Non-Linear Algorithms in Supervised Classical Machine Learning

Monalisa Patel, MD, John B. C. Tan, PhD, Fu-Sheng Chou, MD, PhD

Our June data science column discussed the differences between bottom-up statistical modeling and top-down machine learning modeling. In machine learning, we provide the machine with input and output data (the training data set), we choose one or more algorithms, and we provide a range of hyperparameters to the machine. The machine will "learn" the data and figure how to associate the input data with the output data. After "learning", we then test the machine with the test data set to assess the model's performance. We also used linear regression as an example to demonstrate how regularization (LASSO and Ridge) is used in machine learning to improve "prediction." This month, we would like to take a deeper dive into discussing two major types of classical machine learning and introducing several "non-linear" algorithms to the readers.

"This month, we would like to take a deeper dive into discussing two major types of classical machine learning and introducing several "non-linear" algorithms to the readers."

Supervised vs. Unsupervised Learning

The difference between supervised vs. unsupervised learning is whether output data is provided to the machine. In other words, supervised learning requires both input and output data, but unsupervised learning requires only input data. As you may have already guessed, unsupervised learning is performed to allow the machine to "learn" the patterns hidden in the data, with the goal of grouping observations that are similar together. A commonly used algorithm in unsupervised training is called "K-means clustering, " where K was provided for the machine to cluster the data into K groups. This type of machine learning can be used to study novel disease states or novel cell types based on clinical or molecular data (1).

"In supervised learning, on the other hand, both input and output data are provided. Based on that, the machine then creates a prediction model, which will create predictions on a new set of data (the testing data set)."

In supervised learning, on the other hand, both input and output data are provided. Based on that, the machine then creates a prediction model, which will create predictions on a new set of data (the testing data set). This supervised learning methodology will lead to classification (categorical data type, e.g., BPD or not) or regression (continuous data type, e.g., cognitive scores on Bayley-III). Some algorithms are only used for classification, some are only for regression, and some are applicable to both classification and regression.

Non-linear supervised machine learning algorithms

K-nearest neighbors

K-nearest neighbors (KNN) is a model that classifies data points based on the points that are most similar (closest) to it. It trains the machine to make an "educated guess" on how an unclassified point should be classified. It uses proximity as a representation for "sameness". The algorithm takes a bunch of labeled points and uses them to learn how to label other points. To label a new point, it looks at the labeled points closest to that new point (those are its nearest neighbors). Once it checks with the "K" number of nearest neighbors, it assigns a label based on whichever label most of the neighbors have. Using the geometric distance to decide which is the nearest item may not always be reasonable or even possible; the type of the input may, for example, be text, where it is not clear how the items are drawn in a geometric representation and

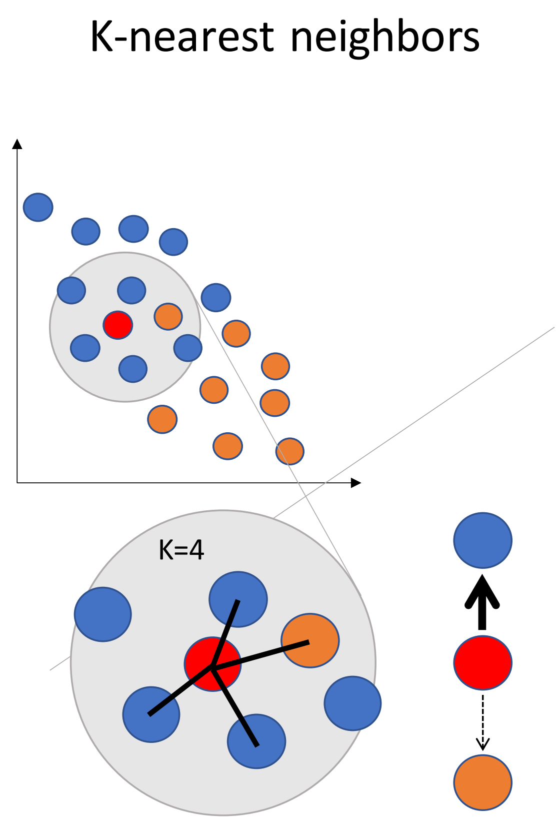


Figure 1. Visual depiction of the K-nearest neighbors algorithm.

