## Week 1 – Introduction

1. What is the US healthcare cost in 2019?
   1. $3.6 trillion
2. What is the biggest portion of healthcare waste in the US?
   1. Unnecessary care
3. How many people died each year because of preventable errors in US healthcare?
   1. 200K to 400K
4. Find the example data science applications for lower healthcare cost
   1. Predictive models for healthcare utilisation
   2. Early detection for heart failure
5. Which type of healthcare data are considered large in terms of data volume?
   1. Genomic data
   2. Medical imaging data
6. Which type of healthcare data are considered fastest in velocity?
   1. Real-time monitoring data from intensive care units
7. What are diagnosis applications?
   1. Patient triaging
   2. Heart failure detection
   3. Medical imaging analysis
8. What are outcome prediction applications?
   1. Hospital readmission prediction
   2. Length of stay prediction
   3. Mortality prediction
9. What are drug discovery and development applications?
   1. Molecule generation
   2. Clinical trial recruitment
   3. Molecule property prediction
10. Which one is a public health application?
    1. Predicting COVID19 cases at different locations in the US

## Week 2 – Health Data

1. Which of the following is true about electronic health records (EHR)?
   1. Outpatient EHR data are viewed as point events.
   2. Inpatient EHR data are viewed as interval events.
   3. EHR data contain longitudinal patient records.

EHR data is often siloed and incomplete. Data from a single patient can be scattered across multiple hospitals’ EHR’s.

1. Which of the following is true about clinical notes?
   1. They can provide a detailed description of patient status.
   2. Most EHR systems provide clinical notes functionality.
   3. Clinical notes can contain sensitive protected health information.

The unstructured format makes it more difficult for algorithms to process clinical notes.

1. Which of the following are the limitations of claims data?
   1. Coding errors can commonly occur in the claims data.
   2. Since claims data are for billing purposes, they do not accurately reflect patient status.
2. Which of the following are true?
   1. EHR are richer than claims.
   2. EHR captures the medication prescription information but does not capture whether the prescription are filled.
   3. Continuous signals provide objective assessments of patients.
3. What are examples of imaging data?
   1. X-rays
   2. Computed tomography
   3. Magnetic resonance imaging
4. What are the limitations of imaging data?
   1. Ground truth labels are lacking in those images.
   2. The high-resolution imaging data are challenging to process.
5. What is true about medical literature data?
   1. They are difficult to parse because of the natural language format.
6. Which of the following is a medical ontology for medications?
   1. RxNorm
7. What are clinical trial data?
   1. Trial protocols
   2. Trial eligibility criteria
   3. Data in clinical trial manage systems
8. What’s true about drug data?
   1. Drugs are often represented in molecule structures.
   2. Drug data are standard.
   3. ChEMBL is a large bioactivity database.

3D molecule structures are important by still uncommon.

## Week 3 – Machine Learning Basics

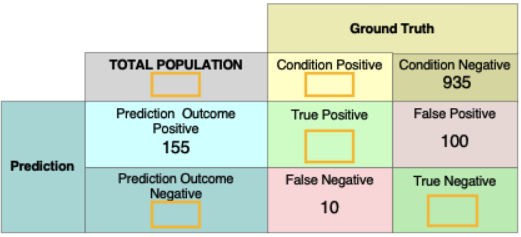
1. What are the steps in a clinical predictive modelling pipeline?
   1. Define prediction target
   2. Cohort construction
   3. Feature construction
   4. Prediction model
   5. Model Evaluation
2. How do you know if a prediction target is possible?
   1. Human performance as the target goal
   2. Prior experience from previous projects
   3. Results from related publications
3. What is true about retrospective studies?
   1. Retrospective studies need to handle a lot of noise in the data.
   2. The dataset used in a retrospective study is often already collected for different purposes than the study itself.

For example, the data used in prospective studies are often secondary data such as HER data, where the primary purpose of such data is not to support research studies but clinical practice.

1. Which of the following are true about cohort construction?
   1. A cohort study is about identifying a group of patients who are exposed to the risk of the prediction target.
   2. A case-control study will first identify the case patients then match them to the control patients.
2. Which of the following are true about feature construction in building a predictive model?
   1. The shorter the observation window, the harder the prediction task is.
   2. Longer prediction can limit the amount of data that can be used for building the predictive model.

