Summary

Arthur J. Redfern
arthur.redfern@utdallas.edu

Big Picture

Why Study This Topic?

- Almost any problem you can think of can be cast as 1 of 2 types of problems
  - Extraction of information from data (e.g., classification or regression)
  - Generation of data from information
- These 2 problem types can be put into a general framework
  - Pre processing - optional domain specific processing
  - Information extraction or data generation - input output mapping
  - Post processing - optional domain specific processing
- xNNs do a great job at input output mapping for many problems of interest
  - xNNs are universal approximators so if a mapping exists they can approximate it (under some relatively mild technical conditions)
  - So potentially they can be used to solve almost any problem you can think of
  - And in practice they're a key component of the top performing methods for many important tasks

What's The Problem With This?

- We don't know the input output mapping that we need
  - It's not given what the optimal network structure is to approximate the input output mapping (design) and how to find the parameters that control the network structure mapping (training)
  - It's not given how to practically and efficiently realize network structures (implementation)
  - It's not given what different applications require in terms of pre and post processing and along with different design and training strategies (vision, speech, language, games, art, ...)

The Purpose Of This Course

- So the obvious purpose of this course is
  - To review the basic math needed for pre processing, xNNs and post processing
  - To learn network design strategies to accomplish different goals
  - To learn end to end network training from data
  - To learn efficient network implementations
  - To learn specific requirements of common applications

- However, an equally important purpose of this course is to learn how to learn
  - The last paper isn’t written, there are many more each day
  - You have a good base to work from
  - You have the potential to understand new ideas from others and create new ideas yourselves
  - Your ability to do this improves with practice

Topics We Covered

We Covered A Lot Of Material

That's by design: if you want to do meaningful work in this area, it helps to know a lot about a lot

- The first 3 topics in the linear algebra class were sets, fields and vectors
- The last 4 flagship topics in the application classes were object based image segmentation, speech to text transduction, language to language translation and playing Atari and Go
- And somewhere in the middle we talked a little bit about semiconductor device physics

The Course Was Divided Into 3 Sections

- Math
    - Linear algebra
    - Calculus
    - Probability / info theory
    - Algorithms
- Networks
    - Design
    - Training
    - Implementation
- Applications
    - Vision
    - Speech
    - Language
    - Games
    - Art

Homework

- 1 purpose was to complement in class material through reading and expand theory through problems

- Another purpose was to provide an opportunity for practical experience
    - TensorFlow and the Colab environment by default
    - Templates to provide a starting point and connect to theory
    - A gradual ramping up in terms of complexity to build confidence
    - Some amount of training pain to better appreciate why we discuss implementations

Topics We Didn’t Cover

There’s A Lot Of Material We Didn’t Cover

- Not sure if this is good or bad
  - Perhaps it’s just the reality of a large field viewed through the lens of a 1 semester class
- Realize that these uncovered topics are everywhere
  - Within the topics areas that we covered
  - Topics that we didn’t cover
- Realize that there are new developments every day in this field
- Again, your only hope is to learn how to learn
  - Hopefully this course contributes positively to your ability to do that
  - In general, the more stuff you know the easier it is to learn something new (as it’s likely similar to something you already know)

Ex: Networks

- While it feels like we covered a lot of network topics
  - Layers, design, training, implementation
- There’s a lot more that we didn’t cover
  - For example, networks with an external memory developed in the context of language applications, interesting to think about in the context of other applications
- On the upside, the network information presented during this semester should provide a good basis for understanding new network designs

Ex: Vision, Speech, Language And Games

- While it feels like we covered a lot of vision topics
  - Classification, pixel segmentation, multiple object detection, object based segmentation, depth estimation and motion estimation
- There's a lot more that we didn't cover
  - Variations within the topics that we covered
  - Other topics that we didn't cover like point clouds, facial recognition, handwriting recognition, optimizing object boundaries, ...
- The same observation is true for speech, language and games
- On the upside, the vision, speech, language and games info presented during this semester should provide a good basis for understanding different topics in these fields

