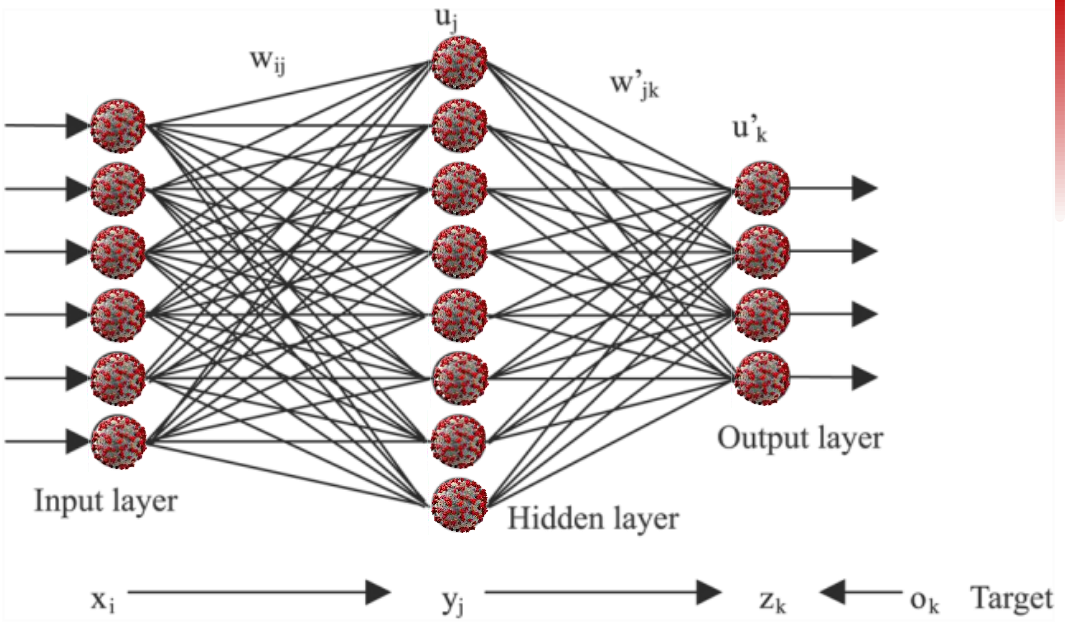
McMaster University ENGINEERING

# Chemical Engineering 4H03

## Covid-19 Edition

Jake Nease
McMaster University



# First Things First…

- I would appreciate you letting me know of:
  - Any **erors** you find in the course notes, assignments, tests, *etc*.
  - Any suggestions you have to improve the course
  - *This is especially true for our second-ever offering of 4H!*

- Either of these can be done by writing to:
  - neasej@mcmaster.ca
  - che4h3instructor@gmail.com

- Anonymous comments and suggestions can be made on our Google Survey – 4H03 Course Feedback

# Credits

- A great deal of this course content is derived from a certain graduate course I have taken and Kevin Dunn's class materials
  - ConnectMV is the name of his company
  - *Portions of this work are copyright of ConnectMV*
  - You may remember Mr. Dunn's material (:3) from MATLS 3J03

- There are others to thank, too
  - Your TAs, who are working on new content
  - Dr. Brandon Corbett
  - Many more!

# Hiya. This is me.

- Graduated ChE & Mgmt in 2011
- Went from MASc to PhD in 2013
- Have a fair amount of teaching experience
  - ChE 4E03, 4N04, 2/3E04, 3G04, 4G03, 4H03, ENG 1C03, IBEHS 4A03

- I am <u>always</u> happy to meet
  - Make an appointment at neasej@mcmaster.ca
  - I can be reached at **905-599-3165** or **x27337**
  - Fire me a **teams message** too if you prefer – we can video chat!
  - **I am also the undergrad chair** for chem eng!
- Office is BSB/B105… If my door is open, you are welcome to come in
  - I guess this is for future years…?
  - This invitation extends beyond 4H03, if you need!

