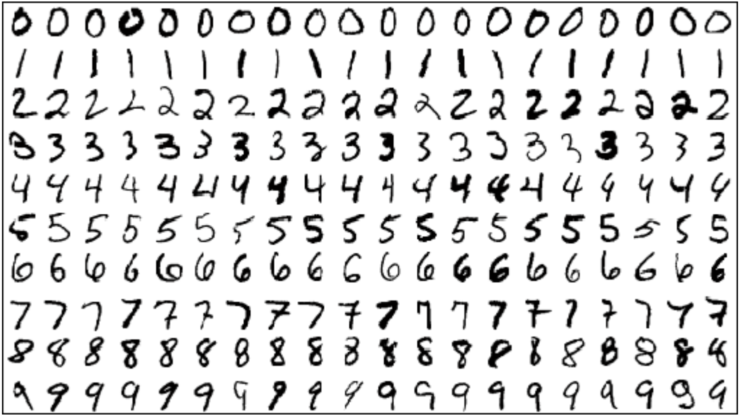
Neural networks and Chinese character recognition

Jeremy Reizenstein

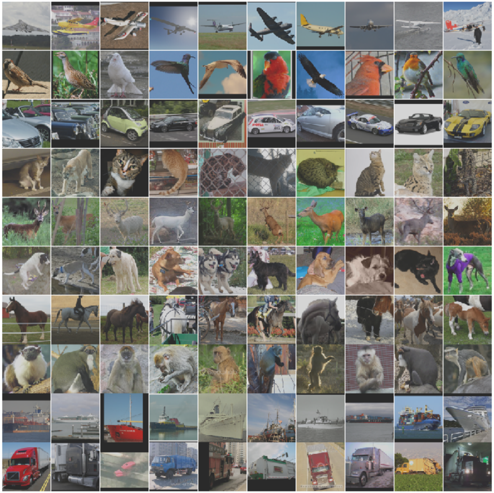
May 2016

Table of Contents

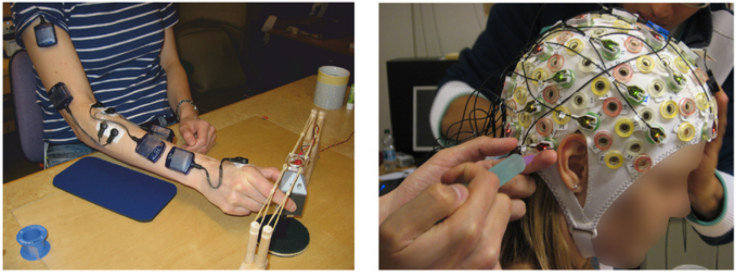
Example: Mnist

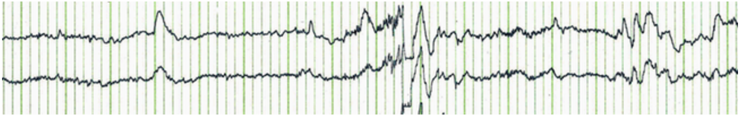


Example: Imagenet



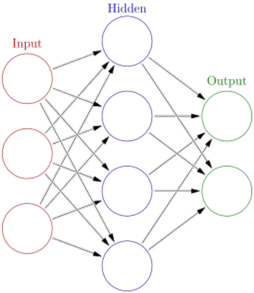
Example: EEG

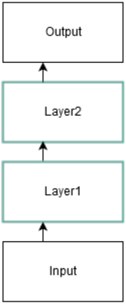

[1]



[1] Multi-channel EEG recordings during 3,936 grasp and lift trials, Luciw, Jarocka and Edin, Scientific Data 1, 2014; kaggle 2014

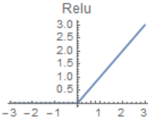
Neural Networks

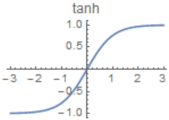




$$H = \sigma(\mathbf{W}_1 I + \mathbf{b}_1)$$

$$O = \sigma(\mathbf{W}_2 H + \mathbf{b}_2)$$





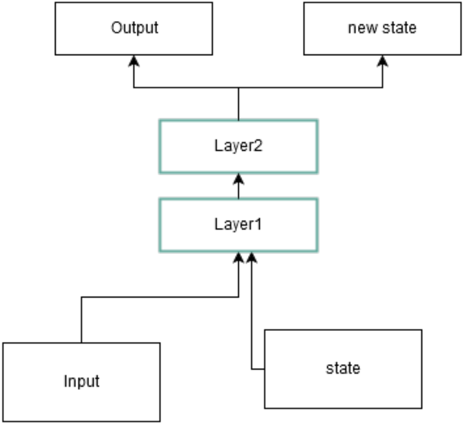
Typical approach to supervised machine learning

$$\text{input} \xrightarrow{\quad\text{representation}\quad} \mathbb{R}^k \xrightarrow{\quad\text{classification}\quad} \text{labels}$$

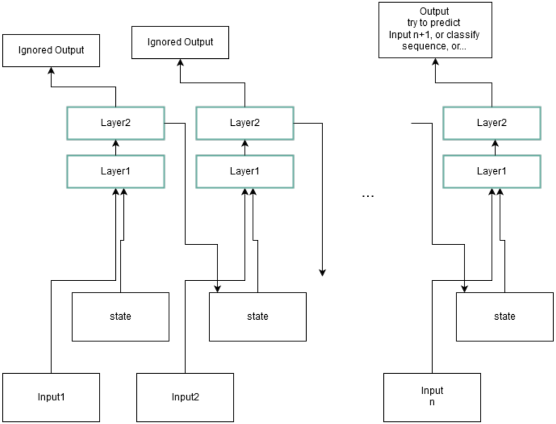
Timeseries and paths

Data which varies in time doesn't live in a fixed $\mathbb{R}^k$.
Time-structure of the data matters

