Introduction to Machine Learning for Longitudinal Medical Data

Orlando Doehring, Ph.D.

Unit 2, 2A Bollo Lane, London W4 5LE, UK

orlando.doehring@phastar.com

Machine learning for healthcare 1

- Machine Learning (ML): Learning from data to extract structures/patterns and make predictions
- Aim: Briefly illustrate 3 ML models for sequential healthcare data along with a case-study for multi-label disease prediction from clinical measurements
- Many medical records are sequential in nature
  - Longitudinal data, e.g. blood pressure indexed by time
  - Spatial data, e.g. imaging data which forms a sequence of coloured 3D RGB pixels
  - Natural Language Processing data, e.g. a list of words in a paragraph

[References: 1-3]

Machine learning for healthcare 2

- Previously, focus on shallow learning techniques for regression/classification without focus on sequential data, such as
  - Logistic Regression: Generalized Linear Model using a Logit link function
  - Support Vector Machines: Regression/Classification for non-linear data based on ideas from statistical learning theory
- Models for sequential data we consider are either:
  - linked (hidden + observed) Markov Chains, such as Hidden Markov Models (HMM), or
  - layered Neural Networks (NN), such as Recurrent Neural Networks (RNNs) for natural language processing or Convolutional Neural Network (CNNs) for image classification

[References: 1-3]

Nature of healthcare datasets

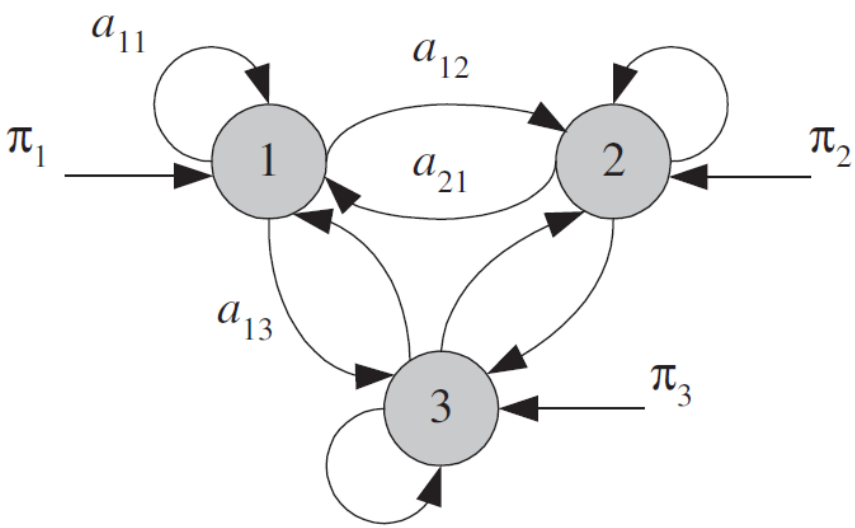
- Surge in biomedical datasets, such as
  - Bioinformatics: Disease diagnosis from microarray data, drug discovery from molecular compounds
  - Medical imaging: Brain reconstruction, organ segmentation, tumour detection
  - Sensing: Anomaly detection, human activity recognition from images, wearable devices
  - Public health: Prediction of epidemic alerts from social media data and meta-information in mobile devices
  - Healthcare Informatics: Length of hospital stay, unexpected readmission probability within next days, mortality prediction from electronic health records (EHR)

