Mathematical Institute

Using signatures and rough path theory to analyse a psychiatric data stream

Imanol Perez
(Joint work with T. Lyons and K. Saunders)

Mathematical Institute
University of Oxford

SIAM-IMA Conference on May 3

Oxford Mathematics

UNIVERSITY OF OXFORD
Mathematical Institute

Definition (Signature of a continuous path)

Let $J \subset \mathbb{R}$ be a compact interval, and let $X \in \mathcal{V}^p(J, \mathbb{R}^d)$ so that the following integration makes sense. The signature of X is defined as

$$S(X) = (1, X^1, X^2, \ldots) \in \bigoplus_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n}$$

where

$$X^n = \int_{\substack{u_1 < u_2 < \ldots < u_n \\ u_i \in J, i = 1, \ldots, n}} \ldots \int dX_{u_1} \otimes \ldots \otimes dX_{u_n} \quad \forall n \geq 1.$$

Definition (Truncated signature of a continuous path)

Similarly, we define, for $n \geq 0$,

$$S^n(X) := (1, X^1, X^2, \ldots, X^n).$$

- We have two data sets: a known set of known input-output pairs (the *training set*), $\{X_i, Y_i\}_i$, which is used to train the model, and a set of inputs that is used for testing (the *out-of-sample set*).

- We have two data sets: a known set of known input-output pairs (the *training set*), $\{X_i, Y_i\}_i$, which is used to train the model, and a set of inputs that is used for testing (the *out-of-sample set*).
- Features play an important role in machine learning.

Theorem (B. Hambly, T. Lyons)

The signature of a path with bounded variation is unique up to tree-like equivalence.

UNIVERSITY OF OXFORD
Mathematical Institute

Theorem

Let $X \in \mathcal{V}^1([0,T],\mathbb{R}^d)$ be a path with bounded variation. Then, given a multi-index $I = (i_1,\dots,i_n)$ with components in $\{1,\dots,d\}$ we have

$$\left\| \int_{\substack{u_1<u_2<\ldots<u_n \\ u_i\in J, i=1,\ldots,n}} dX_{u_1}^{i_1} \dots dX_{u_n}^{i_n} \right\| \leq \frac{\|X\|_{1,[0,T]}^n}{n!}.$$

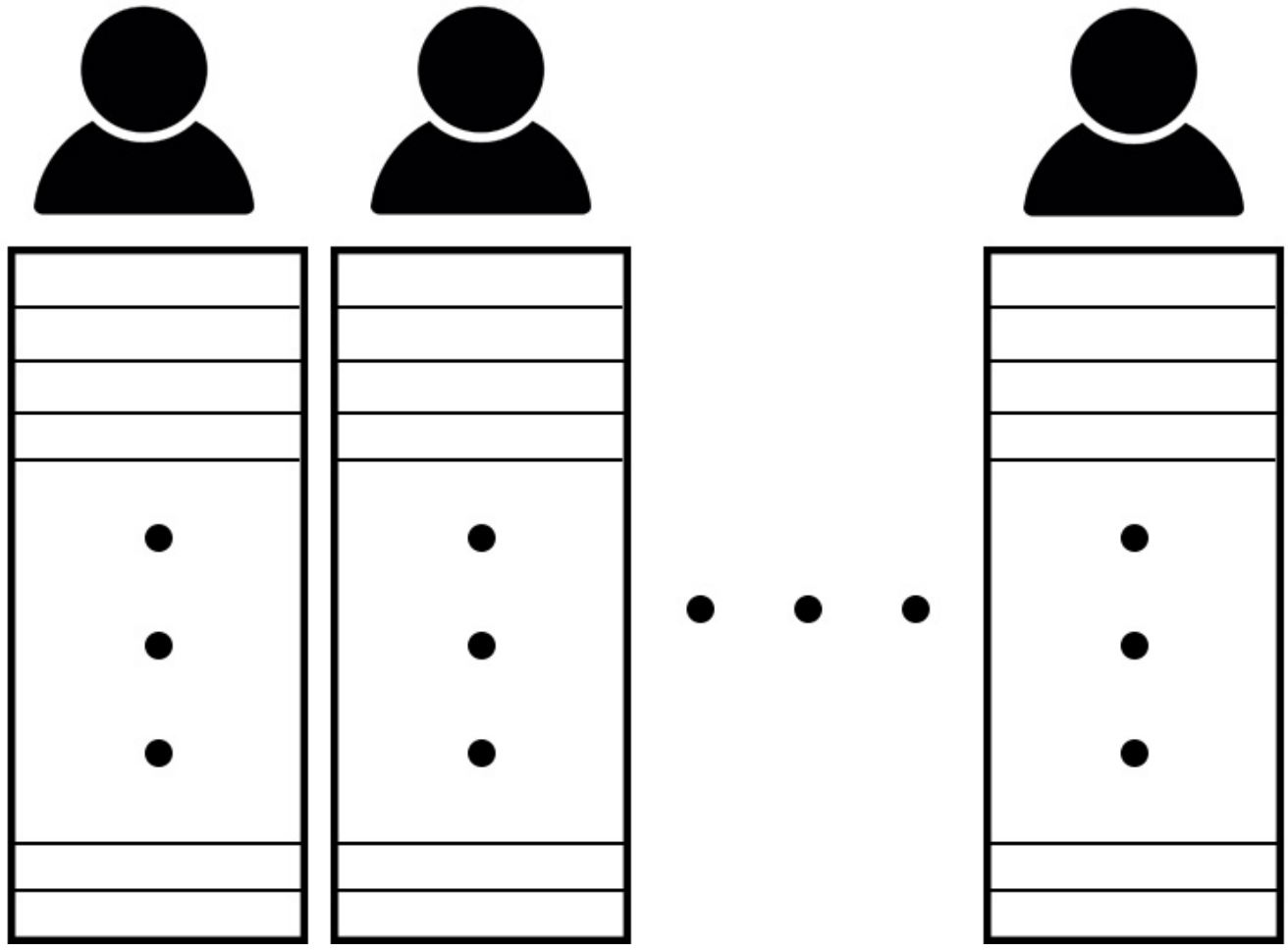
- Given a training set $\{R_i, Y_i\}_{i=0}^N$, of input-output pairs, where $R_i = \{t_{ij}, r_{ij}\}_j$ is a stream of data, construct a new set $\{X_i, Y_i\}_{i=0}^N$ with $X_i \in \mathcal{V}^1$.

