

Welcome to Data Structures & Algorithms

Or, how to pass technical interviews given by programmers

Terence Parr
MSDS program
University of San Francisco

Course contents

<https://github.com/parr/msds689>

- A formula for problem-solving simple algorithm problems
- How to read code
- Core data structures, a unifying perspective
- Algorithm complexity analysis
- “*So much recursion!*” – MSDS2019 student comment
- Walking and searching data structures
- Sorting (with all of my dirty tricks)
- Graphs and graph algorithms

Course projects

- Convert htable project to object-oriented version (8%)
 - With some extensions
 - ...and using somebody else's code from two years ago!
 - hint: it's kinda stinky code. ha!
- kmeans clustering, kmeans++ initial point selection (20%)
 - Spectral clustering using Breiman's unsupervised learning trick for RFs
 - Image compression applications
- Feature importance and selection (22%)
 - Permutation and drop column
 - Automatic feature selection
- Work as hard or as little as you want (I give no unit tests)
 - grader will assign check -, check, check+ based upon your reports

<https://github.com/parr/msds689/blob/master/projects/oohtable/oohtable.md>

<https://github.com/parr/msds689/blob/master/projects/kmeans/kmeans.md>

<https://github.com/parr/msds689/blob/master/projects/featimp/featimp.md>

Student evaluation

Please note grader will take at least a week to grade projects, but I'll grade exams quickly

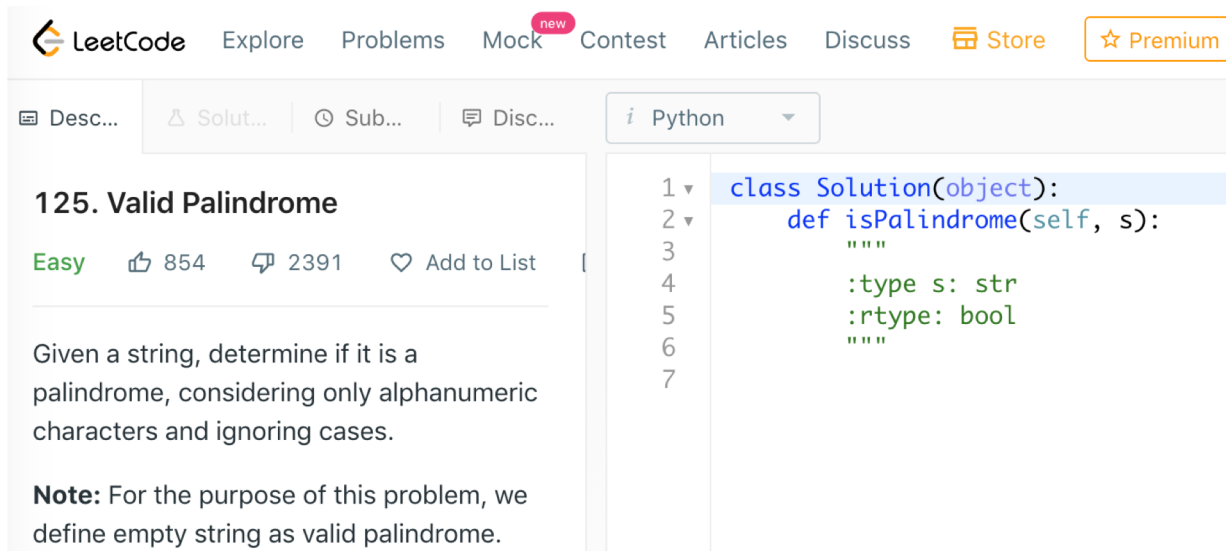
Artifact	Grade Weight	Due date
OO hashtable	8%	Fri, Jan 31 11:59pm
Clustering	20%	Sun, Feb 16 11:59pm
Feature selection and importance	22%	Wed, Mar 4 11:59pm
Exam 1	20%	3:15PM-4:15PM Thur Feb 13
Exam 2	30%	10AM-11:00AM Fri Mar 6

last day of class



Extra things you can do

- Lots of little practice quizzes; e.g.,
<https://github.com/parrt/msds689/blob/master/labs/quiz-oo.ipynb>
- LeetCode algorithm and data structures challenges. e.g.,
<https://leetcode.com/problems/valid-palindrome/>



The screenshot shows the LeetCode website interface. At the top, there's a navigation bar with links: LeetCode, Explore, Problems, Mock (with a 'new' badge), Contest, Articles, Discuss, Store, and Premium. Below this, the 'Valid Palindrome' problem (125) is displayed. The problem is marked as 'Easy' and has 854 likes and 2391 discussions. The description states: 'Given a string, determine if it is a palindrome, considering only alphanumeric characters and ignoring cases.' A note specifies: 'Note: For the purpose of this problem, we define empty string as valid palindrome.' On the right, a Python code editor shows a class solution:

```
1 class Solution(object):
2     def isPalindrome(self, s):
3         """
4         :type s: str
5         :rtype: bool
6         """
7
```

Resources

- A great free book on [algorithms by Jeff Erickson](#)
- Kleinberg and Tardos, *Algorithm Design*
 - Please see compressed pdf kleinberg-common-running-times.7z in Canvas course files area (do not post material publicly please)
- A very useful set of [programming-concepts-for-data-science](#) and [data science coding questions](#) by former USF MSDS student [Shikhar Gupta](#)
- [10 steps to solving a programming problem](#)
- A review [OO notebook](#) and [Operator overloading notebook](#)

Administrivia

- The usual academic honesty rules will be enforced; in projects, reports, exams or any other artifact; [Honor Code](#)
 - Do not represent another person's work as your own
 - Don't leave your laptop unattended/unlocked; others can take a picture of your code or simply use a USB key quickly
- Students with Disabilities
 - If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF [Student Disability Services](#) (SDS) for information about accommodations.
 - More details on the course syllabus: <https://github.com/parrt/msds689>

Why this course, why now?

- At least for the moment, many of the people interviewing you will be programmers, pretending to be data scientists
- What do they know? Programming, data structures, and algorithms
- Being able to organize data within a machine or cross machines is a key skill for a data scientist
- The larger the data, the more critical it is to understand how to measure algorithm performance and how to design efficient solutions
- Optimally, this course would be much earlier, but the timing is good for your interviewing and was only spot we could jam this course in

How to get a job

1. Be accomplished, be interesting
 2. Know lots of people
 3. Mine social network looking for job
- When then fails or simultaneously
 - Cold apply to jobs via the web