

# Predicting (Economic) Trends: Why Signature Method in Machine Learning

Vladik Kreinovich<sup>1</sup> and Chon Van Le<sup>2</sup>

<sup>1</sup>University of Texas at El Paso  
El Paso, Texas 79968, USA  
vladik@utep.edu

<sup>2</sup>International University of Ho-Chi-Minh City  
Vietnam, lvchon@hcmiu.edu.vn

## 1. Predictions are important

- Prediction is one of the main objective of science.
- Economic predictions are one of the main objectives of econometrics.
- We want to predict what will happen to economy if we do not interfere.
- If we do not like this prediction, we need to decide what action to take to improve the economy.
- For that, we need to be able to predict what will happen if we undertake different actions.

## 2. In some cases, we can predict the trend but not the timing

- In some cases, we are able to predict exactly what will happen at different future moments of time.
- For example, what will be the GDP next year.
- In other cases, there are too many factors affecting the situation.
- For example, in agriculture, a lot depends on current weather patterns.
- In many such situations, it is possible to predict the trend – but not the timing.
- For example, we may predict that under an appropriate fiscal policy, the economy will improve.
- That at some future moment of time, the GDP will grow by 20% and the unemployment will decrease to half of the current value.
- But we cannot predict whether this will happen in 3 years or in 6 years.

### 3. In some cases, we can predict the trend but not the timing (cont-d)

- In such situations, what is important for prediction is also:
  - not when exactly different events happen in the past – the exact timing is too much affected by random events to be useful –
  - but rather what was the state of economy at different moments in the past.
- What is important, e.g., is that when the crops decreased by 30%, the unemployment grew by 20%.
- This show how the country's economy depends on its agriculture sector – but it does not matter whether this happened 3 or 4 years ago.

#### 4. Let us describe this situation in precise terms

- The state of the economy at each moment  $t$  can be characterized by the values of several characteristics  $x = (x_1, \dots, x_n)$  at  $t$ .
- In these terms, what we know is:
  - how, in the past, the state of the economy changed,
  - i.e., what were the values  $x_i(t)$  of all these characteristics for all the past moments of time  $t$ .
- These moments range from the moment  $T_0$  when we started recording these values to the current moment  $T$ .
- Based on this information, we want to predict how the state  $x(t)$  will change in the future, for moments  $t > T$ .
- This prediction should not depend on the exact duration of each state, only on the general trend.

## 5. Let us describe this situation in precise terms (cont-d)

- In other words, we should get the same prediction:
  - based on the actual values  $x(t)$  and
  - based on the values  $X(t) = x(\tau(t))$  for any increasing function  $\tau(t)$ .
- The processes may have been slower than they actually were – in this case, we may have  $\tau(t) = c \cdot t$  for some  $c < 1$ .
- The processes may have been faster than they actually were – in this case, we may have  $\tau(t) = c \cdot t$  for some  $c > 1$ .
- The processes may have been slower at some periods of time and faster at other – in this case, the dependence  $\tau(t)$  is nonlinear.
- In all these cases, whether we use the original records  $x(t)$  or re-scaled records  $X(t) = x(\tau(t))$ , we should get the exact same predicted trend.

## 6. Additional requirement: predictions often depend only on changes, not on the initial state

- Another reasonable assumption is that the trend's predictions should depend only on the relative changes, not on the actual initial state.
- For example, it is important to know that the GDP declined by 20% and the unemployment increased by 80%.
- However, it does not matter that much:
  - whether we talk about a big country with a large population and large GDP
  - or we talk about a smaller country with a smaller population and a smaller GDP.
- In precise terms, this means that what is important is:
  - not the actual values  $x_i(t)$ ,
  - but rather the ratios  $x_i(t)/x_i(T_0)$  describing how these values changed.

