We are interested in the prediction of future observations of a dynamical system. More precisely, we aim to study the robustness of neural networks trained on time series data. We assume that every object is smooth (i.e., C^∞).

1 Embedding dynamical systems

A dynamical system is defined by:

- A state space M modeled as a m-dimensional compact manifold.
- A complete vector field $X \in \Gamma(TM)$.

The integral curves of X are the set of curves verifying $\dot{\gamma}(t) = X_{\gamma(t)}$ for all $t \in \mathbb{R}$. The images of the integral curves of X can be seen as a 1-dimensional foliation $\mathcal{F}$ on M where each leaf $L \in \mathcal{F}$ can be associated with an integral curve γ_L such that $\gamma_L(\mathbb{R}) = L$ There is a natural diffeomorphism between the set of integral curves of X and the state space M, given by $\gamma \mapsto \gamma(0)$. From now on, we will identify the integral curves γ with points of M. Given any integral curve γ and any $\tau \in \mathbb{R}$, we define the curve $s_\tau\gamma : t \mapsto \gamma(t-\tau)$. For any $\tau \in \mathbb{R}$, we have $s_\tau\gamma(\mathbb{R}) = L_\gamma$.

Let $\phi : M \to \mathbb{R}$ be an observation function and $\tau \in \mathbb{R}$. We define the map $\theta : M \to \mathbb{R}^k$ by

$$\theta(\gamma) = \begin{pmatrix} \phi(s_\tau\gamma) & \phi(s_{2\tau}\gamma) & \dots & \phi(s_{k\tau}\gamma) \end{pmatrix}^T.$$

We can state Takens' embedding theorem [1] as follows:

Theorem 1: Takens' embedding

Let $k > 2m$.
For an open and dense subset of $C^\infty(M, \mathbb{R}) \times \Gamma(TM)$ and for any $\tau \in \mathbb{R}$, the map θ is an embedding.

On $\theta(M) \subset \mathbb{R}^k$, there exists a vector field Y defined by $Y = \theta_* X$. This is the conjugate dynamics of X on $\theta(M)$. Y is well defined since θ is a diffeomorphism.

2 Integrability of the image of a regular distribution

Let $\theta : M \to N$ be a diffeomorphism. Let X and Y be two non-vanishing vector fields. We have that $[\theta_* X, \theta_* Y] = \theta_*[X, Y]$. $\theta_* X$, $\theta_* Y$, and $\theta_*[X, Y]$ are all well-defined vector fields since θ is a diffeomorphism. Let $g \in C^\infty(N, \mathbb{R})$ and $p \in M$. Given any $X \in \Gamma(TM)$, we have that $(\theta_* X)_{\theta(p)}(g) = X_p(g \circ \theta)$ by definition of pushforward. This can be written as an equality between functions: $(\theta_* X)(g) \circ \theta = X(g \circ \theta)$. We have:

$$\begin{aligned}
(\theta_*[X, Y])_{\theta(p)}(g) &= [X, Y]_p(g \circ \theta) \\
&= X_p(Y(g \circ \theta)) - Y_p(X(g \circ \theta)) \\
&= X_p((\theta_* Y)(g) \circ \theta) - Y_p((\theta_* X)(g) \circ \theta) \\
&= (\theta_* X)_{\theta(p)}((\theta_* Y)(g)) - (\theta_* Y)_{\theta(p)}((\theta_* X)(g)) \\
&= [\theta_* X, \theta_* Y]_{\theta(p)}(g)
\end{aligned}$$

Let $\mathcal{F}$ be a foliation on M and P the associated integrable regular distribution. Since P is integrable, it is involutive i.e., given any $X_1, X_2 \in \Gamma(P)$, we have $[X_1, X_2] \in \Gamma(P)$. Now, define by $Q = \theta_*(P)$ a distribution on N i.e., $Y \in \Gamma(Q) \iff \exists X \in \Gamma(P), Y = \theta_* X$. Let $Y_1, Y_2 \in \Gamma(Q)$. There exists $X_1, X_2 \in \Gamma(P)$ such that $Y_i = \theta_*(X_i), i = 1, 2$. Hence $[Y_1, Y_2] = [\theta_* X_1, \theta_* X_2] = \theta_*[X_1, X_2]$ and since $[X_1, X_2] \in \Gamma(P)$ then $[Y_1, Y_2] \in \Gamma(Q)$. Since θ_* is an isomorphism, Q is a regular distribution. We have shown that it is involutive so by Frobenius' theorem it is integrable. Hence, there exists a foliation $\mathcal{G} = \theta(\mathcal{F})$.

3 Foliations on the input space of recurrent neural networks

3.1 First approach

Let $\theta : M \to \mathbb{R}^K$ be a Takens embedding and let $p(x, h, w)$ be a recurrent neural network parametrized by $x \in \mathbb{R}^K$, $h \in \mathbb{R}^H$, and w. The parameters w are the weights of the network and are assumed to be fixed so we will write $p(x, h, w) = (x, h) \in \Theta$ where $\Theta = \mathbb{R}^K \times \mathbb{R}^H$ is the parameter space. (x, h) defines a probability distribution over the sample space $\mathbb{R}^H$.

On Θ, define the sequence $p_t = (x_t, h_t)$ with $h_t = \int_{\mathbb{R}} p(x_{t-1}, h_{t-1}, w)(y)dy$ and $h_0 = 0$. The $K-1$ first components of x_t are equal to the $K-1$ last components of x_{t-1}, the last component of x_t is the next measure of the dynamical system. Define the log-likelihood by $l(y, x, h) = \log(p(x, h, w)(y))$. Define the score vector $s(x, h) = \left[(\nabla_x l)^T \quad (\nabla_h l)^T\right]^T = (\partial_i l(x, h))_{1 \leq i \leq K+H}$. The Fisher Information Matrix (FIM) is a $(K+H) \times (K+H)$ positive semi-definite matrix defined by $I(x, h) = \mathbb{E}_{(x,h)}[\partial_i l(x, h) \partial_j l(x, h)]_{ij}$.

3.2 Second approach

The parameter space is $\Theta = \mathbb{R} \times \mathbb{R}^H$ and $x_t \in \mathbb{R}$ is the measure at time t.

4 Questions

Quelle est la bonne approche ?
Quels sont les types de propriétés "métriques" habituellement étudiés ?
Pourquoi y a-t-il une place priviligiée pour les familles exponentielles et les familles de mélange ? Quelle est l'intérêt de la structure de connexions duales du point de vue statistique ?

References

[1] F. Takens, "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, vol. 898, pp. 366–381, 1981.