# Teaching Assistants

## Yingkai Song

- songy65@mcmaster.ca
- Room JHE/370
- Currently a PhD student with Dr. Khan
  - Developing techniques from applied mathematics for improved methods of global dynamic optimization

## Mahir Jalanko

- jalankmd@mcmaster.ca
- Room JHE/370
- Currently doing a PhD with Drs. Mahalec and Mhaskar
  - Improving control strategies for ethylene splitters using machine learning tools

*TA office hours are by appointment*

# Some Procedural Advice

- Read the course outline **carefully**
  - There is a significant amount of information there
  - The TAs and I will answer emails promptly
- If you have **questions**:
  - Email the TAs with me CCd
  - If we can't answer your question, we will set up a meeting as soon as possible
- Please email from your McMaster email address
  - Moreover, please use the che4h3instructor address if possible

# References and Readings

- Tons of resources will be posted on A2L – You are encouraged to consult them at your leisure and bring to class whatever you wish

- Examples of course reading materials include:
  - Old slides from ConnectMV
  - Tons of data sets
  - Hyperlinks to software downloads
  - Software tutorials and guides
  - Computation guides and lessons

- There is no textbook for this course
  - Any machine learning textbook will do
  - Course notes will have sufficient information and references

# Software Tools and Website

- This course will use several **tools** for modeling
  - **MATLAB** – Language for scientific computing
    - Used for PCA/PLS and ANNs
  - **EXCEL** – Critical for data manipulation, visualization
  - **OTHERS** – You may CHOOSE some languages
- This course will use Avenue to Learn for announcements, posts, *etc*.

Welcome to ChE 4H03!

Jake Nease posted on Jan 3, 2021 4:50 PM

Hi all!

Happy new year! I hope you had a chance to get some well-deserved rest over the last couple of weeks and you are ready to take down another term of online learning. For many of you, this may be your last term at McMaster, so I hope we can make it a good one!

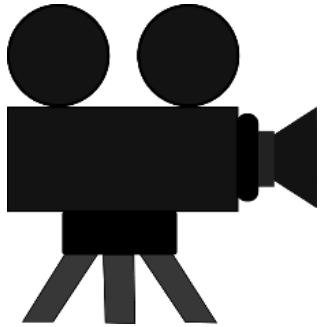
I am thrilled to have the chance to offer 4H03 this term. We had the chance to hit our stride last year with its first-ever offering and I learned a great deal about what is important to students with regards to modeling. I am excited to use that experience to take 4H another step forward this year. We still have some flexibility for coverage, so I will be sure to poll the class for what interests you the most.

I have attached a letter to this announcement detailing some of the important course policies for the term. I will also add the syllabus here over the next day or two. Please check out the letter to get a glimpse of what we are planning for the course.

https://pbs.twimg.com/profile_images/2323205264/h5g3etjnacmazg8oq17z_400x400.jpeg