## 7. Additional requirement: predictions often depend only on changes, not on the initial state (cont-d)

- In other words, the predictions should remain the same whether we use the values  $x_i(t)$  or the values  $c_i \cdot x_i(t)$  for some constants  $c_i$ .
- In many economic situations, it is convenient to use logarithms  $v_i(t) = \ln(x_i(t))$  of the actual values.
- The logarithm  $\ln(c_i \cdot x_i(t))$  of each re-scaled value is equal to the sum  $\ln(x_i(t)) + \ln(c_i) = v_i + C_i$ , where  $C_i \stackrel{\text{def}}{=} \ln(c_i)$ .
- In these terms, we should predict the same trends whether we use the original dependence  $v_i(t)$  or the re-scaled dependence  $v_i(t) + C_i$ .

## 8. Signature method: a brief description

- In such situations, it turned out to be very efficient to replace the original description  $v_i(t)$  with the so-called *signature*.
- The signature is a sequence of the values

$$s_{i_1} = \int \dot{v}_{i_1}(t_1) dt_1,$$
$$s_{i_1, i_2} = \int_{T_0 \leq t_1 \leq t_2 \leq T} \dot{v}_{i_1}(t_1) \cdot \dot{v}_{i_2}(t_2) dt_1 dt_2,$$
$$\dots$$
$$s_{i_1, \dots, i_k} = \int_{T_0 \leq t_1 \leq \dots \leq t_k \leq T} \dot{v}_{i_1}(t_1) \cdot \dots \cdot \dot{v}_{i_k}(t_k) dt_1 \dots dt_k,$$
$$\dots$$

- Here, as usual,  $\dot{v}_i(t)$  indicates the derivative.
- This idea especially useful when we use machine learning.
- Trend predictions based on the signature are much more accurate than if we apply deep learning to the actual record  $v_i(t)$ .

## 9. Known fact: signature has the desired invariance properties

- If we add the same constant vector  $C_i$  to all the values  $v_i(t)$ , the derivatives will not change.
- Thus, the signature values will not change.
- Hence, the signature values remain the same whether we use the original dependence  $v_i(t)$  or the re-scaled dependence  $v_i(t) + C_i$ .
- Similarly, since  $\dot{v}_i(t) \cdot dt_i = dv_i(t_i)$ , each signature value can be represented in the following equivalent form

$$s_{i_1, \dots, i_k} = \int_{T_0 \leq t_1 \leq \dots \leq t_k \leq T} dv_{i_1}(t_1) \dots dv_{i_k}(t_k).$$

- From this expression, it is clear that this value does not change:
  - if we re-scale time,
  - i.e., if replace the original dependence  $v_i(t)$  with the re-scaled dependence  $v_i(\tau(t))$ .

## 10. But why signature?

- A natural question is: why signature and not other characteristics?
- In this talk, we provide a possible explanation of why it makes sense to use signature.

## 11. What we want: reminder.

- We want to find characteristics depending on all the values  $v_i(t)$  for all  $i$  and  $t$ .
- These characteristics should be invariant under re-scaling of time  $t \mapsto \tau(t)$  and under re-scaling of the values  $v_i(t) \mapsto v_i(t) + C_i$ .

## 12. How to achieve independence with respect to re-scaling of values $v_i(t)$

- Independence on re-scaling of values can be achieved if we only consider dependence on the derivatives  $\dot{v}_i(t)$ .
- Indeed, derivatives do not change under such re-scaling.
- Vice versa, once we know the derivatives, we can reconstruct the differences  $v_i(t) - v_i(T_0) = \int_{T_0}^t \dot{v}_i(s) ds$ .
- Thus, we can indeed reconstruct the value  $v_i(t)$  modulo such re-scaling; so:
  - to make sure that our characteristics do not change under re-scaling of values  $v_i(t)$ ,
  - it makes sense to consider characteristics depending on all the values  $\dot{v}_i(t)$  for all  $i$  and  $t$ .
- These characteristics should be invariant under re-scaling of time

$$t \mapsto \tau(t).$$

### 13. How can we describe general characteristics

- Most dependencies are smooth.
- There are seemingly non-smooth processes like phase transition, but in reality, they are smooth too.
- Just the time scale becomes different.
- Similarly in economics, most characteristics smoothly change with time.
- Sometimes the changes are fast and this seem abrupt and discontinuous, but in reality, they are reasonably smooth.