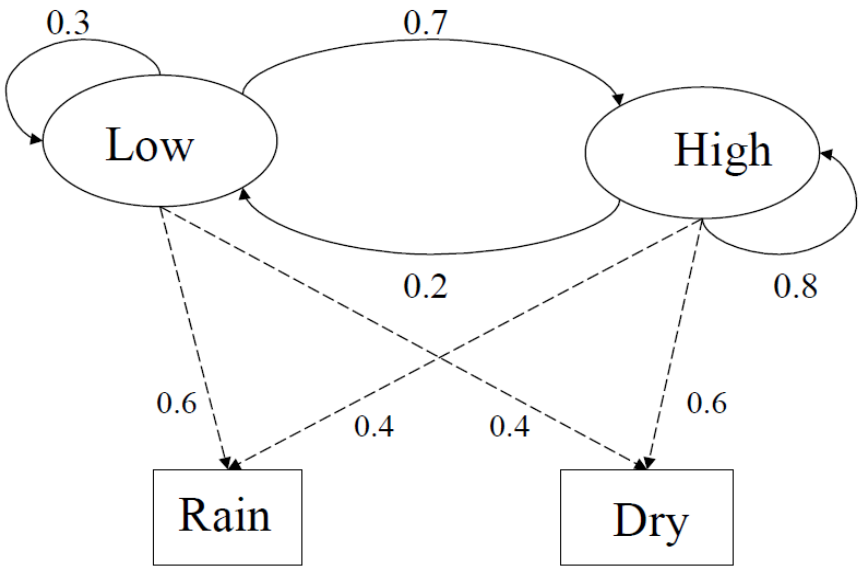
[References: 1-2]

Hidden Markov Models 1

- Two linked 1st order Markov models:
  - latent states (atmospheric pressure low/high)
  - output states (weather: rain/dry)
  - Example: Output sequence O of hourly weather observations

    O = [Rain, Rain, Dry, Rain, Dry] based on last 5 hours





[Reference: 3]

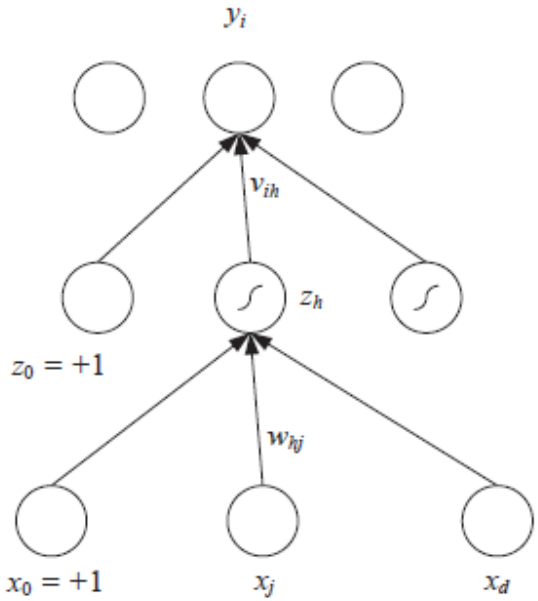
Hidden Markov Models 2

- Three basic learning problems in HMM:
  1. Given model, estimate probability of observed sequence.
  2. Given the model and observed sequence which sequence of hidden states has the highest probability? Viterbi algorithm
  3. Given observed sequences estimate initial, transition and emission probability? Baum-Welch algorithm

[Reference: 3]

Deep Learning: Neural Networks (NNs)

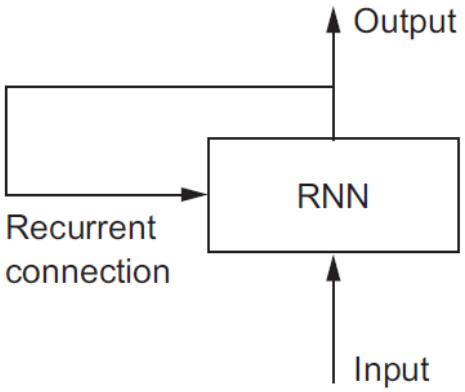
- Single-layer NN / Perceptron: Weighted sum of (dense) neuron variables followed by non-linear activation function, e.g. if zh = logit activation (link) function then logistic regression
- Multi-layer Perceptron (MLP): Several layers of perceptrons stacked onto each other; estimated via back-propagation (chain differentiation rule applied to composite function)

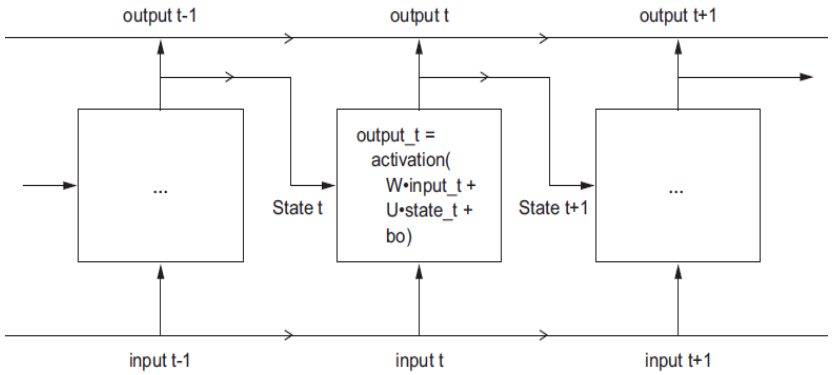


[Reference: 3]

