Machine Learning for Healthcare

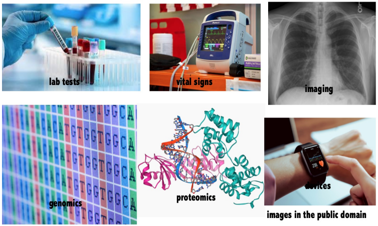
Fatima Batool

June 22, 2024

Plan for this lecture

1. What are unique challenges in healthcare data?
2. Foundation of deep learning neural networks
3. Learning from the network
4. A Case study: prediction COVID19 data
5. Overview and next steps

Medical data challenges

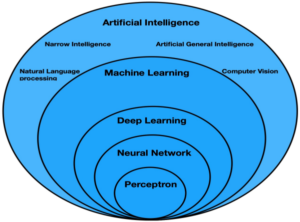


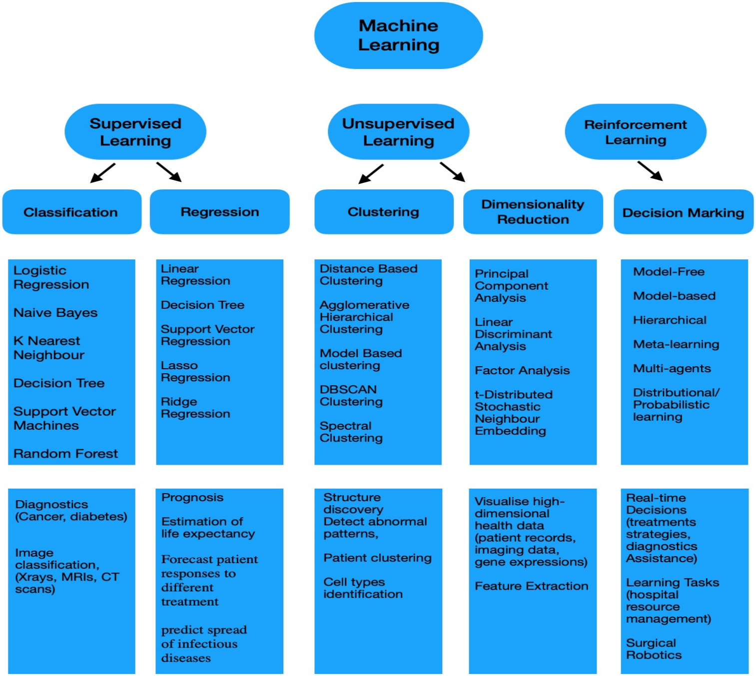
images in the public domain

- Data types from various sources include
  - EHR data
  - both structured and unstructured data
  - imaging and other pixel based diagnostics
  - genomics
  - peripheral sources of data such as wearable devices
- Consistent issues with medical datasets
  - Disorganised data, missing data, heterogeneous data, class imbalance

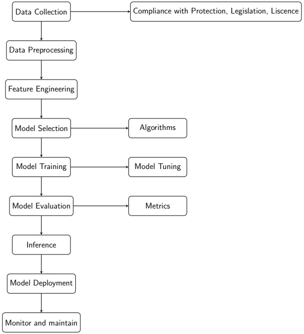
Why machine learning for healthcare?

- Traditional health care task are categorised as:
  - Diagnosis: the identification of disease from it's symptoms
  - Prognosis: the prospect of recovering from disease
  - Treatment: managing and caring for a patient to combat a disease or condition
  - Prevention: measures taken to prevent diseases or injuries rather than curing them or treating their symptoms




Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Dimensionality Reduction
Decision Marking
Logistic Regression
Naive Bayes
K Nearest Neighbour
Decision Tree
Support Vector Machines
Random Forest
Linear Regression
Decision Tree
Support Vector Regression
Lasso Regression
Ridge Regression
Distance Based Clustering
Agglomerative Hierarchical Clustering
Model Based clustering
DBSCAN Clustering
Spectral Clustering
Principal Component Analysis
Linear Discriminant Analysis
Factor Analysis
t-Distributed Stochastic Neighbour Embedding
Model-Free
Model-based
Hierarchical
Meta-learning
Multi-agents
Distributional/ Probabilistic learning
Diagnostics (Cancer, diabetes)
Image classification, (Xrays, MRIs, CT scans)
Prognosis
Estimation of life expectancy
Forecast patient responses to different treatment
predict spread of infectious diseases
Structure discovery
Detect abnormal patterns,
Patient clustering
Cell types identification
Visualise high-dimensional health data (patient records, imaging data, gene expressions)
Feature Extraction
Real-time Decisions (treatments strategies, diagnostics Assistance)
Learning Tasks (hospital resource management)
Surgical Robotics