## 14. How can we describe general characteristics (cont-d)

- In general, a sufficiently smooth function  $b = f(a_1, \dots, a_n)$  of  $n$  inputs  $a_i$  can be described by its Taylor series:

$$b = b_0 + \sum_{i_1=1}^n b_{i_1} \cdot a_{i_1} + \sum_{i_1=1}^n \sum_{i_2=1}^n b_{i_1, i_2} \cdot a_{i_1} \cdot a_{i_2} + \dots +$$
$$\sum_{i_1=1}^n \dots \sum_{i_k=1}^n b_{i_1, \dots, i_k} \cdot a_{i_1} \cdot \dots \cdot a_{i_k} + \dots$$

- In our case, the unknowns  $a_i$  are the values  $\dot{v}_i(t)$  corresponding to different values of  $i$  and  $t$ .
- Theoretically, there are infinitely many moments of time  $t$ .
- However, of course, in practice, we only have values  $v_i(t)$  corresponding to finitely many moments of time  $t_1 < \dots < t_m$ .

## 15. How can we describe general characteristics (cont-d)

- In this case, as approximations to derivatives, we have finite differences

$$\dot{v}_i(t_\ell) \approx \delta v_i(t_\ell) \stackrel{\text{def}}{=} \frac{v_i(t_{\ell+1}) - v_i(t_\ell)}{\Delta t_\ell}.$$

- Here, we denoted  $\Delta t_\ell \stackrel{\text{def}}{=} t_{\ell+1} - t_\ell$ .
- By definition of the derivative, when  $\Delta t_\ell$  tends to 0, the finite difference  $\delta v_i(t_\ell)$  tends to the derivative; thus:
  - when the differences  $\Delta t_\ell$  are sufficiently small – i.e., when the moments  $t_\ell$  are sufficiently close to each other –
  - we can safely assume, for all practical purposes, that the differences  $\delta v_i(t_\ell)$  are equal to the corresponding derivatives.

## 16. How can we describe general characteristics (cont-d)

- In terms of these variables  $\delta v_i(t_\ell)$ , the general Taylor expansion of  $s(\delta v_1(t_1), \dots, \delta v_1(t_m), \delta v_2(t_1), \dots)$  takes the form

$$s = S_0 + \sum_{i_1=1}^n S_{i_1} + \sum_{i_1=1}^n \sum_{i_2=1}^n S_{i_1, i_2} + \dots + \sum_{i_1=1}^n \dots \sum_{i_k=1}^n S_{i_1, \dots, i_k} + \dots$$

- Here we denoted

$$S_{i_1, \dots, i_k} \stackrel{\text{def}}{=} \sum_{\ell_1=1}^m \dots \sum_{\ell_k=1}^m b_{i_1, \dots, i_k, \ell_1, \dots, \ell_k} \cdot \delta v_{i_1}(t_{\ell_1}) \cdot \dots \cdot \delta v_{i_k}(t_{\ell_k}).$$

- Dividing and multiplying each terms in the sum by the product  $\Delta t_{\ell_1} \cdot \dots \cdot \Delta t_{\ell_k}$ , we conclude that  $S_{i_1, \dots, i_k} =$

$$\sum_{\ell_1=1}^m \dots \sum_{\ell_k=1}^m B_{i_1, \dots, i_k}(t_{\ell_1}, \dots, t_{\ell_k}) \cdot \delta v_{i_1}(t_{\ell_1}) \cdot \dots \cdot \delta v_{i_k}(t_{\ell_k}) \cdot \Delta t_{\ell_1} \cdot \dots \cdot \Delta t_{\ell_k}.$$

- Here we denoted  $B_{i_1, \dots, i_k}(t_{\ell_1}, \dots, t_{\ell_k}) \stackrel{\text{def}}{=} \frac{b_{i_1, \dots, i_k, \ell_1, \dots, \ell_k}}{\Delta t_{\ell_1} \cdot \dots \cdot \Delta t_{\ell_k}}$ .