http://seeklogo.com/images/E/excel-logo-974BFF9CB9-seeklogo.com.png

# Recordings

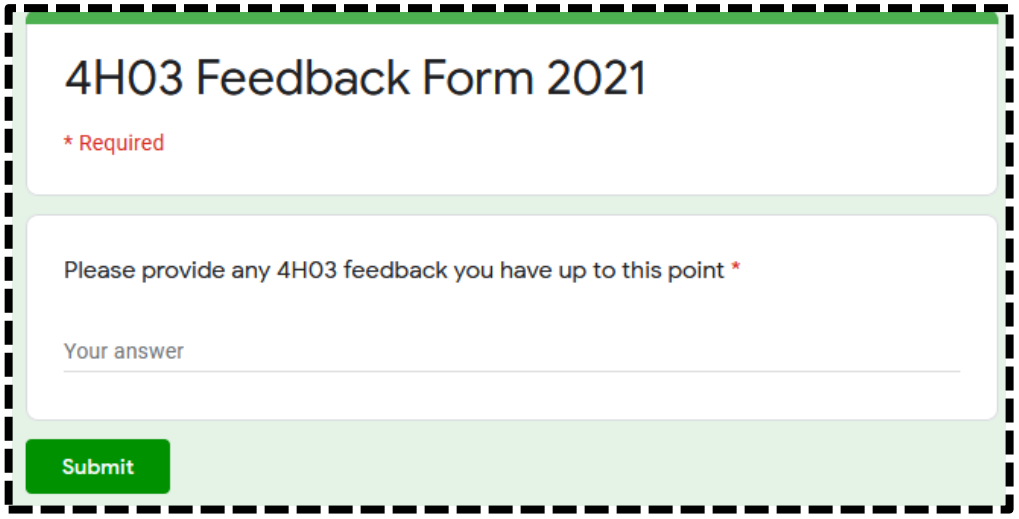


- I will record all classes on video
  - Could be useful if you miss class
  - **No guarantees** on quality
  - Usually available **by the end of the week**
  - All recordings will be posted in the content section on A2L as YouTube links
  - The built-in MS Teams capture software will be used. Your comments in the chat thread will not be recorded

- Please do not use these videos in order to skip class!

# Feedback and Comments

- Comments and feedback are always welcome
  - DO NOT wait for course evaluations!
  - Early suggestions are more important to you and I
  - Feel free to shoot me an email
  - Anonymous comments can be made via this online survey



McMaster University ENGINEERING

# Course Snapshot

## What did I get myself into?

# Scientific Computing

- According to Wikipedia:
  - Scientific computing is the field of study concerned with constructing mathematical models, quantitative analysis techniques and using computers to analyse and solve scientific problems

- **Applications of Scientific Computing**
  - Numerical simulations
    - Reconstruct (and *understand*) known events
    - *Predict* future or unobserved situations
  - **Model fitting and data analysis**
    - Appropriately tune models and equations to reflect reality or physical observations
  - Optimization
    - Do things the best way possible (cheapest, fastest, easiest, *etc*)

https://upload.wikimedia.org/wikipedia/en/thumb/8/80/Wikipedia-logo-v2.svg/1122px-Wikipedia-logo-v2.svg.png

# Learning Objectives

- Demonstrate the concept of identifying the **best model** to explain a data set
- Fit and compute model parameters for principal component analysis (PCA) and partial-least squares (PLS)
- Compare the trade-offs between model **accuracy** and **computational effort**
- Demonstrate the ability to identify if a model is over- or under-fit using statistical significance metrics
- Visualize large data sets to identify trends and key observations
- Provide context to a data set so that a lay-person could understand the key conclusions
- Demonstrate a core understanding of the fundamental background theory behind various big-data driven models and artificial intelligence methods
- Identify **misleading results,** apply appropriate analyses to **judge their accuracy** or applicability, and suggest more appropriate alternatives

# Grading

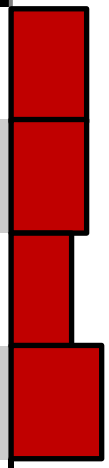
- So, we have the ability to be creative here
- Let's check out some options…

| Weight | Component | Comments |
|---|---|---|
| ???% | Assignments | ??? |
| ???% | Midterm Test(s) | ??? |
| ???% | Course Project | ??? |
| ???% | Final Exam | ??? |
| ???% | Participation | ??? |

# Grading Option 1: The *Classic*

- Feels like a warm hug from an old friend

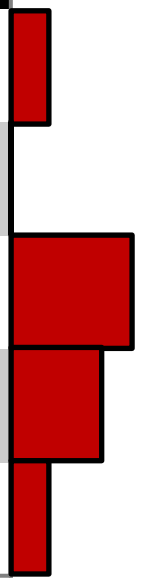
| Weight | Component | Comments |
|---|---|---|
| 25% | Assignments | 5 Assignments at 5% each. Probably EOW and one for each major topic. |
| 25% | Midterm Test(s) | Two tests at 12.5% each, also focusing on major topics in the course. |
| 20% | Course Project | One project done in groups, applying course concepts to a data set of your choice. |
| 30% | Final Exam | A final exam scheduled by the Registrar accounting for whatever is left. |
| 0% | Participation | Better for those of us that do not want to come to class. |