- Given a training set $\{R_i, Y_i\}_{i=0}^N$, of input-output pairs, where $R_i = \{t_{ij}, r_{ij}\}_j$ is a stream of data, construct a new set $\{X_i, Y_i\}_{i=0}^N$ with $X_i \in \mathcal{V}^1$.
- Compute $\{S^n(X_i), Y_i\}_{i=0}^N$ for some $n \in \mathbb{N}$.

- Given a training set $\{R_i, Y_i\}_{i=0}^N$, of input-output pairs, where $R_i = \{t_{ij}, r_{ij}\}_j$ is a stream of data, construct a new set $\{X_i, Y_i\}_{i=0}^N$ with $X_i \in \mathcal{V}^1$.
- Compute $\{S^n(X_i), Y_i\}_{i=0}^N$ for some $n \in \mathbb{N}$.
- Apply regression against the truncated signature.

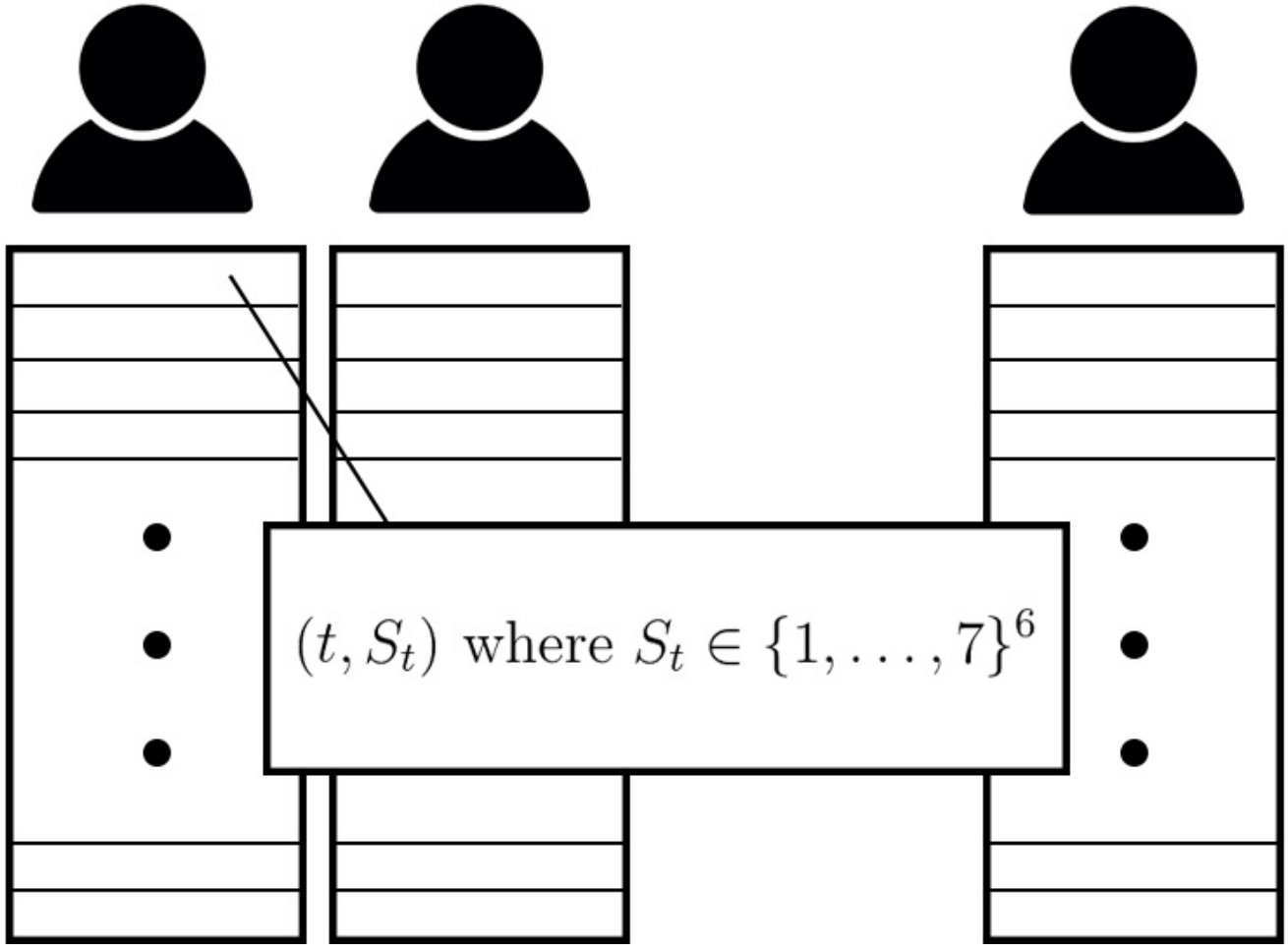
Application to psychiatric data

The problem



Application to psychiatric data

The problem

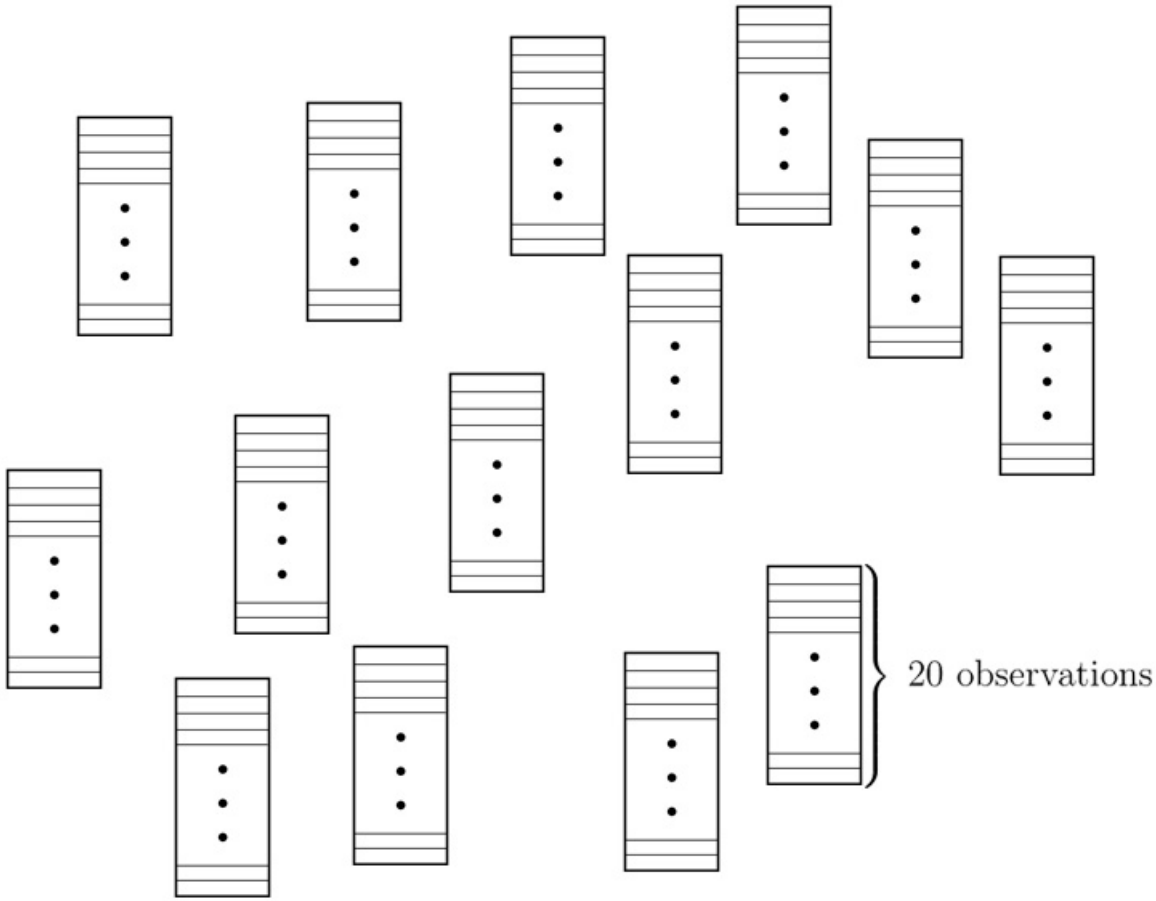


- Given information about a participant, can we tell if he or she is healthy, has bipolar disorder or has borderline personality disorder?

- Given information about a participant, can we tell if he or she is healthy, has bipolar disorder or has borderline personality disorder?
- Given a participant and information about the last few days, can we predict the mood the following day?