Recurrent Neural Networks

- Application Examples:
  - Time series classification, such as topic or author identification of medical article or sentiment analysis (positive, neutral, negative) of patient feedback
  - Time series comparisons, such as how closely related two patients are
- Similar to MLPs but on top of neuron variable there is a state variable fed into z_h
  - State variable at time t+1 is the output of z_h at time t, i.e. z_h is applied to weighted in input sum plus weighted state sum (RNN unrolled over time)

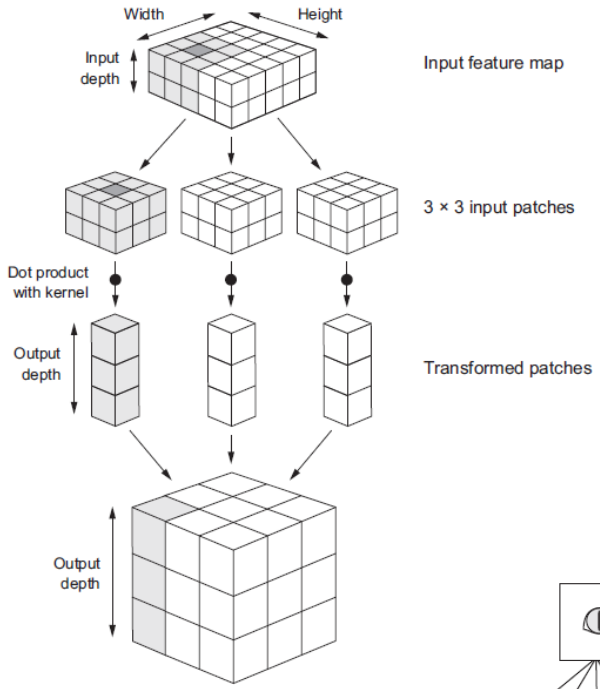


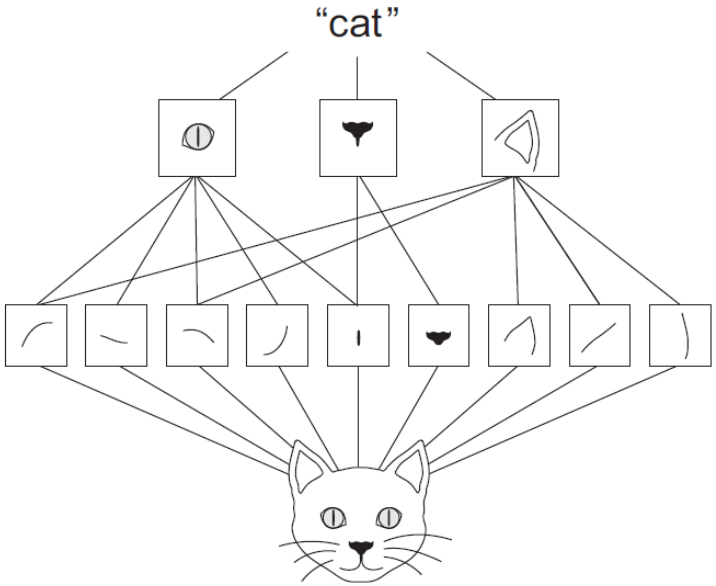


[References: 4-6]

Convolutional Neural Network 1

- While densely connected MLPs learn global patterns CNNs learn increasingly complex spatial hierarchies of translation-invariant local patterns (e.g. filters of 3x3) which require fewer data points and offer more generalization power
  - First layers: style (edges, textures, colours)
  - Higher layers: content, higher-level macro structure of the image





[Reference: 6]

Convolutional Neural Network 2 (Neural style transfer)

