

UPV - CajonDesastre Ideas PINNs

2023-05-23 (Tuesday) | 16:43

Type: #notebook

Main: Main - UPV

Tags: #DeepLearning , #idea

Relates: ToDo Miscelanea, Topic - PINNS, UPV - PINNS

Status:

About

Todo lo que sobra y no se donde ponerlo para que no se pierda sobre ideas etc de PINNS y lo de la UPV.

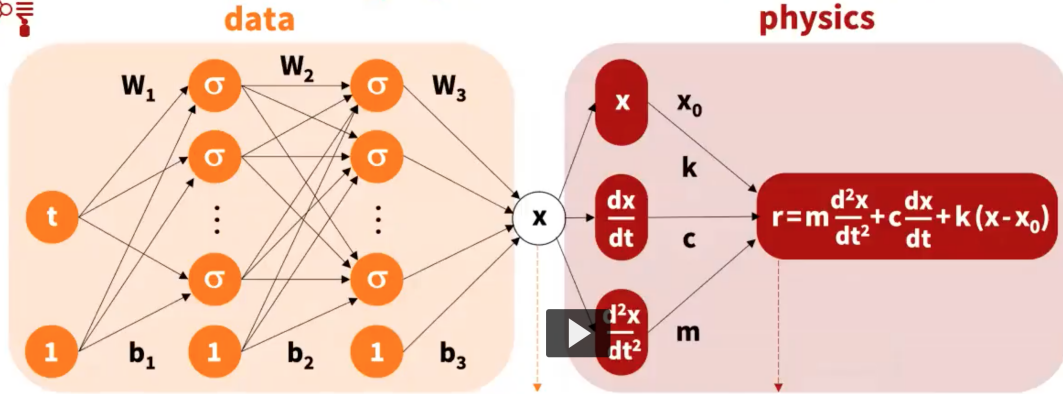
Ideas

- DL - GraphNeuralNetworks
- Uso de jax y cosas desarrolladas entorno a esto:
 - ! PUEDE QUE SEA MUY RELEVANTE SI QUEREMOS OPTIMIZAR Y ACELERAR
 - equinox: elegant NNs
 - DiffraX: numerical ODE/SDE solver.
- 🌟🌟 Temas sobre Neural Differential Equations -> <https://arxiv.org/pdf/2202.02435.pdf>
- Transformers con información física ¿? DL - PINNs + Transformers
- Las DeepONets parece algo tochísimo en la linea de las PINNs.
- Self-adaptative physics informed neural nets. Esto estaba entre lo que me paso MJ al principio de un curso de noseque:



self-adaptive physics informed neural nets

physics informed neural networks



- minimize loss function $L(\Theta; t) = (1 - \epsilon) L_{\text{data}} + \epsilon L_{\text{phys}} \rightarrow \min$
- error between model and data $L_{\text{data}}(\Theta; t) = \frac{1}{n_{\text{trn}}} \sum ||x(t_i) - \hat{x}_i||^2$
- error in physics residual $L_{\text{phys}}(\Theta; t) = \frac{1}{n_{\text{smp}}} \sum ||r(t_i)||^2$
- learn parameters and weight $\Theta = \{W_k, b_k, c, k, x_0\}$ and $\epsilon(t) \approx 10^{-3}$

En esta imagen vemos temas de self-adaptative PINNS donde la importancia de cada valor para la pinns se ajusta, esto podr a usarse para descubrir par metros o ecuaciones constitutivas.

  Ideas trabajo TFM Ruben PINNS >

En el trabajo de Ruben habla de las funciones de activaci n. Esto es muy importante para que las derivadas existan, y la soluci n est  bien construida en el espacio de las funciones de activaci n. Si nuestra soluci n es peri dica, alomejor ser a conveniente algo periodico como funci n de activaci n, no se.

$$\text{SoftModulusQ}(x) = \begin{cases} x^2(2 - |x|), & \text{if } x \leq 1 \\ |x|, & \text{if } x > 1 \end{cases} \quad (24)$$

$$\text{CAF}(x) = \gamma_1 \cos(x) + \gamma_2 \tanh(x) + \gamma_3 \sigma(x) + \gamma_4 \text{SoftModulusQ}(x) \quad (25)$$

Tabla 5: Funciones de activaci n utilizadas en las diferentes arquitecturas PINN para los diferentes experimentos.