# Grading Option 2: The *Graduate*

- Graduate classes are fun
- More self-directed and research, less testing

| Weight | Component | Comments |
|---|---|---|
| **15%** | **Assignments** | 1-2% each, and just participation-based. |
| **0%** | **Midterm Test(s)** | We'll skip the tests. Who likes tests? |
| **40%** | **Course Projects** | Two projects – each on one method presented in each "half" of the course. |
| **30%** | **Final Exam** | A final take-home problem set. |
| **15%** | **Participation*** | **Friday presentations**. Randomly generated groups will give 10-minute paper reviews three times. Includes peer evals. |

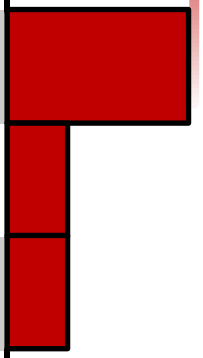


*This may require some scheduling wizardry for those of us on co-op

# Grading Option 3: The *Minnick*

- My office neighbour is Matt Minnick.
- Matt Minnick likes weekly tests.

| Weight | Component | Comments |
|---|---|---|
| 0% | **Assignments** | Not going to be used in this scheme. |
| 60% | **Midterm Test(s)** | Six tests written approximately biweekly. 10% each. |
| 20% | **Course Project** | One project done in groups, applying course concepts to a data set of your choice. |
| 20% | **Final Exam** | A final take-home problem set. |
| 0% | **Participation** | Better for those of us that do not want to come to class. |



# Grading Option 4 – Open to Suggestions

- Feel free to email me a concoction of your own
- **We must decide this by the end of the first week**

| Weight | Component | Comments |
|---|---|---|
| ???% | Assignments | ??? |
| ???% | Midterm Test(s) | ??? |
| ???% | Course Project | ??? |
| ???% | Final Exam | ??? |
| ???% | Participation | ??? |

# Class Structure

- Weekly lecture structure will largely depend on grading scheme chosen
  - Either way, will likely include some [light] readings
- This course will primarily use .ppt slides
  - Depending on resources chosen, we might mix thing up
- Please **be ready to participate** whenever possible
  - We may do some in-class demos
  - Following along is not mandatory, but will likely be useful

# Project

- The ONLY thing that is necessary in the grading scheme is a term-long project
  - A great way to develop application skills!
  - Project will be done in groups
  - Auditing students encouraged to participate

- Some important notes (details to follow)
  - Proposal due mid-Feb
  - Presentations toward end of term
    - TBD optimally based on your schedules
  - Reports due on the last day of class
    - 12 pages MAX

# Tests and Examinations

- MIDTERMS will be take-home problems
  - I will NOT use proctoring software
  - Coverage depends on grading scheme chosen
  - Likely will have 2-3 days to complete a short set of problems individually
- FINAL EXAM will be released in April
  - Coverage depends on grading scheme
  - Will also be take-home problems
- Midterm(s) and final:
  - By virtue of online learning, all tests will be “open everything”
  - Caveat: I could ask you to use software to generate plots, gather data, *etc*.

# Tentative Course Schedule

| Week | Anticipated Topics | Anticipated Module Content (subject to change) |
|---|---|---|
| **Week 01** | Course Overview<br>Review of Modeling Core Concepts | • Course overview<br>• Basic stats review<br>• Sampling methods<br>• Language and lingo<br>• Visualizing data |
| **Week 02** | Multivariate Regression<br>Basis Function Regression | • Review of SSE optimization<br>• Goodness of fit<br>• Training/testing sets<br>• Weakness: dependent sets of data |
| **Week 03** | Dimension Reduction (Latent Variable Methods) | • Principal Component Analysis (PCA)<br>• Purposes of PCA<br>• Defining loadings, scores, and visualization |
| **Week 04** | Dimension Reduction (Latent Variable Methods) | • PCA derivation and the NIPALS algorithm<br>• Interpreting scores<br>• Application: Sports data mining<br>• Application: Soft sensors and equipment health |

