PINNs #3

February 25, 2022

Ozgur Taylan Turan

Physics-Informed Kernel Regression

Past

Whole report is coming with the results of non-linear observations as well!

Physics-Informed Neural Networks

- Side-project after David's Coffee-talk

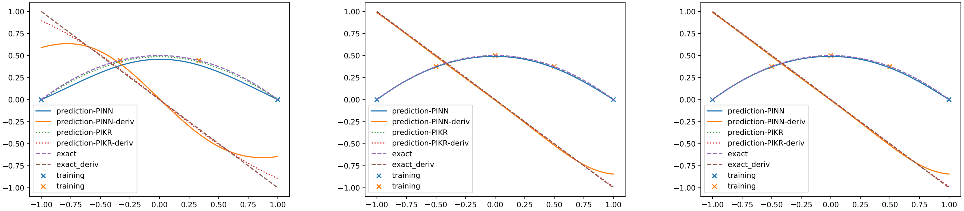
Main Idea:

- Solve $ODE : y'' = -1$ with BCs $y(-1) = 0$ and $y(1) = 0$, where $y(x)$ and $y' = \frac{dy}{dx}$
- Assume your solution $y(x)$ is given by a function $\mathcal{M}(x, \mathbf{w})$,
- Then,solution can be obtained by minimizing $\mathcal{L}_{\text{total}} := \mathcal{L}_{\text{domain}} + \mathcal{L}_{\text{boundary}}$
- $\mathcal{L}_{\text{domain}} := MSE(\mathcal{M}''(x) - 1)$
- $\mathcal{L}_{\text{bc}} := MSE(\mathcal{M}(-1) + \mathcal{M}(+1))$

It works!

- Works beautifully if you sample enough points (x)
- But, why just neural networks, we can put any model there?

Let's look at the kernelized Linear Regression



- If you don't sample enough data you do not simply satisfy the equation at given points although you have the flexibility!
- Is it a bad idea to use the same loss minimization for determining kernel parameters?

Why not use kernel methods not considered?

Connections to kernel methods

Many of the presented NN-based techniques have a close asymptotic connection to kernel methods, which can be exploited to produce new insight and understanding. For example, as demonstrated in REFS[76,77], the training dynamics of PINNs can be understood as a kernel regression method as the width of the network goes to infinity. More generally, NN methods can be rigorously interpreted as kernel methods in which the underlying warping kernel is also learned from data[78,79]. Warping kernels are a special kind of kernels that were initially introduced to model non-stationary spatial structures in geostatistics[80] and have been also used to interpret residual NN models[27,80]. Furthermore, PINNs can be viewed as solving PDEs in a reproducing kernel Hilbert space spanned by a feature map (parametrized by the initial layers of the network), where the latter is also learned from data. Further connections can be made by studying the intimate connection between statistical inference techniques and numerical approximation. Existing works have explored these connections in the context of solving PDEs and inverse problems[81], optimal recovery[82] and Bayesian numerical analysis[83–88]. Connections between kernel methods and NNs can be established even for large and complicated architectures, such as attention-based transformers[89], whereas operator-valued kernel methods[90] could offer a viable path of analysing and interpreting deep learning tools for learning nonlinear operators. In summary, analysing

Polynomial regression

PINNS w/ Polynomial Reg.

$y'' = -1 \longrightarrow y = -x^2/2 + c_1 x + c_2$

$\forall\ y(0) = 0$

$y(1) = 0 \Rightarrow \boxed{y_{analy} = 0.5(1 - x^2)}$

OKAY, Now lets try to apply PINN like method w/ model

$M(x, w) = w_3 + w_2 x + w_1 x^2 \Rightarrow$ Expressive enough for our ~~model~~, eqn.

- Loss "least squares":

Note $M''(w) = 2w_1$ & $L_{total} = L_{bound} + L_{domain}$ & take a point in domain...

$$L_{total} := \underbrace{(2w_1 + 1)^2}_{domain} + \underbrace{(w_3 - w_2 + w_1)^2 + (w_3 + w_2 + w_1)^2}_{Bound}$$

$\underset{w_1, w_2, w_3}{\mathrm{argmin}}\ L_{total} \Rightarrow w_1 = -1/2,\ w_2 = 0,\ w_3 = 1/2$

Thus, $y_{PINN} := 0.5(1 - x^2)$

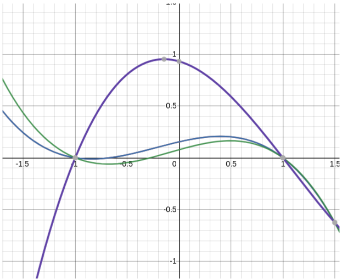
? But is it enough to have a function that is twice diff. for y'' of any type?

? What happens if you have more complex M?

↳ apparently the minimizer of your ~~func~~ loss does not satisfy the point constraint in the middle...

↳ you need regularization of large amounts to get reasonable solutions...

Polynomial regression, increase model complexity!



Regularize the loss from the domain!