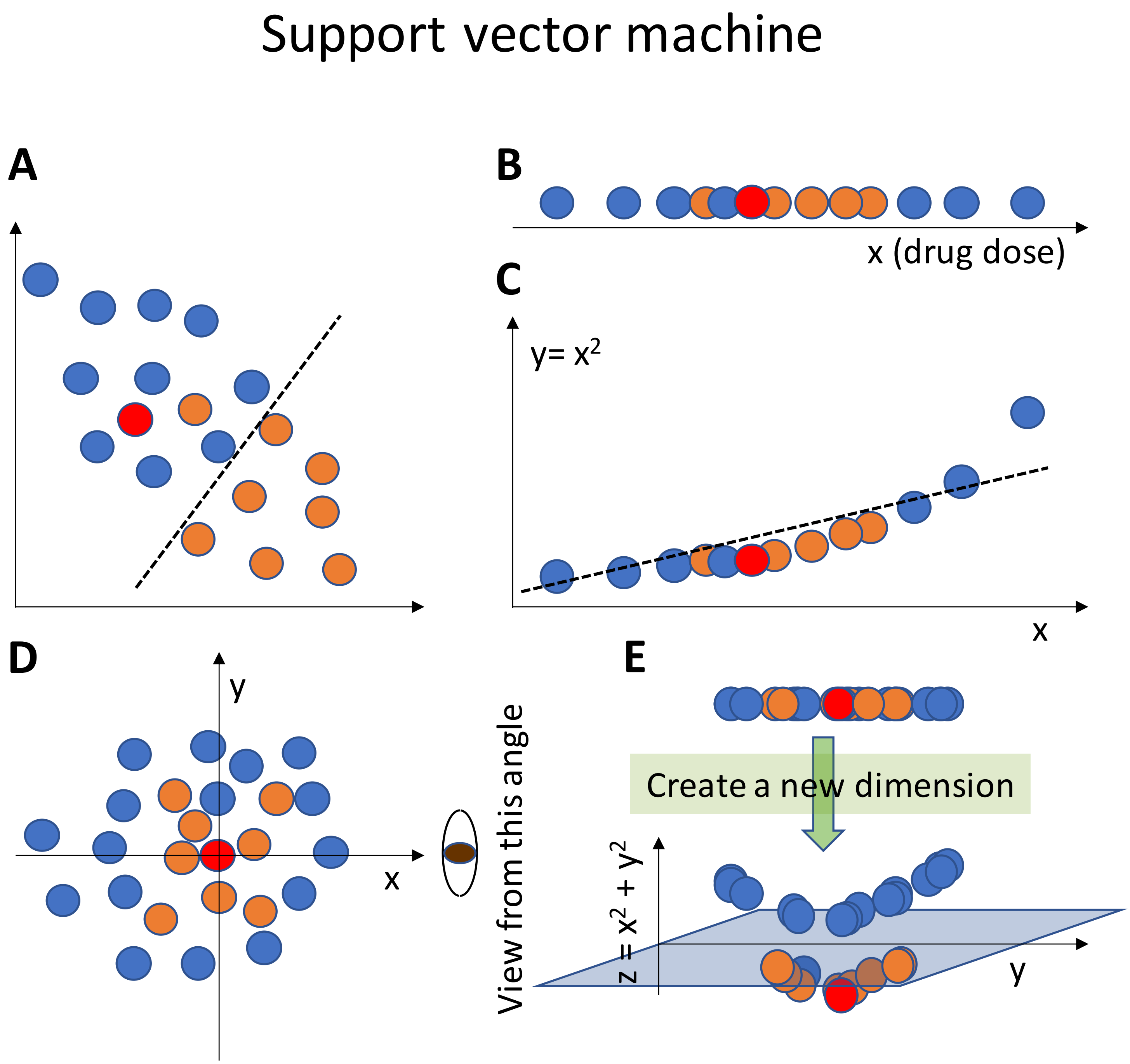


Figure 2. Visual depiction of support vector machine algorithms. (A) linear support vector machine. (B, C) Kernelized support vector machine with polynomial transformation. (B) Kernelized support vector machine with radial transformation in a 3-dimensional space. Note that in reality, a radial kernel is in a space with infinite dimensions, which is not illustratable.

how distances should be measured. One should therefore choose the distance metric on a case-by-case basis. KNN was used in the algorithm to predict diabetic retinopathy (2,3). Disadvantages include the susceptibility to overfitting (thus poor generalization), therefore not performing well in a high dimension dataset (as opposed to supporting vector machine, see below). Figure 1 showed a visual depiction of KNN.

