necessarily physical shape). Do your rehearsals and practice so you are ready for action. Be ready to improvise.

Example

In a linear regression model, how does R^2 differ from Adjusted R^2 ?

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$R^2 Adjusted = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Discussion ? Where to Start ? Recitation of Learned Facts ?

Reproducibility and Availability

Crisis

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

Figure 5: Replicating Published Work

Use Reproducible Research Tools

- Sweave
- knitr
- Markdown
- Jupyter

Use Git !!

- Great for Backing Up Your Code
- Can handle multiple versions
- Can Share with others
- Others can modify and submit changes
- Issue Tracking
- Contains a Wiki

Learning Analysis and Programming

Like Learning to Cook

- Start with a good recipe
- Demonstration (by a competent cook)
- Observation (by students)
- Duplication (basic repetition)
- Imitation (repetition with changes)
- Fear of Messing Things Up

Just Eat Frozen Food or Eat Out

- Unhealthy
- Little Variation
- Expensive
- No Self Sufficiency
- No Independence

Creating Interesting Material

Think of an encompassing project

- For advanced students give them the project all at once
- For Intro classes cover each subtopic per week + Homework and Labs + Maybe a final project
- For Intermediate classes do a combination + Have mid term and final projects + Some supporting homework

Yum !

1 Assignment

Create a recipe to cook and serve Fettucine Carbonara to a group of six people for dinner tonight.

- The recipe must present options to accomodate wide variation in the ultimate number of diners. That is you must provide preparation options from 1 diner up to 10.
- You must make the pasta yourself. Store bought pasta is not permissible for purposes of this assignment and use thereof will result in a failing grade.
- The pasta must be hand-rolled (machine or hand crank not acceptable) and properly hung to exhbit neat folds to simplify boiling. Spinach may be integrated into the egg-flour mix to provide extra color.
- When served, the pasta must be firm to bite but not so undercooked that it exhibits a chewy, rubbery texture.
- The pasta should be lightly coated with a creamy parmesian sauce seasoned with pepper. There should be no after taste except for a slight hint of the shallots and garlic used as a foundation for the sauce.
- You may adorn the pasta with sun dried tomatoes and or other vegetables that complement the flavor of the sauce. However, attention must be given to aesthetics and unappealing color combinations will result in point deductions.
- While the recipe mentions fettucine specifically, other types of pasta will be acceptable such as ravioli and shapghetti as long as there is no compromise in flavor and presentation.

Make the Pasta

Figure 7: Mmmm.. Making Pasta

Cut the Pasta

Figure 8: Cut the Pasta

Dry the Pasta

Make the Sauce

Boil Pasta

Programming Assignment

1 Assignment

Write an R package that implements user selected methods (from a known list) to find the square root of a number and more generally the roots of a polynomial.

- This package should pass any all tests required by CRAN and provide compreheneisve end user documentation with realistic examples.
- The package method should return an S4 object complete with helpful accessor methods including visualiztion plots and information on time to convergence for a given input set.
- Your method(s) must accomodate input in vectors, lists, matricies, and data frames. Extensive error checking is expected and required to insure that user input is reasonable.
- You may implement any number of supporting methods within the object but these should be private and not expose to the user. To this end your package implementation should reflect intelligent use of the NAMESPACE file.
- You should provide an option to trigger a performance timing method to help the user understand what methods provide optimal convergence for a given input set.
- The package must be implemented using devtools which also implies the availability of your package on both CRAN and a public Github repository

You Gotta Break Things Up

Solving for $\sqrt{S}$ is the same as solving $f(x) = x^2 - S = 0$

We can use Newton's Method to iterate towards an answer:

$$x_{n+1} = x_n - \frac{f(x)_n}{f'(x)_n} = x_n - \frac{x_n^2 - S}{2x_n} = \frac{1}{2}(x_n + \frac{S}{x_n})$$

Make a first guess and then compute x_{n+1} until x_{n+1}^2 is close enough to S within some specified tolerance

A Beginning Approach

```
Make a guess (e.g 2) of the square root of some number (e.g
Specify a tolerance level (e.g. 0.0001)

while the absolute value of the square of that guess minus
      the number  is >= than tolerance
          compute a new guess
```

```
somenum <- 16
tolerance <- 0.0001
guess <- 2

# Note we have to use Absolute value since we
# are concerned with the magnitude of the difference

while( abs((guess^2)-somenum) >= tolerance) {
  guess <- (guess + (somenum/guess)) * 0.5
}
guess
```

Segment the work into functions

```
somenum <- 16
tolerance <- 0.0001
guess <- 2

# Write a Function to judge quality of computed guess

compare <- function(guess,target) {
   diff <- abs((guess^2)-target)
   return(diff)
}

while(compare(guess,target) >= tolerance) {
  guess <- (guess + (target/guess)) * 0.5
}

guess
```

