

Recognizing characters by their rough-path signatures

Supervisor: Ben Graham

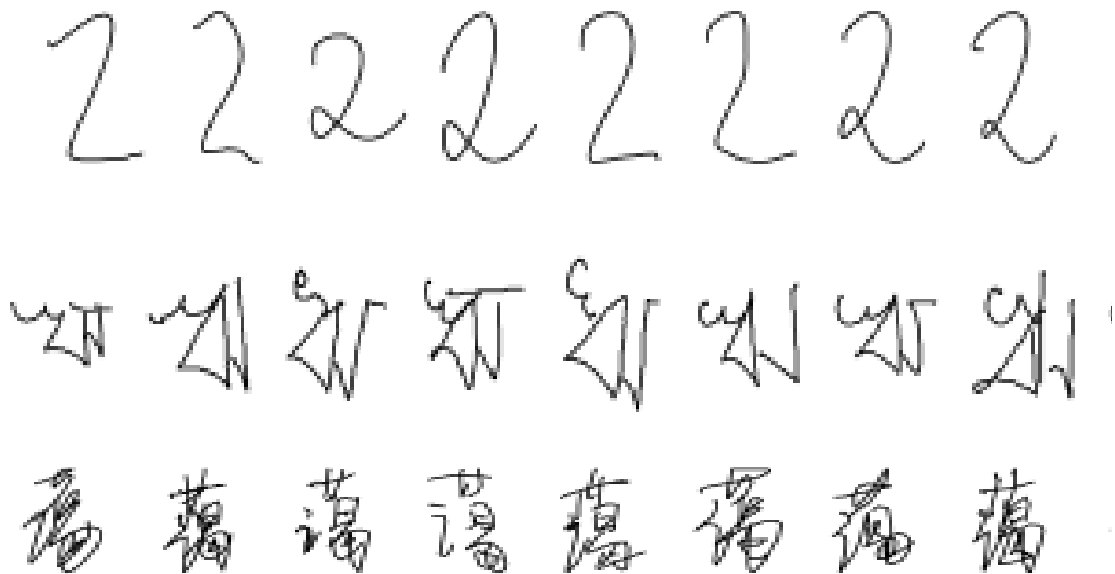
The “online character recognition problem” is the task of classifying a collection of paths representing a handwritten character. Rough paths theory provides a tool, the signature of iterated integrals,

$$X_{0,1}^n = \int_{0 < u_1 < \dots < u_n < 1} 1 dX(u_1) \otimes \dots \otimes dX(u_n) \in \mathbb{R}^{d^n}$$

for describing the shape of a path. If you think of the path as a driving signal for a differential equation

$$dY(t)f(Y(t)) = dX(t)$$

then the signature characterizes the *effect* of the path. The major challenge is to identify pairs of signatures, that may be far apart in a Euclidean sense, but that look similar to a human reading them.



Experience with programming in Python required.

- Terry Lyons and Zhongmin Qian, System Control and Rough Paths, Clarendon Press, Oxford Mathematical Monographs
- Pen-Based Recognition of Handwritten Digits Data Set
<http://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits>
- Online Handwritten Assamese Characters Dataset Data Set
<http://archive.ics.uci.edu/ml/datasets/Online+Handwritten+Assamese+Characters+Dataset>