Support vector machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mainly used in classification problems. In the SVM algorithm, we plot each data element as a point in n-dimensional space (where n is the number of features you have).

The value of each feature is the value of a particular coordinate. Then, we perform classification by finding a point/line/plane (the hyperplane) that differentiates the two classes very well. Support vectors are simply the coordinates of individual observations that are used to build a "soft margin" for classification, allowing some degree of misclassification in exchange for better generalizability (the bias-variance trade-off) (Figure 2A). Cross-validation is used to determine how much misclassification is allowed to have the best bias-variance balance. SVM can be linear or kernelized. When linear SVM is applied, it is assumed that the data are separable by a hyperplane that is one dimension lower than the data (Figure 2A).

"What if a hyperplane cannot easily separate the data? This is when kernel functions come into play. In the face of data that a hyperplane cannot be easily separated, kernel functions allow us to add dimensions to the data to allow for hyperplane separation, in other words."

What if a hyperplane cannot easily separate the data? This is when kernel functions come into play. In the face of data that a hyperplane cannot be easily separated, kernel functions allow us to add dimensions to the data to allow for hyperplane separation, in other words. Additional features are created (but not really from the computational standpoint, thanks to kernel tricks) to increase the dimensionality of the data, allowing us to separate data that were once inseparable. For example, in Figure 2B, where the x-axis denotes drug dose, the doses that are too high or too low result in one outcome (blue), and the doses in the middle result in the other outcome (orange). In this case, there is no single point that can successfully classify the outcome (into blue and orange). What a kernel function can do here is to bring in another dimension. For example, in Figure 2C, a polynomial kernel was introduced by having the y-axis be the drug dose to the power of two. In the two-dimensional space, the distance between two data points is calculated to identify the support vectors (the data points that establish the margin for each class). Subsequently, a line can be drawn to separate the two classes. In another example, the two-dimensional data are centered (Figure 2D). An imaginary eye is viewing the data from the end of the x-axis to render the view depicted in Figure 2E. Z-axis was added to the data by taking the sum of the square of the features x and y. Now the data can be separated by a horizontal plane that goes between class orange and class blue. This calculation is a simple example of another kernel function called the Gaussian radial basis function (RBF kernel or radial kernel). The radial kernel places the data points in a space with N dimensions where N is infinite, allowing a hyperplane of N-1 dimensions to separate the classes.

SVM can be used in image classification, spam detection, etc. It is very good for a small dataset with high dimensionality (many features or variables). The disadvantages of SVM are that it requires a high level of processing power. Additionally, the precise relationship between input and output features may be challenging to interpret, given high dimensionality. SVM use includes face

