CST 495: Advanced Machine Learning and Big Data

Glenn Bruns
CSUMB

image credit: Martin Pyka
http://www.martinpyka.de/wp-content/uploads/2014/02/Calender_neuron.jpg

Course Description

Students learn to use advanced machine learning methods and big data technologies.

The advanced machine learning methods include support vector machines, ensemble methods, neural nets, and methods for working with time series data.

16 weeks from now, you will be able to use the Python machine learning ecosystem, and will be able to write TensorFlow code for deep learning.

Your growth as a computer scientist

You will grow in 2 ways:

- You will develop understanding and skill with machine learning
- You will solve a bunch of challenging problems

You need to solve challenging problems to become a better problem solver and computer scientist.

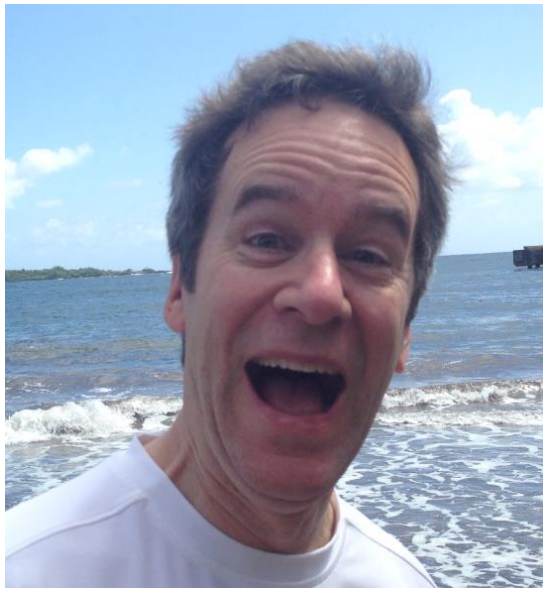
It can be a painful process – do not try to avoid it.

Contacting me

- Slack – best for quick response
- email: gbruns@csumb.edu
- phone: (831)582-4006
- office: BIT 213
- office hours:
  - Mon: 9:30 – 11
  - Tu/Th: 1-2
  - Wed: 12-1
  - Fri: 2-3

About me

- BS in Chem. Engineering from Cal. State Northridge
- Masters in Software Engineering from the Wang Institute, MA
- PhD in Comp. Science from the Univ. of Edinburgh, Scotland
- Software developer for 5+ years
- Comp. science researcher at Univ. of Edinburgh, Bell Labs
- Data Science Consultant at Cisco
- Research areas: data science, software verification, security, …



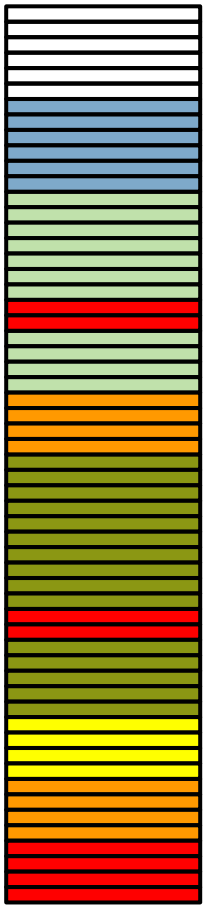
Course Structure

Course is organized into “modules” of 50 minutes.

- 25-30 minute lecture
- 20-25 minute lab

2 modules per class session, break inbetween

Visualization of course



intro + ML with python

linear algebra review

advanced methods

- optimization methods
- support vector machines
- ensemble learning
- dimensionality reduction

time series data

deep learning

projects

quiz and exams

Texts

Aurelien Geron, Hand-On Machine Learning with Scikit-Learn & TensorFlow, O'Reilly. Available from Amazon.

Optional:

- Cowpertwait and Metcalfe, Introductory Time Series with R, Springer.

  (for time series segment of course)

- Rui Miguel Forte, Mastering Predictive Analytics with R, Packt Publishing

  (used in Data Mining course, nice chapter on time series data)

How to learn

Research (see book "Make it Stick") shows these common learning strategies are not very effective:

- rereading text or lecture slides
- massed practice (cramming)
- blocked practice (drilling on a single problem type)

More effective learning strategies are:

- retrieval from memory
- spaced practice
- elaboration

Lab work will use effective learning strategies, and they might sometimes seem painful.

Peerwise

Peerwise is an online tool that allows you to:

- create multiple-choice questions
- answer questions written by other class members
- rate and comment on others' questions

The value is:

- writing questions is a great way to learn
- in writing multiple-choice questions you think about common misunderstandings

peerwise.cs.auckland.ac.nz

- our school is listed as CSU, Monterey Bay
- our course ID is 17698

Slack

Our class has its own Slack workspace!

