Many of these basic principles are very familiar to the pharmacometrics and clinical pharmacology community.

Even before that, other methods like Bayesian statistics and Markov chains were used with a similar idea in mind.

Many of these methods are known to the pharmacomet- rics and clinical pharmacology community by different naming conventions.

The approaches typically used in pharmacometric applications fall into culture 1, where an underlying model is assumed based on phar- macological principles and understanding of drug properties.

We point the interested readers to other articles or books, such as “The Elements of Statistical Learning”9 (referred as ESL), and we refer to examples of their ap- plication in molecular biology, drug discovery, drug development, and clinical pharmacology.

Many of the issues related to data are universal and affect not only ML approaches but any quantitative discipline, including pharmacometrics.

In ML, including time as a distin- guished continuous variable into respective algorithms, remains challenging and is an area of active research.

Such data transforma- tion is an important preprocessing step that can have a profound effect on the model performance.

This could pose difficulties for many ML algorithms, including artificial neural networks and gradient boosting methods.

For this purpose, we can use unsupervised learning methods, like clustering, frequent pattern detection, and dimensionality reduction.

Further, it is able to also identify complex cluster shapes, like the one shown in Figure 2c.

Later, in every step of an iterative process, the two data points with smallest distance are grouped together.

Again, there is no optimal way of selecting such a threshold and many reasonable

solutions may exist.

The “curse of dimensionality” poses challenges on most data analysis approaches, including but not limited to ML.

Although some of these methods, like principal component analysis, have even been de- veloped long before the term ML has been coined,22 others, like t-Dis- tributed Stochastic Neighbor Embedding23 or Uniform Manifold Approximation and Projection,24 were developed recently and address complex challenges arising in data analysis.

In single-cell sequencing, Uniform Manifold Approximation and Projection or t-Distributed Stochastic Neighbor Embedding are used both for data visualization and for subsequent clustering.24 Dimensionality reduction is also used to visualize the high-dimensional chemical space33 or as a prepro- cessing step to improve performance of an ML model.34 Main takeaways • Clustering can be used to understand structure in data by grouping similar observations together.

• k-means clustering is a simple yet powerful tool, however, the number of clusters must be specified in advance.

The general principle of model selection is as follows: When there are enough data, we separate them into three subsets—training, validation, and test sets.

For regression models, we typically use the mean squared error, or other types of average objective functions, to compare model performance on training and test set.

For two-class classification problems, common performances measures are often derived from the “confusion matrix” shown in Figure 5 and briefly described below.

Some of these metrics could be generalized for multiclass problems, where there are more than two different labels in the dataset.

Let us assume we have training dataset with patients suffering either from a harmless cold or an influenza (flu) infection.

They have been used for at least 50 years.38,39 The idea behind decision trees is very intuitive and best represented in a visual form (e.g., Figure 1).

Depending on the problem, de- cision tree leaf nodes have classes, probabilities, or continuous values in case of regression.

Nevertheless, decision trees became the building block for two widely used approaches: Random deci- sion forests and gradient boosting frameworks.

The crucial difference between tree-based gradient boosting and random decision forests is on how trees are created.

Even without hyper- parameter tuning, they usually provide excellent performance with a relatively low computational cost.11 On the other hand, random forests are usually less prone to overfitting45 and re- quire less parameter tuning.46 This makes random decision for- ests attractive for smaller datasets or as a baseline method for benchmarking.

By introducing slack variables, tolerance for the residual term to be greater than ε is made.

The remit used to address this fundamental problem will be described more in two well-known extensions of RNN (long short-term memory (LSTM) and gated recurrent network (GRU)).

Examples of supervised ML applications in clinical pharmacology Models in clinical pharmacology have typically been established by translating physiological and pharmacological principles to systems of differential equations and using expectation-max- imization algorithms to estimate the model parameters.

This mechanistically motivated approach has proven useful in many applications and is a well-established component of drug develop- ment programs.

In a recent study, an ML-type control algorithm was inte- grated with existing structural PK/PD models that are familiar to pharmacometricians and the resulting closed-loop control system was found to outperform a sensor-assisted pump.69 Main takeaways •

Supervised learning methods infer models based on labeled out- put-input pairs of the training dataset.

For example, Bayesian methods are a well-established component of pharmacometric approaches.72,73 It seems, therefore, likely that as statistical and ML approaches be- come more established and more prominent in the pharmaceuti- cal industry, pharmacometricians will be among those who take advantage of these methods.

It is unlikely that these models will be completely replaced by ML approaches in the near future.

It is also important to note that we are still very much at the infancy stage of understanding at which point the merger of larger data with these novel ML methods can be beneficial for performance as compared with more traditional methods.

Overall, we expect that there will never be a universal, one-size- fits-all approach to which modelers from different fields converge.

Representation of Events in Nerve Nets and Finite Automata (RAND Project Air Force, Santa Monica, CA, 1951) <https://apps.dtic.mil/docs/citat

ions/ADA59

6138>.

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13–17, 2016.

Gemma, A. et al. Anticancer drug clustering in lung cancer based on gene expression profiles and sensitivity database.

Hutchinson, L. et al. Models and machines: how deep learning will take clinical pharmacology to the next level.