- Style: Blue-and-yellow circular brushstrokes from “Starry Night” by Vincent Van Gogh
- Content: Buildings in city of Tübingen, Germany

[References: 6]

Pros/Cons: HMMs vs. NNs 1

- HMMs are generative sequence models based on discrete-time Markov Chains:
- Given performance metric [e.g. Accuracy], HMMs can be estimated with fewer data points and it converges more quickly [PRO]
- Scales quadratically with number of hidden states, such that inclusion of several previous time steps in state space quickly increases complexity [CONS]
- Regular HMMs: Difficult to model long-range dependencies due to limited memory [CONS]

[References: 3-6]

Pros/Cons: HMMs vs. NNs 2

- Neural Networks are discriminative models for classification/regression:
  - RNNs explicitly model long-range dependencies; e.g. for speech recognition [PRO]
  - CNNs build a local hierarchy from low-level features (style) to high-level (content/objects) for computer vision [PRO]
  - Non-linear activation functions can model complex behaviour [PRO]
  - Models often need much training data and lots of computing power (GPU farms) for good results [CONS]
  - But could transfer pre-trained models from other tasks in same domain [PRO]

[References: 3-6]

Learning to Diagnose with LSTM Recurrent Neural Networks 1

- Data: Multivariate time series from paediatric intensive care unit (PICU) episodes at Children's Hospital LA
- Each patient visit: Sensor / lab data recorded in Electronic Health Record (EHR) of varying length sequences (hours to several month), irregular sampling, missing data, long-range dependencies
- Each episode: Multivariate time-series of 13 variables, such as diastolic/systolic blood pressure, heart rate, pH, respiratory rate, body temperature, etc.

[References: 7-9]

Learning to Diagnose with LSTM Recurrent Neural Networks 2

- Response: Multi-label, i.e. zero or more diagnoses (128 most common ones out of 429), such as diabetes, asthma, neoplasm, respiratory distress syndrome
- Data quality: Lots of free text with varying quality, detail, correctness; limited data sharing leads to missing data (e.g. partial medication list); unclear patience compliance with treatment regime; misdiagnoses of patients; data falsification
- Method: Long short-term memory (LSTM, variant of RNN) regularized (e.g. dropout to avoid overfitting) with two layers of 128 neurons

[References: 7-9]

Learning to Diagnose with LSTM Recurrent Neural Networks 3

- Experiments: Train on 80% of episodes, 10% for validation / model selection, 10% for testing
- Metrics: Extensions of AUC [TPR (Sensitivity) vs. FPR (1-TNR=1-Specificity)] to multiple labels
- Results: LSTM trained on raw data about the same as a MLP with expert features chosen by clinicians (85% AUC) [3] but typically 5-10% better than logistic regression

[References: 7-9]

Summary

- Covered 3 models: HMMs, RNNs and CNNs to model sequential healthcare data, such
  - as annotation of multi-label disease responses [Longitudinal] ,
  - tumour detection in images [3D RGB Images],
  - evaluation of written/spoken patient feedback [Sentiment in Natural Language Processing]
- NNs can model more complex data but requires more computation power or pre-trained models than HMMs
- NNs can deal with complex sequential data structures with long-ranging dependencies and observations recorded at varying number of time points due irregular sampling, missing data, disease characteristics and patient stays

Questions and References

- [1] "Deep Learning in Bioinformatics" (Min, 2017)
- [2] "Deep Learning for Health Informatics" (Ravi, 2017)
- [3] “Introduction to Machine Learning” (Alpaydin, 2014)
- [4] "A Review of Machine Learning Applied to Medical Time Series" (Westhuizen, 2016)
- [5] "A Critical Review of Recurrent Neural Networks for Sequence Learning" (Lipton, 2015)
- [6] “Deep Learning with Python” (Chollet, 2017)
- [7] “Learning to Diagnose with LSTM Recurrent Neural Networks” (Lipton, 2015)
- [8] "Scalable and accurate deep learning for electronic health records“ (Rajkomar, 2018)
- [9] “Dynamic Mortality Risk Predictions in Pediatric Critical Care Using Recurrent Neural Networks” (Aczon, 2017)

- Any questions?