# Tentative Course Schedule

| Week | Anticipated Topics | Anticipated Module Content (subject to change) |
|---|---|---|
| **Week 05** | Dimension Reduction (Latent Variable Methods) | • Projection of Latent Structures (PLS)<br>• Modeling with PLS<br>• Application: Spectral data |
| **Week 06** | Data Clustering | • K-Means Clustering (algorithm and interpretation)<br>• Integrated clustering and visualization in the reduced-dimension space |
| **RW** | Reading Week | • Suggested Reading: "Infinite Jest" by David Foster Wallace. |
| **Week 07** | Data Clustering | • Expectation Maximization Clustering<br>• Support vector machines |
| **Week 08** | Machine Learning Tools | • Background and core concepts<br>• Supervised learning (AKA regression!) |

# Tentative Course Schedule

| Week | Anticipated Topics | Anticipated Module Content (subject to change) |
|---|---|---|
| **Week 09** | Machine Learning Tools | • Artificial Neural Networks<br>• Single-layer networks<br>• Basis functions |
| **Week 10** | Machine Learning Tools | • Multi-layered networks<br>• Examples and analysis |
| **Week 11** | Machine Learning Tools | • Support Vector Machines for categorization<br>• Predicting sample classifications |
| **Week 12** | Boolean Decisions | • Decision Trees<br>• Random Forests |

# Interesting Data

- Millau Viaduct: connects Paris and Barcelona

*https://divisare-res.cloudinary.com/images/c_limit,f_auto,h_2000,q_auto,w_3000/v1/project_images/5055521/1158_FP207418/foster-partners-nigel-young-millau-viaduct.jpg*

# Interesting Data

- Millau Viaduct: connects Paris and Barcelona

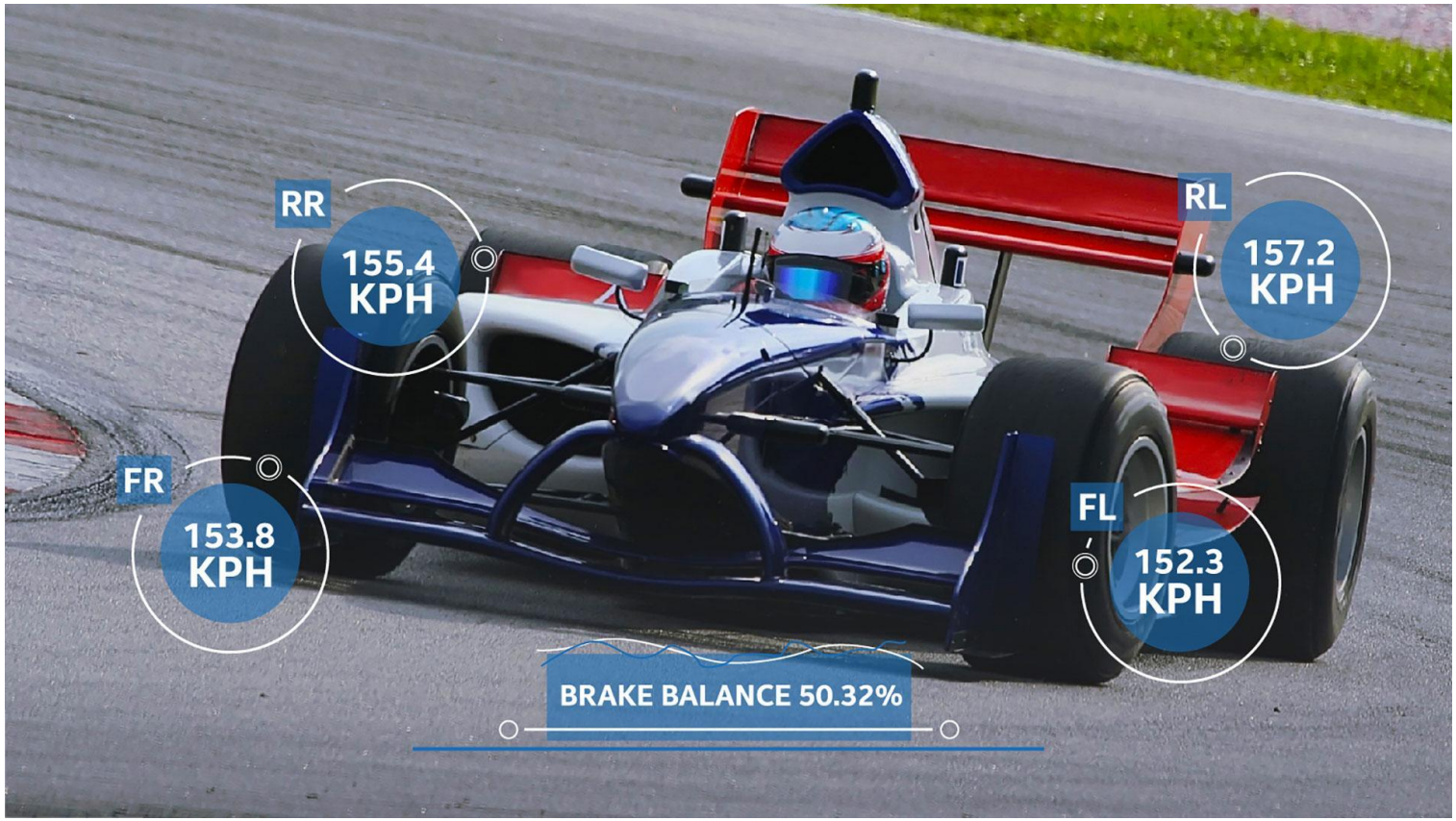
*https://akm-img-a-in.tosshub.com/indiatoday/images/story/201612/millau-viaduct1-647_121416042052.jpg*