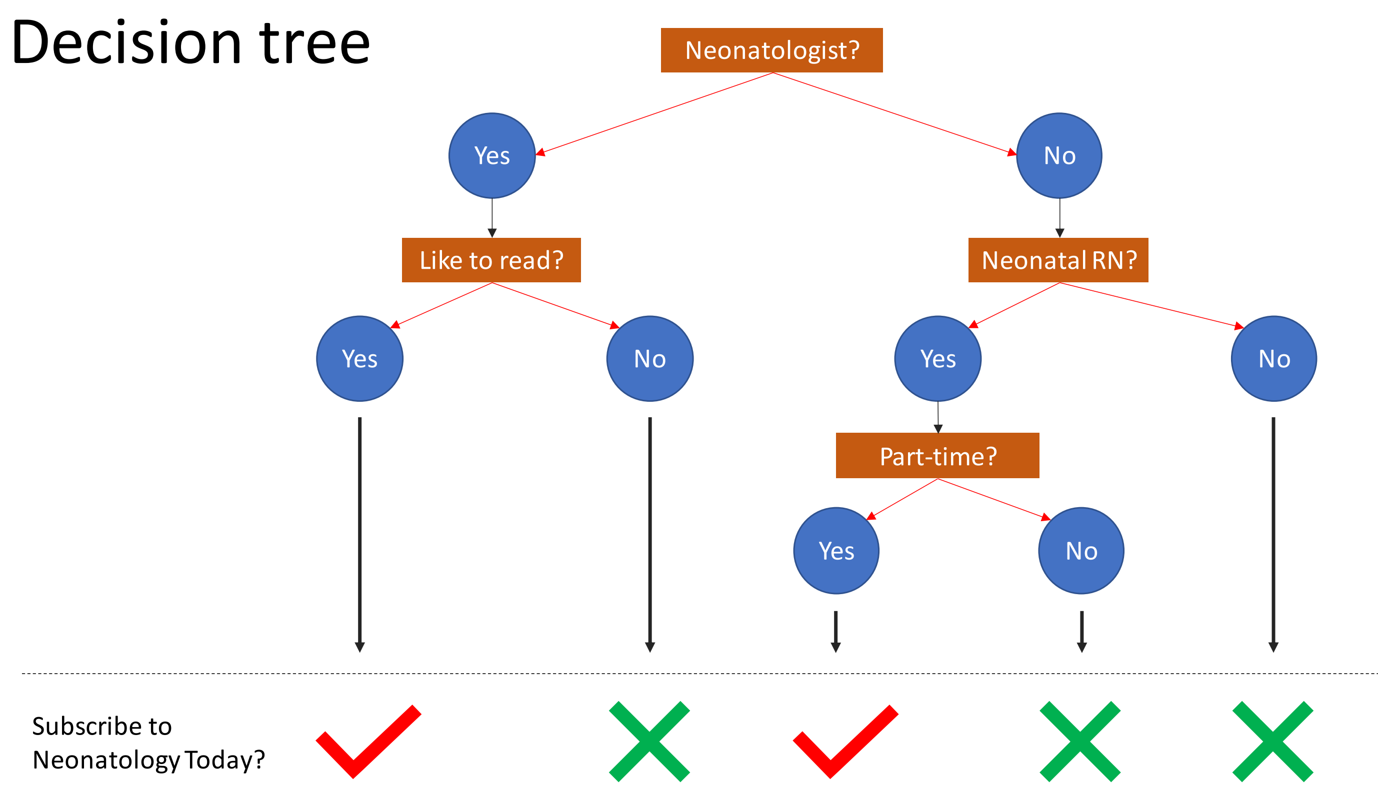


Figure 3. Visual depiction of a decision tree.

recognition, image classification, cancer classification, etc. (4).

Decision tree-based algorithm

Decision trees build classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. It works like a flow chart, separating data points into two similar categories at a time from the "tree trunk" to "branches" to "leaves," where the categories become more finitely similar. This situation creates categories within categories, allowing for organic classification with limited human supervision. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches, and a leaf node represents a classification or decision. An example of a decision tree is depicted in Figure 3. Decision trees can handle both categorical and continuous data types but are particularly useful if the input data are categorical. Decision trees can help physicians identify patients at a higher risk of developing a serious condition or needing an intervention such as mortality or requiring a gastrostomy tube placement (5–7). The disadvantage of decision trees is that they do not work well with smaller datasets, and decision trees are very prone to overfitting with the training data, and therefore not generalizable. They are also very sensitive to small changes in the input data. For the above reasons, the random forest algorithm was developed, in which case hundreds to thousands of trees were generated, each taking a subset of the features in the data set. Random forest takes care of the issues associated with hypersensitivity to small changes in the data and overfitting. However, it does not address another disadvantage with decision tree-based algorithms: the intolerance of missing values in the data set.

"Random forest takes care of the issues associated with hypersensitivity to small changes in the data and overfitting. However, it does not address another disadvantage with decision tree-based algorithms: the intolerance of missing values in the data set."

Summary

In this article, we introduced the definitions of supervised vs. unsupervised learning. We also introduced three non-linear classical machine learning algorithms, namely KNN, SVM, and the decision tree. These non-linear algorithms distinguish themselves from traditional clinical research by rethinking the linear assumptions of the data to quantify the association between the independent and dependent variables. We hope you enjoyed the read.