Machine learning pipeline



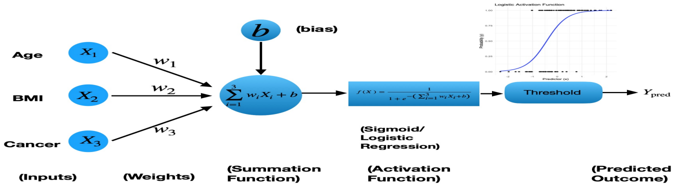
Know your data

Suppose we have:

- Data for $i = 1, 2, \cdots, N$ observations/patients
- Inputs: d predictor variables $X = x_1, x_2, \cdots, x_d$
- Outputs: dichotomous response variable y

Patient Id	Age	Gender	Smoke	$\cdots$
1	49.28	0	Current	$\cdots$
1001	56.82	0	Moderate	$\cdots$
10000	44.59	1	Low	$\cdots$
1007	56.72	1	Never	$\cdots$
100	41.32	1	Never	$\cdots$

Supervised learning problem



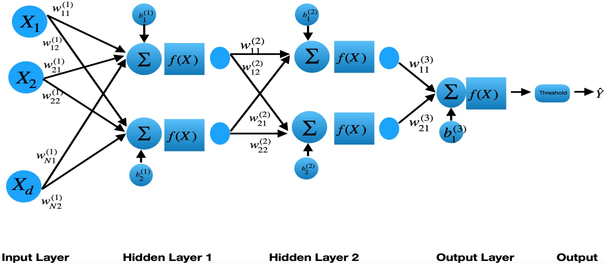
Parameters: $\boldsymbol{w}$ are weights and $\boldsymbol{b}$ are biases.
A neural net is a non-linear function of $\boldsymbol{w}$, $\boldsymbol{b}$ and $\boldsymbol{x}$.

$$a = \sum_{i=1}^{d} w_i x_i + b,$$

$$y(x, w, b) = f(a),$$

$$f(a) = \frac{1}{1 + e^{(-a)}}.$$

MLP architecture



Multilayer neural networks passes the output from one function to other forming a composition of functions, formally:

$$y_k(x, w, b) = h\left(\sum_{j=1}^{M} w_{kj}^{(2)} \; f\left(\sum_{i=1}^{N} w_{ji}^{(1)} x_i + b_j^{(1)} \right) + b_k^{(2)} \right).$$

Training to maximize the parameters from likelihood using stochastic gradient descent optimization.

Training steps

1. Split data into training, validation, and test set
2. Initialise parameters randomly
3. Forward pass (sample a batch of data from training set)
4. Compute loss (make prediction and compute “logloss”, “cross entropy”)
5. Validation during training
6. Backward pass (error backpropagation)
7. Update weights and biases
8. Repeat (optimisation)

Loss function

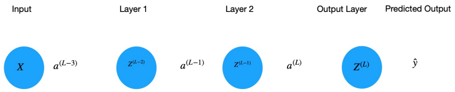
Cross entropy loss function:

$$C = L(\hat{y}, y) = -\sum_{i=1}^{N} y_i log(\hat{y}_i) - \sum_{i=1}^{N} (1 - y_i) log(1 - \hat{y}_i).$$

To compute $\hat{Y}$ we need to compute all weights and biases from all previous layers.
Output of any neuron is dependent on three aspects:

- Bias
- weight
- activation function from previous layer

Gradient descent for neural network



Recall we have called the weighted sum as:

$$a^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)}.$$

and the activation function called on this weighted sum as $f(a^{(L)})$, lets the output of the activation function be called

$$z^{(L)} = f(w^{(L)} a^{(L-1)} + b^{(L)}).$$

Note $a^{(L-1)}$ is calculated by it's own weight and biases in previous layer.