Ex: Art

- A lot of art related work is leading to interesting network structures for generative systems
    - Style transfer
    - GANs
    - Music, images and video
- My goal is to add this topic to the course next semester
    - It will be posted to GitHub if you would like to see it
    - Currently just in outline format
    - Alternatively, if you're interested in this you should consider teaching yourself (all the information is out there, it just takes time to go through and organize it for understanding)
    - No guarantees on any of this though

The Future

Interpolation

These are pretty easy to guess

- Gradually more and more theoretical questions are answered
    - People will quit complaining about a lack of theory
    - Note: there's already a lot of theory and the complaining seems unreasonable even now

- Everything gets better in proportion to gradual advances in computing
    - Maybe slightly faster than process technology but not radically faster
    - Will get a boost for going to lower precisions and building more efficient architectures

- Basic methods are applied to more and more fields
    - Just like there was computational <science> that became <science>, there will be machine learning for <science> that eventually becomes commonplace enough that it's again called <science>
    - It's likely that some developments in some of these fields will propagate back to the core and become used in many fields
    - In general there will be convergence towards a few core techniques with some specialization (e.g., pre and post processing) around them

Extrapolation

These are comments and guesses with a much higher variance and by choice are somewhat constrained

- Theory
  - It would radically change what we're capable of if algorithms are developed with processing complexity proportional to the information rate vs the data rate
- Computation
  - Over the course of a longer time scale what happens if compute capability increased by 1000? 1000000? 1000000000?
  - What happens if a different computing paradigm (maybe quantum) becoming commonplace and let's us design algorithms that solve complementary problems and providing another fundamental step up in AI
- Applications
  - We figure out how to handle more and more complex problems by putting together more and more simple problems
  - To an outsider this starts to feel like general intelligence vs specific / isolated problem solving
- Humanity
  - We become much more closely integrated with AI; everyone's hand is forced in this direction or they're left out / disadvantaged
  - We observe a person over the course of their life and get good at simulating them
  - We question our own intelligence

Your Place In It

- A few suggestions
  - Add a new hobby every few years
  - Figure out what’s at the intersection of what you love to do and what you’re good at
  - Do that
  - Be great (know you can be)
  - Don’t be afraid to change

Next Time

Logistic Changes

- Change the course name to deep learning as that better reflects the content
  - This may have to wait for the next next time (need to chat with the administration)

- Budget 2 classes in the schedule for
  - Big picture white board discussion right after the introduction lecture
    - Data – forward pass – error – backward pass – weight update
    - Applications
  - Software implementation in TensorFlow

- Complete the Art lecture and maybe add an Art class (I need to see how it fits in the schedule)

Content Refinements

- Continue to improve the code examples and homework
  - Improve code examples for design, training, implementation, vision, speech, language, games and art
  - Likely switch over to TensorFlow 2.0 and Keras at that point in time
  - Figure out ways to improve training efficiency
- Continue to improve the slides
  - I have a few hundred notes on smaller refinements to make
  - It would also be nice to do some general notational cleanup in a few places
  - And I have a relatively large re write / streamlining of the implementation content planned
- Continue to look for ways to use the board vs slides whenever possible

Consider An Expanded Offering

Organization changes to make it a 4 credit 1 semester or 3 credit 2 semester class

- Math
  - Linear algebra
  - Calculus
  - Probability
  - Algorithms
- Network
  - Design
  - Implementation
- Understanding
  - Supervised learning
  - Images
  - Sounds
  - Words
- Acting
  - Reinforcement learning
  - Games
  - Controls
- Creating
  - Unsupervised learning
  - Images
  - Sounds
  - Words

One More Thing

Thank You!

- You’re all special to me and I appreciate you sticking with me though this
- I wish you all the best for happiness in your lives and careers
  - You’re capable of great things
  - I look forward to seeing what you accomplish academically and professionally
- Keep in touch!

The Final Slide

Number of slides

Introduction	56
Linear algebra	111
Algorithms	32
Probability	87
Calculus	55
Design	158
Training	152
Implementation	185
Vision	114
Language	91
Speech	122
Games	95
Art	20
Summary	25
Total	**1303**