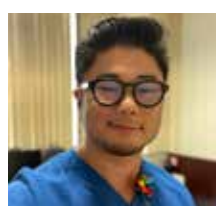
References

1. *Alashwal H, El Halaby M, Crouse JJ, Abdalla A, Moustafa AA. The Application of Unsupervised Clustering Methods to Alzheimer's Disease. Front Comput Neurosci. 2019 May 24;13:31.*
2. *Faruque MdF, Asaduzzaman, Sarker IH. Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). 2019. p. 1–4.*
3. *Alehegn M, Joshi RR, Mulay P. Diabetes Analysis And Prediction Using Random Forest, KNN, Naïve Bayes, And J48: An Ensemble Approach. 2019;8(09):9.*
4. *Real-Life Applications of SVM (Support Vector Machines) - DataFlair [Internet]. [cited 2021 Jul 11]. Available from: https://data-flair.training/blogs/applications-of-svm/*
5. *The Ad-Hoc PEG Tube Study Group, Beaulieu P, Bochanski PG, Clark CE, Doherty K, Frank JW, et al. When to Recommend a PEG Tube: A Decision Tree for Clinicians from a Catholic Perspective. Linacre Q. 2012 Feb;79(1):25–40.*
6. *Bae J-M. The clinical decision analysis using a decision tree. Epidemiol Health. 2014 Oct 30;36:e2014025.*
7. *Figure 1 Decision tree for the selection of the appropriate tube system... [Internet]. ResearchGate. [cited 2021 Jul 11]. Available from: https://www.researchgate.net/figure/Decision-tree-for-the-selection-of-the-appropriate-tube-system-for-enteral-nutrition-for_fig2_7504811*

Disclosure: The authors identify no conflict of interest

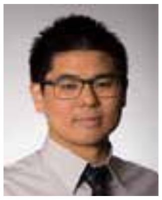
NT

Monalisa Patel, MD
Division of Neonatology,
Department of Pediatrics,
Children's Hospital Orange County
Orange, CA
Email: dr.monalisa.patel08@gmail.com



John B. C. Tan, PhD
Data Scientist
Assistant Professor of Pediatrics
Division of Neonatology, Department of Pediatrics
Loma Linda University Children's Hospital
Email: JBTan@llu.edu

Corresponding Author



Fu-Sheng Chou, MD, PhD -
Senior Associate Editor,
Director, Digital Enterprise
Neonatology Today
Assistant Professor of Pediatrics
Division of Neonatology, Department of Pediatrics
Loma Linda University Children's Hospital
Email: FChou@llu.edu

Neonatology Today's Digital Presence

Neonatology Today's now has a digital presence. The site is operational now and defines the future look of our digital web presence. By clicking on this https://www.neonatologytoday.org/web/., researchers can download individual manuscripts both in digital format and as part of the original PDF (print journal). While the PDF version of Neonatology Today will continue in its present form, we envision that the entire website will be migrated to this format in the next several months. We encourage you to take a look, "kick the wheels," and let us know where we still need to improve.. We are working towards making the website more functional for subscribers, reviewers, authors and anyone else. Although we have not yet applied for inclusion in the National Library of Medicine Database (Pub-Med), this new format meets several of the important metrics for this ultimate goal. As of December, 2020, NT has its own account with CrossRef and will assign DOI to all published material.

As we indicated last month, we look forward to a number of new features as well.

1. An online submission portal: Submitting a manuscript online will be easier than before. Rather than submitting by email, we will have a devoted online submission portal that will have the ability to handle any size manuscript and any number of graphics and other support files. We will have an online tracking system that will make it easier to track manuscripts in terms of where they are in the review process.
2. Reviewers will be able to review the manuscript online. This portal will shorten the time from receipt of review to getting feedback to the submitting authors.
3. An archive search will be available for journals older than 2012.
4. A new section called news and views will enable the submission of commentary on publications from other journals or news sources. We anticipate that this will be available as soon as the site completes the beta phase
5. Sponsors will be able to sign up directly on the website and submit content for both the digital and PDF issues of Neonatology Today.

Neonatology Today will continue to promote our Academic True Open Model (ATOM), never a charge to publish and never a charge to subscribe.

If there are any questions about the new website, please email Dr. Chou directly at:

fu-sheng.chou@neonatologytoday.net