## 17. How can we describe general characteristics (cont-d)

- Taking into account that, for the practical purposes, the differences  $\delta v_{i_k}(t_{\ell_k})$  are equal to the derivatives  $\dot{v}_i(t_\ell)$ , we get  $S_{i_1, \dots, i_k} =$

$$\sum_{\ell_1=1}^m \dots \sum_{\ell_k=1}^m B_{i_1, \dots, i_k}(t_{\ell_1}, \dots, t_{\ell_k}) \cdot \dot{v}_{i_1}(t_{\ell_1}) \cdot \dots \cdot \dot{v}_{i_k}(t_{\ell_k}) \cdot \Delta t_{\ell_1} \cdot \dots \cdot \Delta t_{\ell_k}.$$

- This expression is the integral sum for the integral  $I_{i_1, \dots, i_k} =$

$$\int_{t_1=T_0}^T \dots \int_{t_k=T_0}^T B_{i_1, \dots, i_k}(t_1, \dots, t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \dots \cdot \dot{v}_{i_k}(t_k) dt_1 \dots dt_k.$$

- When the differences  $\Delta t_\ell$  are small, for practical purposes, the integral sum  $S_{i_1, \dots, i_k}$  is equal to the integral  $I_{i_1, \dots, i_k}$ .
- So, we conclude that a generic representation of a characteristic  $s$  has the form

$$s = S_0 + \sum_{i_1=1}^n I_{i_1} + \sum_{i_1=1}^n \sum_{i_2=1}^n I_{i_1, i_2} + \dots + \sum_{i_1=1}^n \dots \sum_{i_k=1}^n I_{i_1, \dots, i_k} + \dots$$

## 18. When is the general expression invariant under re-scaling of time?

- We want to come up with a description of all possible invariant characteristics.
- So, we need to find out which expressions are invariant under re-scaling of time  $t \mapsto \tau(t)$ .
- To find out which expressions are thus invariant, let us first take into account – as we did in the previous section – that  $\dot{v}_i(t) \cdot dt_i = dv_i(t)$ .
- Then, the expression for each integral takes the equivalent form

$$I_{i_1, \dots, i_k} = \int_{t_1=T_0}^T \dots \int_{t_k=T_0}^T B_{i_1, \dots, i_k}(t_1, \dots, t_k) dv_{i_1}(t_1) \dots dv_{i_k}(t_k).$$

- If we re-scale time, we get an expression

$$I_{i_1, \dots, i_k}^\tau = \int_{t_1=T_0}^T \dots \int_{t_k=T_0}^T B_{i_1, \dots, i_k}(\tau(t_1), \dots, \tau(t_k)) dv_{i_1}(t_1) \dots dv_{i_k}(t_k).$$

## 19. When is the general expression invariant under re-scaling of time (cont-d)

- Invariance means that we should have  $I_{i_1, \dots, i_k} = I_{i_1, \dots, i_k}^\tau$  for all possible functions  $\dot{v}_i(t)$ .
- This means that the coefficients  $B_{i_1, \dots, i_k}$  should be the same in both cases, i.e., that we should have

$$B_{i_1, \dots, i_k}(t_1, \dots, t_k) = B_{i_1, \dots, i_k}(\tau(t_1), \dots, \tau(t_k)) \text{ for all } \tau(t).$$

- One can easily check that every two tuples  $t_1 < \dots < t_k$  and  $t'_1 < \dots < t'_k$  can be obtained from each other by some increasing function.
- Thus, for all such tuples, the value  $B_{i_1, \dots, i_k}(t_1, \dots, t_k)$  is the same; we will denote it by  $\mathbf{B}_{i_1, \dots, i_k}$ .

## 20. When is the general expression invariant under re-scaling of time (cont-d)

- Similarly:
  - for any other ordering of the moments  $t_{\pi(1)} < \dots < t_{\pi(k)}$  corresponding for any permutation  $\pi : \{1, \dots, k\} \mapsto \{1, \dots, k\}$ ,
  - the value  $B_{i_1, \dots, i_k}(t_1, \dots, t_k)$  depends only on this permutation and is the same for all tuples  $(t_1, \dots, t_k)$  for which this ordering is true.
- We will denote this common value by  $\mathbf{B}_{i_{\pi(1)}, \dots, i_{\pi(k)}}$ .
- The whole domain of all possible tuples  $(t_1, \dots, t_k)$  – over which the integral is computed – can be divided into sub-domains.
- These sub-domains correspond to different orders between  $t_i$ .

