

# Machine learning methods in psychiatry: a brief introduction

Zhirou Zhou,<sup>1</sup> Tsung-Chin Wu,<sup>2</sup> Bokai Wang ,<sup>1</sup> Hongyue Wang,<sup>1</sup> Xin M Tu,<sup>3,4</sup> Changyong Feng<sup>1</sup>

**To cite:** Zhou Z, Wu T-C, Wang B, *et al.* Machine learning methods in psychiatry: a brief introduction. *General Psychiatry* 2020;**33**:e100171. doi:10.1136/gpsych-2019-100171

Received 04 November 2019  
Accepted 05 November 2019

## ABSTRACT

Machine learning (ML) techniques have been widely used to address mental health questions. We discuss two main aspects of ML in psychiatry in this paper, that is, supervised learning and unsupervised learning. Examples are used to illustrate how ML has been implemented in recent mental health research.

## INTRODUCTION

The development of high technologies has significantly changed the research and treatment methods in psychiatry. Advanced technologies such as social media, smartphones and wearable devices have enabled psychiatric clinicians and researchers to collect a wide range data of subjects/patients within a relatively short period of time to monitor the psychical status of clients or patients,<sup>1</sup> and to offer more accurate and personalised treatments. While enjoying the convenience brought to us by the advanced technologies, we are facing the challenge of analysing the large data set generated from them, and making good prediction of some outcomes for a new subject. Unlike the traditional statistical methods which try to find a good fit of the data to interpret the association between the outcome and some potential features, medical researchers and clinicians are interested in the prediction of treatment methods (for example, the dosage of a drug) and treatment outcomes (eg, 5-year survival probability) given a comprehensive measurement of different features of a patient.

Machine learning (ML) takes advantage of advanced statistical methods and computer science techniques, and has been implemented to analyse ‘big data’ nowadays.<sup>2</sup> The common types of ML techniques used in the psychiatric field include supervised learning (SL) and unsupervised learning (USL).<sup>3</sup>

SL is used for data type with a labelled response variable. The purpose of SL is to develop a model for which the outcome can be formulated as a function of the features (covariates) so that the model can make a prediction of the outcome in the future when

only the features are given. For instance, suppose we are interested in identifying a patient with either major depressive disorder or no depression, based on the measurement of some factors of patients. SL methods try to build a model between the outcome (eg, depression or not) and a series of features, such as age, gender, education background, work type and so on, which are collected from different data sources. Commonly used examples of SL algorithms include logistic regression (LR) and support vector machine (SVM);<sup>4</sup> LR was borrowed directly from traditional statistics and SVM was invented by computer scientists.<sup>5</sup> We will discuss LR in detail in the next section.

USL is applied to data without a labelled outcome.<sup>6</sup> The algorithms try to recognise similarities/dissimilarities between subjects through input variables (features) without the aid of a labelled outcome. This is why it is called ‘unsupervised’. One of the most commonly used USL methods is the k-means clustering which minimises within-cluster variances to partition observations into k clusters. The lack of labelling will make USL more challenging, while this could also help to reveal the underlying data structure without a possible prior bias. We will discuss the k-means clustering by concrete example in the later section.

## SUPERVISED LEARNING

LR is a widely used statistical method to model the conditional expectation of a binary outcome with the given covariates. It has also been generalised to the case of polytomous outcomes.<sup>7</sup> For this reason, it is a natural choice for a classification method in multivariate analysis. After estimating the parameters in the regression function based on the training data, we predict the probabilities that a new subject will be assigned to each category in the outcome by substituting its observed features into the regression function.



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York, USA

<sup>2</sup>Department of Mathematics, University of California San Diego, La Jolla, California, USA

<sup>3</sup>Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA

<sup>4</sup>Naval Health Research Center, San Diego, California, USA

## Correspondence to

Professor Changyong Feng;  
Changyong\_Feng@URMC.  
Rochester.edu

LR is one of the most popular SL tools in biomedical studies. Recently, Lee *et al* developed an LR model for adolescent suicide attempt prediction using sociodemographic characteristics, risk behaviours and psychological variables.<sup>8</sup> This study is based on a sample of 247 222 subjects in the Korea Youth Risk Behavior Web-based Survey. The LR model was used to predict the risk of suicide with 13 different variables selected through univariate analysis screening.<sup>9</sup>