# Interesting Data

- Millau Viaduct: connects Paris and Barcelona
  - Pylons, masts, deck *etc*. all equipped with accelerometers, inclinometers, temperature sensors…
  - Detect movement at the $\mu$m level and detect for oscillations, stresses, strains…
  - Additional sensors gather traffic, weight
  - *Machine learning* allows it to interpret up to 14 different types of vehicles

# Interesting Data

- Formula 1 Race Cars
  - Data collected and interpreted by Intel®




https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/big-data-powers-f1.html

# Interesting Data

- Formula 1 Race Cars
  - Cars can contain 100-200 sensors
  - 2 GB of data transmitted wirelessly per lap
  - 3 TB data transmitted over a race (some data stored locally)
  - Can help diagnose problems before they arise (soft sensor)
  - Pit team can advise driving strategies based on data

- SPORTS data is actually huge
  - I mean, *Moneyball* was an epic movie

https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/big-data-powers-f1.html

# Interesting Data

- Chemical plants
  - TONS of (redundant!) sensors in chemical plants
  - Classics:
    - Pressures
    - Temperatures
    - Flows
  - New-Age:
    - Cameras
    - Acoustics (Hassan can tell you about this!)
    - Spectral data
- Facial recognition
  - Home monitoring, device security…

# Interesting Data

- Financial data
  - Predicting loan default probabilities
  - Mortgage stress-tests
  - Machine learning and decision trees applied to loan approval
- User scores
  - NETFLIX actually released data in 2011 for users to develop a learning recommendation platform
  - Was abandoned due to privacy concerns (obviously)
- Kaggle.com is FULL of amazing data sets
  - Could be good for your course project

# Interesting Data – Previous Projects

- X-Ray and medical imagery
  - Predicting age and gender of children age 2-10 based on bone structure of hands using convolutional neural nets
  - Using MRI data on breast cancer images to identify malignant and benign tumors
- Sports data
  - Determining the "composition" of a winning NBA team using PCA and PLS
- Demographic and weather data
  - Relating demographic data to student performance
  - Modeling dispersion of pollutants based on weather patterns

# Final Words

- I am really excited for 4H03!

- Hopefully you will find this course informative, useful, and interesting
  - I am anticipating a great learning experience with you