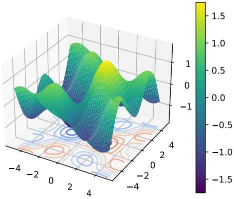
Formulas for computing derivatives

How sensitive is the loss function with respect to weights and biases?

Calculate derivative using chain rule with respect to one training examples:

$$\frac{\partial C_i}{\partial w^{(L)}} = \frac{\partial a^{(L)}}{\partial w^{(L)}} \frac{\partial z^{(L)}}{\partial a^{(L)}} \frac{\partial C_i}{\partial z^{(L)}}, \text{where}$$

$$\frac{\partial C}{\partial w^{(L)}} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial C_i}{\partial w^{(L)}}.$$



Formulas contd.

Similarly sensitivity of the loss/cost function to the biases are:

$$\frac{\partial C_i}{\partial b^{(L)}} = \frac{\partial a^{(L)}}{\partial b^{(L)}} \frac{\partial z^{(L)}}{\partial a^{(L)}} \frac{\partial C_i}{\partial z^{(L)}}, \text{where}$$

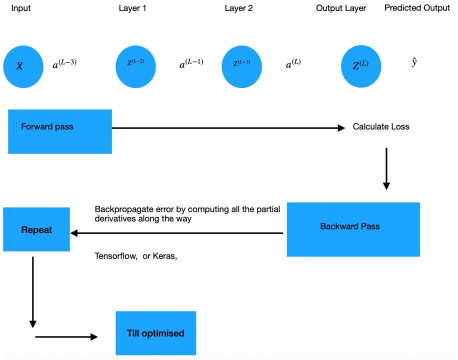
$$\frac{\partial C}{\partial b^{(L)}} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial C_i}{\partial b^{(L)}}.$$

Rules for parameter updates

Update each weight and bias using:

$$w^{(L+1)} = w^{(L)} - \epsilon \frac{\partial C}{\partial w^{(L)}}.$$

$$b^{(L+1)} = b^{(L)} - \epsilon \frac{\partial C}{\partial b^{(L)}}.$$



Exercise

Task: Predict the risk of 3-month COVID-19 death

How would you approach building a neural network model to predict patient outcomes?

Exercise

Task: Predict the risk of 3-month COVID-19 death

How would you approach building a neural network model to predict patient outcomes?

Consider the steps you would take from data preprocessing to model evaluation.

- There are 110 predictors,
- BMI with 0.01% of missing values
- Think about data scaling
- Think about feature selection or no selection

Exercise

Task: Predict the risk of 3-month COVID-19 death

How would you approach building a neural network model to predict patient outcomes?

Consider the steps you would take from data preprocessing to model evaluation.

- There are 110 predictors,
- BMI with 0.01% of missing values
- Think about data scaling
- Think about feature selection or no selection

See Jupyter notebook for the Python code of a two layered MLP on our COVID19 dataset. See what choices are made during the training.

Challenges and considerations

- Data quality and availability: Handling missing data and ensuring data integrity.
- Ethical and privacy concerns: Ensuring patient privacy and addressing algorithms' and user biases.
- Model interpretability: Importance of explainable AI in healthcare.
- Deployment and integration: Challenges in implementing ML models in real-world healthcare settings.
- Other: Workforce displacement, medical-legal responsibilities, and regulations.

Improvements on neural networks

For high performing Neural Network there are many practical aspect. Think about following:

- Setting up data for training model
- Hyper parameters tuning process
- How many layers
- How many units
- Learning rate
- Activation functions
- Mini-Batch size
- Regularization, drop out rate
- Better optimization algorithms (SGD with momentum Adam optimizer)

Recap today's lecture

1. Types of Machine learning tasks with relevance to healthcare
2. Key concepts in neural networks: inputs, size, layers, hidden nodes, weights, bias, activation functions, outputs
3. Feed forward neural network training using stochastic gradient optimisation
4. Challenges of health care datasets using real-world data, Programming code for you to run in your own time
5. Practical aspects and considerations for neural networks improvements