Practical Machine Learning

Special topic: Covid-19 & ML

Lecturer: Nguyen Thi Ngoc Diep, Ph.D.

Contact: ngocdiep \at vnu.edu.vn

How machine learning is helping us

- Identify who is most at risk,
- Diagnose patients,
- Develop drugs faster,
- Predict the spread of the disease,
- Understand viruses better,
- Map where viruses come from, and
- Predict the next pandemic.

https://towardsdatascience.com/fight-covid-19-with-machine-learning-1d1106192d84

In this lecture (workshop)

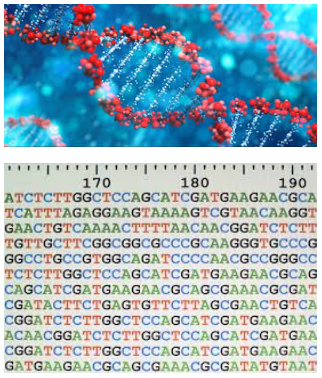
- Identify who is most at risk,
- Diagnose patients,
- Develop drugs faster,
- Predict the spread of the disease,
- Understand viruses better,
- Map where viruses come from, and
- Predict the next pandemic.

https://towardsdatascience.com/fight-covid-19-with-machine-learning-1d1106192d84

Techniques

- Regression models
  - Fitting a numerical sequence
  - Application: Predicting number of Covid-19 cases in New York
- Classification
  - Naive Bayes, Decision Tree, SVM, XGBoost
- High-dimentional visualization
  - PCA, t-SNE
- Clustering

Using genomic data



Libraries

- Numpy, Pandas
- Sklearn, XGB
- Biopython
- Matplotlib
- Seaborn

Sequence classification

- Approach: Context-independent (Bag of words)
- Preprocessing: To format the input sequence in a specific format
  - Tokenize
  - Numericalize
  - Vectorize

Sequence representations

- Context-independent: BoW, Word2Vec, Glove, TFIDF
- Left-to-right context: RNN (GRU, LSTM)
- Left-to-right & Right-to-left (bidirectional):
  - LSTM based: ELMO, ULMFiT...
  - Transformer-based: BERT, XLNet, GPT-2...

Book recommendation