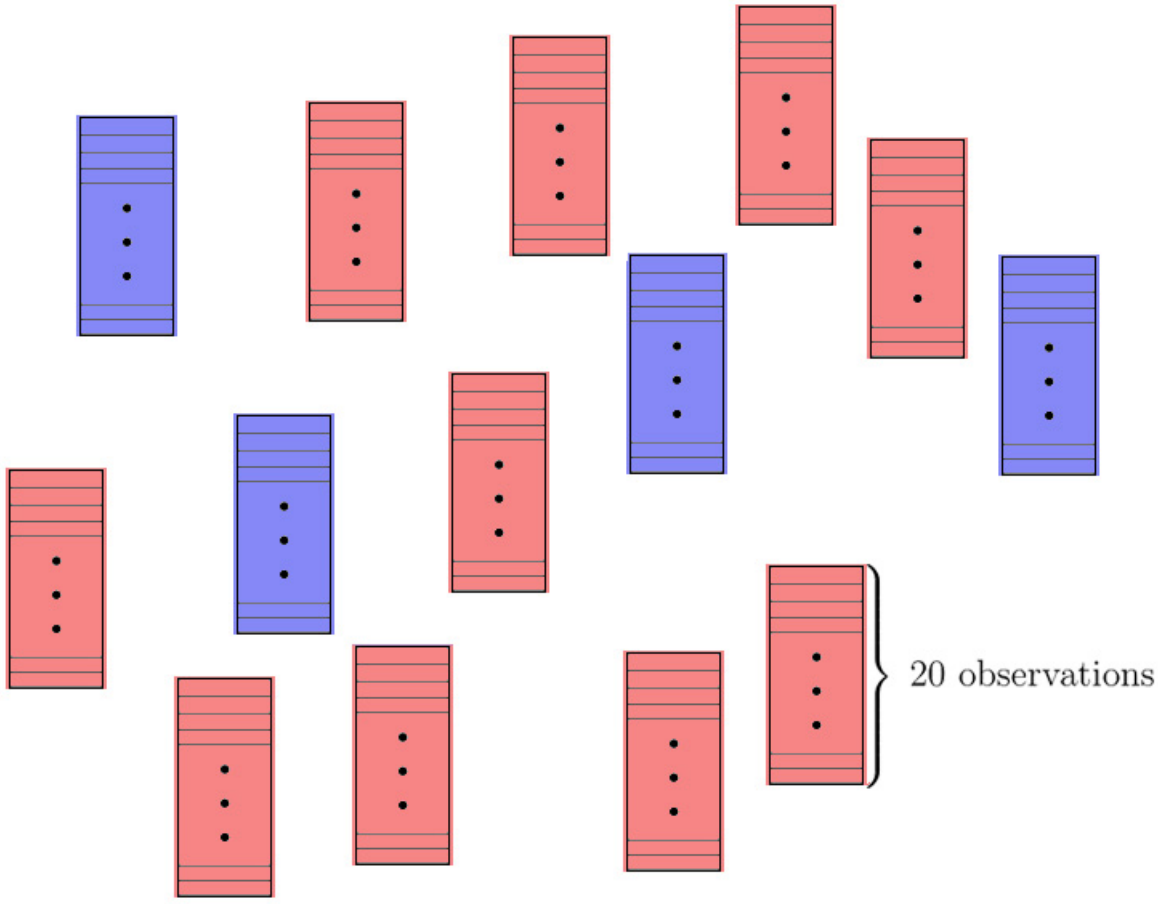
Application to psychiatric data

Methodology



Application to psychiatric data

Methodology



Application to psychiatric data

Methodology

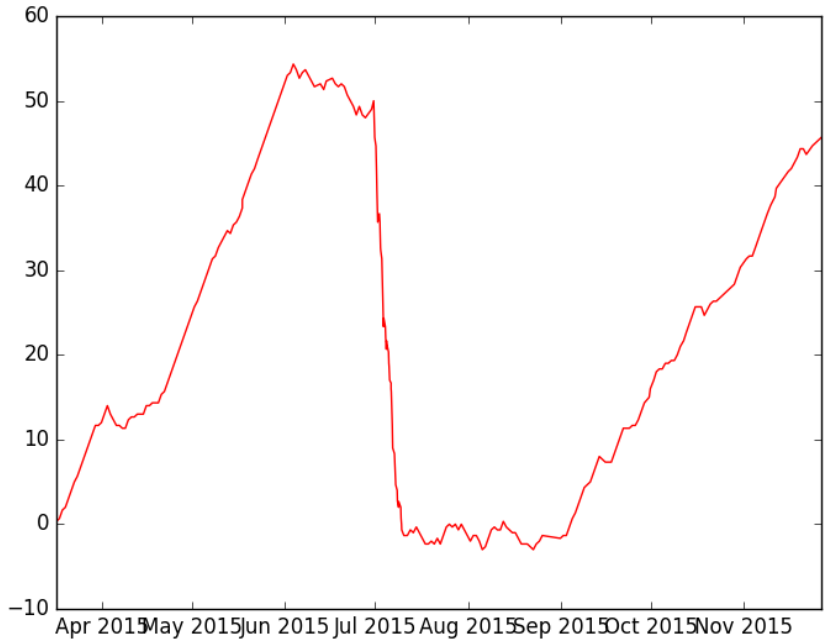


Figure: Normalised path for anxiety scores.

Application to psychiatric data

Predicting if a person is healthy, has bipolar disorder or has borderline disorder

Mathematical Institute

$$\{t_i, S_{t_i}\}_{i=0}^{N} \quad \rightarrow \quad \begin{cases} (-1, 1), & \text{if the partcipant is healthy} \\ (-1, -1), & \text{if the participant is bipolar.} \\ (1, 0), & \text{if the participant is borderline.} \end{cases}$$

Application to psychiatric data

Predicting if a person is healthy, has bipolar disorder or has borderline disorder

Mathematical Institute

Order	Correct guesses
2nd	75%
3rd	70%
4th	69%

Table: Percentage of people correctly classified in the three clinical groups.

$$\{t_i, S_{t_i}\}_{i=0}^{N} \quad \rightarrow \quad S \in \{1, \ldots, 7\}^6$$

where S is the scores of the participant the following recorded observation.

Application to psychiatric data

Predicting the future mood of a person with bipolar disorder

UNIVERSITY OF OXFORD
Mathematical Institute

Category	Healthy	Bipolar	Borderline
Anxious	98%	82%	73%
Elated	89%	86%	78%
Sad	93%	84%	70%
Angry	98%	90%	70%
Irritable	97%	84%	70%
Energetic	89%	82%	75%

Table: Percentage of correct guesses for mood predictions

Thank you!

Thank you!