Topics in Deep Learning - Healthcare, medicine, and the life sciences

Note on course structure: We have a total of 10, 2.5 hour sessions. I suggest we break up the sessions in (roughly): 6-7 ML/DL in healthcare, 2-3 in conformal and model explainability, and 1 for the business pitch

Below are a list of my own publications and which topics/lectures they would work well for
- (1)-2 [article] Implementing AI in Healthcare
- (1)-3 [article] Showing how performance can be misleading (pitfall)
- (1)-4 [dataset] SurvSet
- (1)-6 [article] Example of RWE

(1) Healthcare, medicine, and the life sciences

Lectures
1. Overview: the practice of healthcare, medicine, and life sciences (Goal: Provide a the "landscape" of how ML/AI can be used in HC, and what its applications are, and why the sector hasn't yet been "transformed" or "disrupted")
    a. [lecture notes] This MIT courses first lecture has a lot of good materials on the history, what makes HC unique, etc
    b. [lecture notes] Similar "what makes HC unique" but Marzeh's class
    c. [article] NNets in Clinical medicine from 1996!!! This seems new, but it's not, foreshadowing that this stuff is not easy
    d. [report] **AI has been lagging in HC -** but the upside is very large (see Figure 1 and 6)
    e. [report, article] Potential impacts and applications (Figure 2 for example of hospital applications
    f. High profile examples of failed applications of AI in HC
        i. https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html
        ii. https://www.fiercehealthcare.com/tech/epic-s-widely-used-sepsis-prediction-model-falls-short-among-michigan-medicine-patients
        iii. https://www.wired.co.uk/article/babylon-disrupted-uk-health-system-then-left
2. Implementing AI in healthcare #1 (Goal: Describe what makes healthcare unique, and why it's been so challenging to get these systems adopted)
    a. [article] AI on the front lines
    b. [article] Implementing AI in Healthcare
    c. [article] The Human Body if a Black Box

    d. [article] Do no harm: a roadmap for responsible machine learning for health care
    e. Duke's DIHI is the gold standard for how to integrate data science tools into actual practice
        i. [article] Accelerating health system innovations
        ii. [article] Organizational Governance of Emerging Technologies
        iii. [video] Workflow Integration of Machine Learning Solutions
    f. [article] Implementing machine learning in medicine
3. Implementing AI in healthcare #2 (Goal: Focus on the key pitfalls including bias, risk, and generalization)
    a. [article] Google's diabetic retinopathy (partial) flop - great example of building the solution first, then figuring out how to use it
    b. [article] Barriers to Achieving Economies of Scale in Analysis of EHR Data
    c. [video] Beware of **alert fatigue**
    d. UPenn's sepsis model shows how a "good model" doesn't necessarily move the dial
        i. [article] A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock
        ii. [article] Clinician Perception of a Machine Learning-Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock
4. Survival modeling (Goal: Understand what sort of data is appropriate for survival modeling, and what the different algorithms there are to support this)
    a. [Intro] Lifelines package
    b. [Evaluation metrics] scikit-survival package
    c. [Modeling] See parametric models and CoxPH (from the PySurival package)
    d. [Landscape] Excellent overview of all the topics on ML for Survival, up to Slide 51 covers most of the key elements: censoring, concordance, KM curves, etc
    e. [dataset] SurvSet
5. Protein folding, drug discovery, and medical imaging, and 'Omics (Goal: show the cool range of application for "cutting edge" science)
    a. [Article] Polygenic risk scores from risk scores and stratification
    b. [lecture notes] GWAS
    c. [Video] GWAS, Linkage, Fine-Mapping
    d. [lecture notes] DL for drug discovery
    e. [lecture notes] Protein folding
    f. [lecture notes] Imaging topic from csc2541
6. Commercial applications (Goal: show how predictive modeling is actually being used for bread-and-butter commercial applications for thinking like patient finding or advertising)
    a. [Article 1, 2] How Theratyping expanded market access for cystic fibrosis
    b. [Article] How RWE is shaping regulatory decisions (expanded indications, comparative effectiveness, post-marketing surveillance)
    c. Patient finding and clinical trials
        i. Connecting Patients to Clinical Trials
        ii. Forward Thinking for the Integration of AI into Clinical Trials
    d. [article] Patient and HCP segmentation

    e. [article 1, 2] "Real-world evidence"
    f. [article] Next-best actions
    g. [article] MLHOps - architecture for hosting and maintaining a ML model in a HC organization

(2) Conformal prediction & model explainability

Lectures

1. "White box" model explainability (Goal: understand the classical problem of statistical testing and how we can extend this to "double dipping" methods of POSI)
    a. Statistical testing (e.g. p-values and confidence intervals)
        i. [lecture notes] Exact CIs
        ii. [lecture notes 1, 2] See section 5 from either
    b. Multiple hypothesis correction
        i. [lecture notes, blog post w/ notebook]
    c. Selective inference (aka post-selection inference)
        i. [course notes 1]
        ii. [article] PNAS
        iii. [math 1, 2]
        iv. [video]
        v. [python package]
2. "Black box" model explainability (Goal: understand the different methods for "understanding" how a non-linear model works for wrappers on black-box approaches)
    a. Local methods (more examples here)
        i. LIME
        ii. SHAP
    b. Global methods (more examples in here)
        i. HRT (vs Permutation)
        ii. Knock-offs
    c. General references [1, 2]
3. "Prediction intervals" (Goal: understand how to provide quantifications of uncertainty around the actual a prediction - i.e. moving beyond a "point prediction")
    a. Classical PI for (G)LMs
        i. [ciTools] Technically R-based, but has the useful math and description
    b. Conformal regression, classification, and conditional conformal
        i. [Article] Gentle intro
        ii. [Notebooks] Gentle intro
        iii. [Video 1, 2, 3] Gentle Intro
        iv. [Lecture and notebooks] Tutorial on Conformal Prediction
        v. [Video 1, 2] More technical details
        vi. [Book] Practical guide
        vii. [Collection] Resources

(3) Business thinking and organizational structure

Lectures

1. All things business (Goal: understand what are the key factors in successful HC companies powered by AI, how they can express their comparative advantage - the “pitch” - and organizational structures that enable data science to work at an enterprise level)
    a. Business pitch
        i. Writing Business Plans for a Life Science Startup or Clinical Program
        ii. Pitch deck
    b. [video, slides] Creating a “triple win” for innovative ideas (strategic planning)
    c. [blog] Do organizations need an AI Center for Excellence?
    d. Successful HC/life science startups with AI part of their core business model
        i. Flatiron health
        ii. Tempus Labs
        iii. 23andMe
        iv. Recursion Pharma

Original description found on website

In this course, participants will apply machine learning techniques to software applications using the Python language. Successful participants will apply each step of the machine learning workflow in relevant industry applications, as well as be exposed to cutting-edge techniques and the underlying theory. Technical topics include: Machine Learning in Practice, Data Preparation and Feature Engineering, Supervised and Unsupervised Learning, Model Evaluation, and Ensemble Learning. Participants will also work on other skills, such as networking and building a community of influence, mastering the interview process, including behavioral and technical interviews.

Learning Outcomes

- *Propose and present a business pitch for a machine learning project in a real-world business setting through a business case study with an industry partner or at their current organization.*
- *Design, formulate, and construct a full comprehensive lifecycle project in machine learning via a course project by managing a timeline*
- *Apply, deploy, and implement each step of the machine learning lifecycle in Python programming language, and debug errors and iterate on improvements.*

Course delivery

Three weeks of instructor-led live webinars, each lasting 2.5 hours (totaling 25 hours)