Funci�n de activaci�n
$\cos(x)$
$\tan(x)$
$\sin(x)$
$\tanh(x)$
$\sigma(x)$
$ x \approx \text{SoftModulusQ}(x)$
$\text{CAF}(x)$

Tambi n pone limitaciones sobre la funci n de p rdida, pues tiene que cumplir: 1 con los

datos, 2 con la física, 3 con restricciones de la física que no están explícitamente en las ecuaciones, por ejemplo, la solución buscada en el caso de las órbitas del problema de ruben, no podía tener excentricidades negativas etc...

Respecto de la función de coste introducida para el entrenamiento consta de diferentes términos a identificar en la ecuación 28. Enunciémoslos y describamos brevemente su razón de ser.

1. Penalización de la predicción de la forma de onda \hat{y} respecto del valor exacto y . Este término hace que la red neuronal imite los datos de la forma de onda.
2. Penalización de soluciones no físicas mediante los términos $\langle (\dot{p})_+^2 \rangle$ y $\langle (\ddot{p})_+^2 \rangle$. De la ecuación 20, se ve que la distancia entre los dos agujeros negros es proporcional a r . Debido a la pérdida de energía de las ondas gravitacionales emitidas, $r(t)$ converge a cero a un ratio que crece con la evolución del sistema.
3. Penalización por excentricidades negativas ya que solo tienen sentido físico soluciones positivas.
4. Penalización por valores de *semi-latus rectum* p negativos.
5. Penalización de los pesos grandes.
6. Condición de estabilidad de la órbita $p - 6 - 2e \geq 0$. Esta puede deducirse de las soluciones 23 pues el contenido de la raíz $\sqrt{p - 6 - 2e \cos(\chi)}$ debe de ser positiva.

Realizamos una búsqueda de la mejor función de la tabla 6 para penalizar el error de los datos y la mejor propuesta $L1$ o $L2$ para la penalización de los pesos. Posteriormente, realizaremos una búsqueda por los hiperparámetros de la función $\gamma_i \forall i \in [1, 2, 3, 4, 5, 6]$.

$$L(y, w) = \frac{1}{N} \left(\gamma_1 L_{\text{datos}}(y, \hat{y}) + \gamma_2 \langle (\dot{p})_+^2 \rangle + \gamma_3 \langle (\ddot{p})_+^2 \rangle + \gamma_4 \sum_{i=1}^N (e - e_0)^2 \right) + \frac{1}{N} (\gamma_5 \max(e[p \geq 6 + 2e[1]], -e[1], 0) + \gamma_6 L_{\text{pesos}}(w)) \quad (28)$$

En la ecuación anterior hemos utilizado una notación vectorial de cómo se haría en código el término $\max(e[p \geq 6 + 2e[1]], -e[1], 0)$ para facilitar la comprensión de la expresión.

Esto se puede mejor en contexto en el trabajo.

- Topological features (TDA-Net): En este paper hablar de converting a given tensor (i.e., a multidimensional array such as a 2d image), to a vectorized representation of its persistence diagram. Como funciona esto? Parece muy interesante. -> [TDA-net.pdf](#)
Todo esto va en la línea del
- Las PINNs reducen el espacio de parámetros, al parecer hay otras formas de también hacer eso, noseque de los pesos en la esfera unidad: <https://arxiv.org/pdf/1805.08340.pdf>

- Añadir el Gradient Checking de alguna manera?

3.5 Good practices

❑ **Overfitting small batch** – When debugging a model, it is often useful to make quick tests to see if there is any major issue with the architecture of the model itself. In particular, in order to make sure that the model can be properly trained, a mini-batch is passed inside the network to see if it can overfit on it. If it cannot, it means that the model is either too complex or not complex enough to even overfit on a small batch, let alone a normal-sized training set.

❑ **Gradient checking** – Gradient checking is a method used during the implementation of the backward pass of a neural network. It compares the value of the analytical gradient to the numerical gradient at given points and plays the role of a sanity-check for correctness.

	Numerical gradient	Analytical gradient
Formula	$\frac{df}{dx}(x) \approx \frac{f(x+h) - f(x-h)}{2h}$	$\frac{df}{dx}(x) = f'(x)$
Comments	<ul style="list-style-type: none">- Expensive; loss has to be computed two times per dimension- Used to verify correctness of analytical implementation- Trade-off in choosing h not too small (numerical instability) nor too large (poor gradient approx.)	<ul style="list-style-type: none">- 'Exact' result- Direct computation- Used in the final implementation

- Imagen chula de la loss: (a) Schematic of a PINN for solving inverse problem in photonics based... | Download Scientific Diagram