## 21. When is the general expression invariant under re-scaling of time (cont-d)

- For example, for  $k = 2$ , we divide the domain of all the pairs  $(t_1, t_2)$  into two sub-domains:
  - the set of all the pairs for which  $t_1 < t_2$  that corresponds to the identity permutation  $\pi(i) = i$ , and
  - the set of all the pairs for which  $t_2 < t_1$  that corresponds to swap  $\pi(1) = 2$  and  $\pi(2) = 1$ .
- On the sub-domain  $D_{k,\pi}$  corresponding to the identity permutation  $\pi$ , the value  $B_{i_1,\dots,i_k}(t_1, \dots, t_k)$  is a constant  $\mathbf{B}_{i_1,\dots,i_k}$ .
- So the integral over this sub-domain has the form

$$\int_{D_{k,\pi}} B_{i_1,\dots,i_k}(t_1, \dots, t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \dots \cdot \dot{v}_{i_k}(t_k) dt_1 \dots dt_k =$$
$$\mathbf{B}_{i_1,\dots,i_k} \cdot \int_{T_0 < t_1 < \dots < t_k < T} dv_1(t_1) \dots dv_{i_k}(t_k).$$

## 22. When is the general expression invariant under re-scaling of time (cont-d)

- As usual, in the integration, the integral over measure-0 parts corresponding to possible equalities such as  $t_1 = T_0$ ,  $t_1 = t_2$ , is 0.
- So, we can say that

$$\int_{D_{k,\pi}} B_{i_1,\dots,i_k}(t_1,\dots,t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \dots \cdot \dot{v}_{i_k}(t_k) dt_1 \dots dt_k =$$
$$\mathbf{B}_{i_1,\dots,i_k} \cdot \int_{T_0 \leq t_1 \leq \dots \leq t_k \leq T} dv_1(t_1) \dots dv_{i_k}(t_k).$$

- The integral in the right-hand side is exactly one of the signature values  $s_{i_1,\dots,i_k}$ , so we get

$$\int_{D_{k,\pi}} B_{i_1,\dots,i_k}(t_1,\dots,t_k) \cdot \dot{v}_{i_1}(t_1) \cdot \dots \cdot \dot{v}_{i_k}(t_k) dt_1 \dots dt_k =$$
$$\mathbf{B}_{i_1,\dots,i_k} \cdot s_{i_1,\dots,i_k}.$$

## 23. When is the general expression invariant under re-scaling of time (cont-d)

- The integral  $I_{i_1, \dots, i_k}$  over the whole set of tuples  $(t_1, \dots, t_k)$  is equal to the sum of the integrals over all sub-domains.

- So, we get

$$I_{i_1, \dots, i_k} = \sum_{\pi} \mathbf{B}_{i_{\pi(1)}, \dots, i_{\pi(k)}} \cdot s_{i_{\pi(1)}, \dots, i_{\pi(k)}}.$$

- So, in the invariant case, the general expression for a characteristic takes the form

$$s = S_0 + \sum_{i_1=1}^n \mathbf{B}_{i_1} \cdot s_{i_1} + \sum_{i_1=1}^n \sum_{i_2=1}^n \mathbf{B}_{i_1, i_2} \cdot s_{i_1, i_2} + \dots +$$
$$\sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbf{B}_{i_1, \dots, i_k} \cdot s_{i_1, \dots, i_k} + \dots$$

- So, it is a linear combination of the signature values.

## 24. Conclusion

- It was known that signature values are invariant.
- What we have shown is that any other invariant characteristic is nothing else but a linear combination of signature values.
- In this sense, signature values is all that we can extract from the data, they provide full information about the inputs.
- This explains why signature values are so successful – since they provide full information about the input.

## 25. Acknowledgments

- This work was supported in part by the National Science Foundation grants:
  - 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and
  - HRD-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported by the AT&T Fellowship in Information Technology.
- It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.