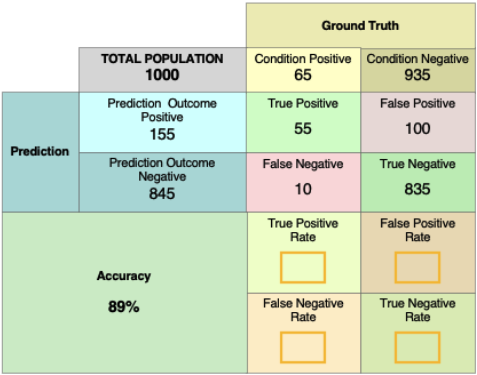
Some patients with shorter history will disqualify them in predictive modelling because of a long prediction window.

1. Which of the following are standard/good practice for building clinical predictive models?
   1. Cross-validation are most commonly used for evaluating deep learning models.
   2. Validation and Test sets can be small but should contain realistic samples with high-quality labels.
   3. Training data can be large and flexible, even with potentially noisy data.
2. What is the time complexity of K-means algorithm given n is # of points, k is # of clusters, d is the dimensionality of each point, and i is the number of clustering iterations?
   1. n\*d\*i\*k
3. Fill in the Blanks in this order: Total Population, Condition Positive, True Positive, Prediction Outcome Negative, True Negative.



* 1. TN = Actual Negative – FP = 935 – 100 = 835
  2. Predicted Negative = TN + FN = 835 + 10 = 845
  3. TP = Predicted Positive – FP = 155 – 100 = 55
  4. Actual Positive = TP + FN = 55 + 10 = 65
  5. Total Population = TP + FP + FN + TN = 55 + 100 + 10 + 835 = 1000

1. What is the definition of recall or sensitivity?
   1. True positive / (true positive + false negative)
2. Fill in the blanks in this order: True Positive Rate, False Positive Rate, False Negative Rate, True Negative Rate.



* 1. TPR = TP / Actual Positive = 55 / 65 = 84.6%
  2. FPR = FP / Actual Negative = 100 / 935 = 10.7%
  3. FNR = FN / Actual Positive = 10 / 65 = 15.4%
  4. TNR = TN / Actual Negative = 835 / 935 = 89.3%

## Week 4 – Deep Neural Networks

1. What is true about activation functions?
   1. Activation functions describe non-linear transformation
   2. Activation functions are specified by the user when setting up the neural network architectures
   3. ReLU is able to cope with vanishing gradient problems better than Sigmoid and Tanh.
2. What is true about gradient descent?
   1. Log-likelihood and likelihood function has the same optimal but log-likelihood is often easier to manipulate.
   2. Gradient descent is an optimization method for optimizing model parameters
   3. Stochastic gradient descent is a variant of the gradient descent method that is popular for neural networks training.

Gradient descent is a general method for optimisation, which is commonly used in many machine learning methods.

1. In forward computation, how is weight used?
   1. Connect neuron x2 from the input layer to output neuron h1 in the second layer.
2. In the general form of forward computation, the weight matrix and bias vector are used to connect?
   1. Pre-activation and activation
3. What is true about back propagation?
   1. Back propagation is an efficient way to compute derivatives on parameters on a neural network.
4. Which is true about Multilayer Neural Networks?
   1. An activation function is applied before the linear combination .
   2. The linear combination of layer 2 is computed as
   3. Multilayer neural networks are computed more efficiently on GPU.
5. Which is true about the readmission study using DNN?
   1. Multiple layers of DNN can help construct better features before the final classification layer.
   2. Separate DNNs are trained for the 5 different disease cohorts.
   3. DNN models achieved better accuracy than logistic regression models in this study.
6. Which of the forward computation equations are correct?
7. Why do you think a DNN is a good model for QSAR applications?
   1. The problem is very complex, with large numbers of observations and features.
8. What is true about hospital readmission?
   1. Readmission often indicates low-quality in the original admission.
   2. Different neural models can be trained for readmission for different diseases.
   3. Readmission can happen due to non-clinical reasons such as social determinants of health.