Recurrent Neural Networks

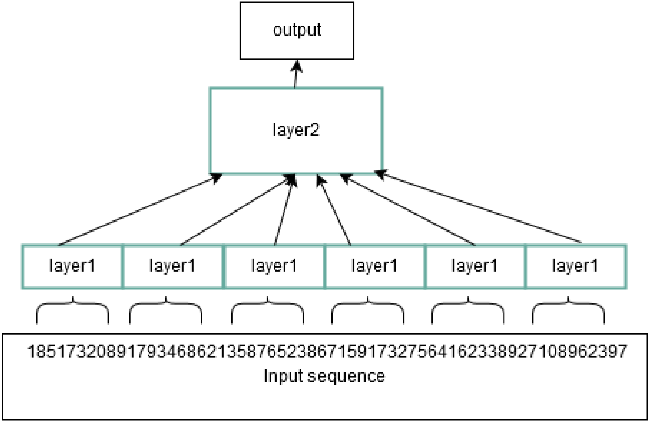


Recurrent Neural Networks

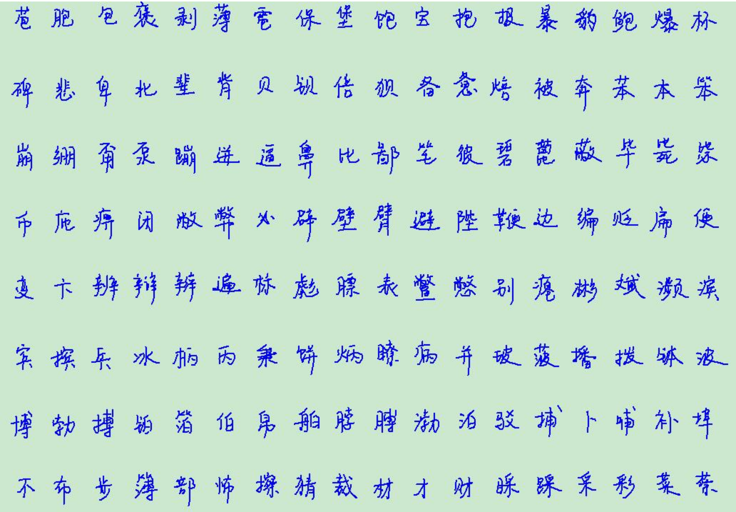


http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Convolutional Neural Networks

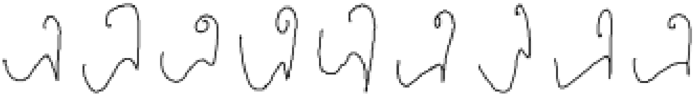


Chinese - the Casia online dataset



http://www.nlpr.ia.ac.cn/databases/handwriting/Home.html

Online handwriting



Typical approaches

$$\text{input} \xrightarrow{\text{representation}} \mathbb{R}^k \xrightarrow{\text{classification}} \text{labels}$$

$$\text{input} \times t \xrightarrow{\text{representation}} \mathbb{R}^{k\times t} \xrightarrow{\text{RNN}} \text{labels}$$

Signatures

The signature of a path is a set of iterated integrals.
Consider a path in $\mathbb{R}^3$ parameterised by the variable t ranging from 0 to 1, given by

$$t \mapsto (f_1(t), f_2(t), f_3(t))$$

Then, for example, the element 2,3 of the signature is

$$\int_0^1 \left[\int_0^t f_2'(s)\,ds\right] f_3'(t)\,dt = \int_0^1 \int_0^t df_2(s)\,df_3(t)$$

and element 2,1,2 of the signature is

$$\int_0^1 \int_0^t \int_0^s df_2(r)\,df_1(s)\,df_2(t).$$

Signatures

The mth level of the signature of a path in $\mathbb{R}^d$ is the d^m values of the elements with m integrated integrals. It takes values in $(\mathbb{R}^d)^{\otimes m}$. Given a piecewise linear path, it is easy to compute the first m levels of its signature using a theorem called Chen's identity.

Log-Signature demonstration

There is redundancy in the signature. The log signature is a transformation of the same information as the signature which is not redundant. For example, in $\mathbb{R}^2$, the first four levels of the signature look like this

$$(\cdot\,\cdot) + \begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix} + \left(\begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix}\;\begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix}\right) + \begin{pmatrix}\left(\begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix}\;\begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix}\right)\\\left(\begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix}\;\begin{pmatrix}(\cdot\,\cdot)\\(\cdot\,\cdot)\end{pmatrix}\right)\end{pmatrix}$$

- that is $2+4+8+16=30$ numbers while the log signature is only $2+1+2+3=8$ numbers.

Questions

A signature is a nice representation of a path of arbitrary length. When is it good enough? What properties of a complicated path can be derived from some levels of its signature? How to balance a representation using more levels of the signature versus chopping the path up and looking at signatures of the chunks?

Thanks!

EPSRC
Engineering and Physical Sciences
Research Council

centre for
complexity
science

WARWICK
THE UNIVERSITY OF WARWICK

Dr Ben Graham