Create functions with sane arguments and defaults

```
mySqrt <- function(target=16,guess=2,tolerance=0.0001,verbo
  while( abs((guess^2)-target) >= tolerance) {
    guess <- (guess + (target/guess)) * 0.5
    if (verbose) {
      print(guess)
    }
  }
  return(guess)
}
```

Functions that embed other functions

```
mySqrt <- function(target=16,guess=2,tolerance=0.0001,verbo
  compare <- function(guess,target) {
   diff <- abs((guess^2)-target)
   return(diff)
  }

  while(compare(guess,target) >= tolerance) {
    guess <- (guess + (target/guess)) * 0.5
    if (verbose) {
      print(guess)
    }
  }
  return(guess)
}
mySqrt(16,2,0.0001)
```

Vectorize the function

```
sapply(c(16,22,49,39),mySqrt)

# or

round(sapply(c(16,22,49,39),mySqrt),2)

# or

Sqrt <- function(input=c(16,22,49,39),rnd=2) {
  retvec <- round(sapply(input,mySqrt),rnd)
  return(retvec)
}
```

Now for the Dark Side

Another Crisis ?

The New York Times

U.S.

As Computer Coding Classes Swell, So Does Cheating

By JESS BIDGOOD and JEREMY B. MERRILL MAY 29, 2017

Academic Dishonesty

The New York Times

Business Day
Technology

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

Scandal Over Cheating At M.I.T. Stirs Debate On Limits of Teamwork

By FOX BUTTERFIELD,
Published: May 22, 1991

CAMBRIDGE, Mass., May 21— The largest cheating scandal in the history of the Massachusetts Institute of Technology has touched off a campuswide debate over how far collaborative work on scientific projects can go before it crosses the line into cheating.

The scandal, in an entry-level computer programming course, resulted in 73 students being disciplined this spring. It has shocked much of the faculty to discover that so many high-caliber students would feel compelled to cheat. But many M.I.T. students say they are not surprised, given the heavy workload and the competitive pressures.

Academic Dishonesty

- Lectures on Honor Code Are Helpful
- Keeps the Honest Person Honest
- It Can Still Happen
- Google Is Easy To Use
- Stack Overflow Knowingly Aids and Abets Cheating
- Cheating is Easy to Spot

Academic Dishonesty

- Competition
- Start the assignment late
- Later assignments rely on earlier work
- Very easy to get help from others
- Students circulate past assignments
- Claims of language confusion

Academic Dishonesty

In 2011 this assignment was given in the "Paradigms for Computing'' class at Stanford

1 Assignment - Vectorizing A Function

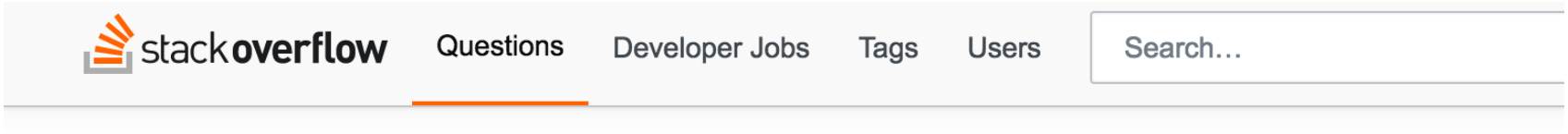
In mathematics (and, in particular, functional analysis) convolution is a mathematical operation on two functions (f and g) to produce a third function, that is typically viewed as a modified version of one of the original functions, giving the integral of the pointwise multiplication of the two functions as a function of the amount that one of the original functions is translated.

Here is a function written in R that implements convolution on two given vectors x and y. Since R is an interpreted language the double for loop will be slow. For vectors of increasing lengthe the performance will slow down significantly. See if you can vectorize this function to improve speed of processing.

```
convolveSlow <- function(x, y) {
nx <- length(x); ny <- length(y)
xy <- numeric(nx + ny - 1)
for(i in seq(length = nx)) {
 xi <- x[[i]]
        for(j in seq(length = ny)) {
            ij <- i+j-1
            xy[[ij]] <- xy[[ij]] + xi * y[[j]]
        }
```

Academic Dishonesty

Within 1 day this was found on the Stack Overflow site:



9

3

I have a R code that can do convolution of two functions...

```
convolveSlow <- function(x, y) {
nx <- length(x); ny <- length(y)
xy <- numeric(nx + ny - 1)
for(i in seq(length = nx)) {
 xi <- x[[i]]
        for(j in seq(length = ny)) {
            ij <- i+j-1
            xy[[ij]] <- xy[[ij]] + xi * y[[j]]
        }
    }
    xy
}
```

Is there a way to remove the two for loops and make the code run faster?

Thank you San

r loops