For simplicity, assume the behaviour of adolescent suicide attempt (a binary outcome variable  $Y$ , with 1 for 'Yes' and 0 for 'No') is strongly associated with age (a continuous variable  $X_1$ ), gender (a binary variable  $X_2$ , with 1 for male and 0 for female), experience of violence (a binary variable  $X_3$  with 1 for 'Yes' and 0 for 'No'), feelings of sadness (a binary variable  $X_4$ , with 1 for 'Yes' and 0 for 'No') and current alcohol drinking (a binary variable  $X_5$ , with 1 for 'Yes' and 0 for 'No') after variable selection. The LR model assumes that the conditional distribution of  $Y$  given the covariates is of the form

$$P(Y = 1|X_1, \dots, X_5) = \frac{1}{1 + \exp(-\mu)}$$

where

$$\mu = \beta_0 + \sum_{i=1}^5 \beta_i X_i$$

is called the linear predictor, a linear combination of covariates.<sup>10</sup>

Suppose we have a new subject with features in table 1, and want to predict the probability of committing suicide.

Then we have,

$$\begin{aligned} \mu &= \beta_0 + \sum_{i=1}^5 \beta_i X_i \\ &= (-0.285) \times 15 + 0.403 \times 0 + 1.257 \\ &\quad \times 1 + 1.760 \times 1 + 0.382 \times 1 - 2.497 \\ &= -3.373 \end{aligned}$$

and,

$$P(Y = 1|X_1, \dots, X_5) = \frac{1}{1 + \exp(-\mu)} = \frac{1}{1 + \exp(3.373)} \approx 3.32\%$$

Therefore, the suggested probability of this person's suicide was calculated as 3.32%, which belongs to the high-risk group (>0.12%).<sup>8</sup>

Variables	Coefficients	Covariate values of the subject
Age	-0.285	15
Gender: female	0.403	0
Experience of violence: yes	1.257	1
Feeling of sadness: yes	1.760	1
Current alcohol drinking: yes	0.382	1
Intercept	-2.497	

### UNSUPERVISED LEARNING

K-means clustering is a statistical technique that has been used to recognise the patterns of different types or levels of severity of a specific illness, based on related variables with no outcome labels provided. Fuente-Tomas *et al* proposed an easy-to-use, cluster-based severity classification for bipolar disorder (BD) that may help clinicians in the processes of personalised medicine and shared decision-making.<sup>11</sup> In this study, 224 subjects with a diagnosis of BD (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision) under ambulatory treatment were classified into five different clusters based on 12 variables from five domains.

In their method part, k-means clustering was used to reduce the dimension of four types of variables including patients' sociodemographic and BD characteristics, psychometric instruments and laboratory results. By the algorithm of k-means clustering, the criterion is calculated by within-point scatter,

$$W(C) = \frac{1}{2} \sum_{j=1}^k \sum_{C(i)=j} \sum_{C(i')=j} \|x_i - x_{i'}\|^2 = \sum_{j=1}^k n_j \sum_{C(i)=j} \|x_i - \bar{x}_j\|^2$$

where  $\bar{x}_j = (\bar{x}_{1j}, \dots, \bar{x}_{pj})$  is the mean vector of the  $j$ th cluster,  $n_j = \sum_{i=1}^n I(C(i) = j)$ , and  $C(i)$  refers to all clusters that subject  $i$  belongs to. The number of clustering  $k$  can be found out by the Elbow method,<sup>12</sup> a heuristic method that helps interpret the consistency within cluster analysis. K-means clustering aims to minimise the criterion by assigning  $n$  observations to  $k$  clusters in such a way that within each cluster, the variance between the observations and the cluster mean is minimised.

The variables were selected by testing the between-group difference using  $\chi^2$  test or one way analysis of variance. Along with other variables chosen by expert criteria, 12 variables were included in the global severity formula.<sup>11</sup> The variables included in the global severity formula are listed in box 1.