CST 463, Fall 18

cst463-f18.slack.com

It is required that you join the workspace, because:

- I make class announcements on Slack
- Discussion with classmates will happen there
- I respond most quickly to Slack messages

Invitation to the workspace is at the top of our iLearn page.

Course policies

- I design the course to try to maximize learning and involvement
- Policies are based on learning research
- There will be a lot of work, but I hope it will be engaging

During class

- Please ask questions! No need to raise hand
- Phones -- put off desk and silent
- Laptops -- closed during lecture
- Please move close to the front of room

Multiple studies show students who take handwritten notes perform the best.

theverge.com/2017/11/27/16703904/laptop-learning-lecture

Lecture slides and lecture videos will be posted to iLearn for every class!

Labs and active learning

Labs are a key part of class.

Please:

- focus on the lab work
- work with a neighbor
- ask me if you need information

Please don't:

- do other things during lab time
- leave the class during lab
- stop working before the end of class

Attendance

- Attendance is required
- Late arrival = absent
- You get two unexcused absences without a penalty, but each additional unexcused absence will result in a deduction of 3 points (out of 100 possible) from your overall course score.
- Let me know well in advance if you have any special circumstances

I'm sorry if it seems harsh, but attendance is associated with increased learning.

Also, would you miss or be late for meetings at work?

Homework

- Assigned weekly, usually at beginning of week
- At least 50% of homework will be graded
- Homework not accepted past due date!
- Don't ask me to accept late homework. I will drop your lowest homework grade dropped to account for the time when you:
  - fall asleep
  - lose your internet just before the homework is due
  - think the iLearn homework closed 10 seconds earlier than it should have

Expect to spend about 8 hours per week on homework for this class. Please let me know if you're spending much more!

Office visits

Please drop by my office every so often.

If you haven't passed a class with me before, you must schedule a visit with me during the first four weeks of class.

If you ever want a recommendation from me, it will help you a lot if we know each other.

Scheduling office visits: See scheduling link on our iLearn page or my CSUMB page.

Visits blocks are 10 minutes. Schedule two of them if you need a longer visit.

Exams

Let me know well in advance if you cannot attend an exam. If you have a compelling reason then a makeup will be arranged.

Getting help and academic integrity

On labs and projects you work together.

On homework you work on your own.

If you need help or hints:

- See me during our office hours
- Contact me by Slack
- Ask your classmates for hints on Slack – so that all may benefit

Cheating of any kind won't be tolerated in this class. If cheating is detected on homework or exams then zero points will be given for the assignment and the students involved will be reported to the CSUMB Office of Judicial Affairs and Community Standards.

Cheating includes plagiarism and working together on homework.

Please see csumb.edu/policy/academic-integrity-policy

Reading

It’s important to spend time on the assigned readings.

- The author covers many things we don’t have time to cover in lecture
- You need to learn to read dense technical material.

Read the material slowly.
Think about what you are reading; take notes.
Don’t try to read many pages in one sitting.

Evaluation

- **Homework** (25% total).
- **Projects** (20% total).
- **Mid-terms** (25% total). Two 65-minute mid-terms.
- **Final** (20% total). A one hour, 45 minute final.
- **Participation** (10% total).
  - based on quality and quantity of your Peerwise contributions

Final course score can be reduced by unexcused absence.

5% penalty on overall class score if you don't schedule an office visit for first four weeks of class.

Grades

[97,100]	→ A+
[93, 97)	→ A
[90, 93)	→ A-
[87, 90)	→ B+
[83, 87)	→ B
[80, 83)	→ B-
[77, 80)	→ C+
[70, 77)	→ C
[60, 70)	→ D
[0, 60)	→ F

[a,b] means ≥ a and ≤ b
[a,b) means ≥ a and < b

A letter grade of C or better is needed to satisfy graduation requirements.

Academic Support

csumb.edu/scd/academic-support-services

- Above and beyond CS workshop (ABCS)
  - Fridays 2-5, BIT 113 (juniors, seniors)
  - FB representatives help with resume writing, applying, interview prep; hiring orgs will present
- TA help
- Center for Student Success (CSS)
  - one-on-one support, study skills workshops, etc.
  - Library, 3rd floor, Suite 3180

Students with Disabilities

Students with disabilities who require accommodations such as time extensions or alternate media/format must present current verification from Student Disability Resources as soon as possible.

Please schedule an appointment to discuss specifics with me. If you think a disability may impact your performance in this class, meet with SDR professional staff in the Health and Wellness Services, Building #80. Contact SDR: (831) 582 - 3672 or Email: student_disability_resources@csumb.edu

Some things to remember

- Have fun
- Focus on learning and mastery
- Start homework early; spend enough time on it
- Look at the detailed course outline on iLearn
- Please call me Prof. Bruns or Dr. Bruns
- Drop by my office to chat

- I am also committed to your success and will do all I can to help you.
- You must take responsibility for your learning!