Since the 12 selected variables can take values from 0 to 1 and all of them have equal weights, the sum of all variables need to be multiplied by 10/12 so that the result could represent the severity from low (0) to high (10).

$$\begin{aligned} \text{Severity} &= \frac{10}{12} \times (\text{HospN} + \text{SuicAttN} + \\ &\quad \text{ComPD} + \text{BMI} + \text{MetS} + \text{IllnessN} + \text{SCIP}_{T+} + \\ &\quad \text{PD} \times \text{BD} + \text{FAST}_T + \text{FAST}_{\text{leisure}} + \text{SF}_{PF} + \text{SF}_{MH}). \end{aligned}$$

The severity clusters were defined by the 5th, 25th, 50th, 75th and 95th percentiles of the score calculated by this formula. Patients can be classified into different clusters by recognising which range (defined by the centiles) their global severity score falls into.

### CONCLUSION AND DISCUSSION

In this paper we give a brief introduction of two ML methods, SL and USL, through LR and k-means clustering. Examples have been used to show how they can

## Box 1 Variables in the global severity formula

### Variables

#### (1) Clinical characteristics of the bipolar disorder (BD)

Number of hospitalisations (*HospN*)

Number of suicide attempts (*SuicAttN*)

Comorbid personality disorder (*ComPD*)

#### (2) Physical health

Body mass index (*BM*)

Metabolic syndrome (*MetS*)

Number of comorbid physical illnesses (*IllnessN*)

#### (3) Cognition

Screen for Cognitive Impairment in Psychiatry score (*SCIP<sub>T4</sub>*)

#### (4) Real-world functioning

Permanently disabled due to BD (*PD × BD*)

Functioning Assessment Short Test Total Score (*FAST<sub>T</sub>*)

Functioning Assessment Short Test Leisure Time Subscale Score (*FAST<sub>leisure</sub>*)

#### (5) Health-related quality of life

SF-36 Physical Functioning Scale Score (*SF<sub>PF</sub>*)

SF-36 Mental Health Scale Score (*SF<sub>MH</sub>*).

be used in practice. As data structures are getting more and more complicated in mental health studies, we need advanced and flexible methods to analyse the data and to offer precise and personalised treatments for patients. We believe ML as a combination of statistical methods and computer science will play an important role in psychiatry. We plan to introduce some other developments of ML methods in the following issues.

**Correction notice** This article has been corrected since it was first published. The second equation under the section heading 'Unsupervised Learning' was missing an end parenthesis. This has since been updated.

**Contributors** ZZ and T-CW: collected the data and wrote the draft. BW and HW: reviewed and revised the draft. XMT: reviewed the article. CF: proposed the topic and reviewed the final draft.



Zhirou Zhou obtained her BEd in Economic Statistics from Beijing University of Technology in 2018. She is currently a masters student in Statistics in the Department of Biostatistics and Computational Biology at the University of Rochester Medical Center. Her research interests include variable selection and causal inference.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Commissioned; internally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID iD

Bokai Wang <http://orcid.org/0000-0002-2998-693X>

### REFERENCES

- Chen M, Mao S, Liu Y. Big data: a survey. *Mobile Netw Appl* 2014;19:171–209.
- Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015;349:255–60.
- Cho G, Yim J, Choi Y, et al. Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investig* 2019;16:262–9.
- Bzdok D, Krzywinski M, Altman N. Machine learning: supervised methods. *Nat Methods* 2018;15:5–6.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273–97.
- Miotto R, Li L, Kidd BA, et al. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 2016;6:26094.
- Agresti A. *An introduction to categorical data analysis*. 3rd edn. New York: Wiley, 2018.
- Lee J, Jang H, Kim J, et al. Development of a suicide index model in general adolescents using the South Korea 2012–2016 national representative survey data. *Sci Rep* 2019;9:1846.
- Wang H, Peng J, Wang B, et al. Inconsistency between univariate and multiple logistic regressions. *Shanghai Arch Psychiatry* 2017;29:124–8.
- McCullagh P, Nelder JA. *Generalized linear models*. 2nd edn. Chapman and Hall/CRC, 1989.
- Fuente-Tomas Ldela, Arranz B, Safont G, et al. Classification of patients with bipolar disorder using k-means clustering. *PLoS One* 2019;14:e0210314.
- Thorndike RL. Who belongs in the family? *Psychometrika* 1953;18:267–76.