Homotopy Perturbation Neural Networks

Preprint · July 2024

DOI: 10.13140/RG.2.2.16035.46885/3

1 author:

Gary Nan Tie
Mu Risk LLC

Homotopy Perturbation Neural Networks

Gary Nan Tie, July 15th 2024

Abstract: Inspired by He's homotopy perturbation method, we introduce a new deep learning technique to parsimoniously solve regression problems. Essentially, a homotopy is constructed that deforms a linear problem to our desired non-linear regression problem.

For reference, we recall the homotopy perturbation method [He, 1999] for solving general non-linear differential equations of the form:

(1) $$A(u) + f(r) = 0, \qquad r \in \Omega$$

with boundary conditions

$$B\left(u, \frac{\partial u}{\partial n}\right) = 0, \qquad r \in \Gamma$$

where A is a general differential operator, B is a boundary operator, $f(r)$ is a known analytic function, Γ is the boundary of the domain Ω.

Suppose $A = L + N$, where L is linear and N is non-linear.

So (1) becomes: (2) $$L(u) + N(u) - f(r) = 0$$

We construct a homotopy $v(r,p) : \Omega \times [0,1] \to \mathbb{R}$ which satisfies: (3) $$H(v,p) := (1-p)[L(v) - L(u_0)] + p[A(v) - f(r)] = 0$$

where $p \in [0,1]$, $r \in \Omega$ and u_0 is an initial approximation of (2), which satisfies the boundary conditions.

Note that, $H(v,0) = L(v) - L(u_0) = 0$

and $H(v,1) = A(v) - f(r) = 0$.

Now use the embedding parameter p as a 'small parameter' and assume that the solution of (3) can be written as a power series in p: $v = \sum_{i=0}^{\infty} p^i v_i$

Setting $p=1$ results in an approximate solution of (2):

$$(4) \quad u = \sum_{i=0}^{\infty} v_i$$

The series in (4) is convergent in most cases, and the rate of convergence depends on $A(v)$.

The homotopy perturbation method motivates the following:

given training data pairs (x_i, y_i) $\quad i = 0, 1, \ldots, n$

we want to learn a regression function $A(x) = y$.

Let L_φ be a linear L-layer FFN (feedforward network)

with parameters φ and identity activation function, and

let N_ψ be a non-linear M-layer FFN with

parameters ψ and activation function $\sigma \neq$ identity.

Suppose $X = \{x_1, x_2, \ldots, x_n\} \subset \mathcal{X} \subseteq \mathbb{R}^m$

define homotopy by: $\quad$ for $j = 0, 1, \ldots, k$

$$H_j : X \longrightarrow \mathbb{R}$$

$$x_i \longmapsto (1 - p_j)\left[L_\varphi(x_i) - L_\varphi(x_0)\right] + p_j\left[L_\varphi(x_i) + N_\psi(x_i) - y_i\right]$$

where $p_j \triangleq \dfrac{j}{k}$, $k \geqslant 2$; a discrete deformation

from $L(x_i) - L(x_0)$ to $L(x_i) + N(x_i) - y_i$.

On average we want $H_j(x_i) = 0$ with small dispersion.

So let $V_j \triangleq \sum_{i=1}^{n} | H_j(x_i) |^2$

and $L(\varphi, \psi) \triangleq \sum_{j=0}^{k} V_j + \lambda [L(x_0) + N(x_0) - y_0]^2, \lambda > 0$

and $L(\hat{\varphi}, \hat{\psi}) \triangleq \min_{\varphi, \psi} L(\varphi, \psi)$.

Then $L_{\hat{\varphi}}(x_i) + N_{\hat{\psi}}(x_i) \approx y_i \quad , \quad i = 1, 2, .., n$

since $H_k(x_i) \approx 0$.

Define regression function $A(x) = L_{\hat{\varphi}}(x) + N_{\hat{\psi}}(x) \quad , \quad x \in \mathcal{X}$.

Summary: We learn $\hat{\varphi}$ and $\hat{\psi}$ so that deformation

H_j from $L(x_i) - L(x_0)$ to $A(x_i) - y_i$

satisfies $H_j(x_i) \approx 0$ for $j = 0, 1, \dots, k$.

In particular, $H_k(x_i) \approx 0$, ie $L_{\hat{\varphi}}(x_i) + N_{\hat{\psi}}(x_i) \approx y_i$, $i = 1, .., n$

so define our regression function to be

$$A(x) \triangleq L_{\hat{\varphi}}(x) + N_{\hat{\psi}}(x) \quad \text{for } x \in \mathcal{X} \supset X.$$

Refining the initial guess x_0

Let $X = \{x_0, x_1, \ldots, x_n\}$ and $Y = \{y_0, y_1, \ldots, y_n\}$

For $j = 0, 1, \ldots, k$ with $k \geqslant 2$,

let
$$H_j(x_i) \triangleq (1-\rho_j)\left[L_\varphi(x_i) - L_\varphi(x_0)\right] + \rho_j\left[L_\varphi(x_i) + N_\psi(x_i) - y_i\right]$$

where $\rho_j \triangleq \frac{j}{k}$ and $x_i \in X \setminus \{x_0\}$,

and let $V_j \triangleq \sum_{x_i \in X \setminus \{x_0\}} |H_j(x_i)|^2$.

$$l(x_0, \varphi, \psi, \lambda) \triangleq \sum_{j=0}^{k} V_j + \lambda\left[L(x_0) + N(x_0) - y_0\right]^2, \quad \lambda > 0$$

$$l(x_0, \hat{\varphi}, \hat{\psi}, \lambda) \triangleq \min_{\varphi, \psi} l(x_0, \varphi, \psi, \lambda)$$

$$A[x_0, \hat{\varphi}, \hat{\psi}, \lambda](x) \triangleq L_{\hat{\varphi}}(x) + N_{\hat{\psi}}(x), \quad x \in \mathcal{X} \supset X$$

$x_{\hat{i}} \triangleq x_i \in X$ that minimizes $\left| A[x_0, \hat{\varphi}, \hat{\psi}, \lambda](x_i) - y_i \right|$

$L(x_0, \hat{\varphi}, \hat{\psi}, \lambda)$ learns $\hat{\varphi}$ and $\hat{\psi}$

over X and Y with initial guess x_0.

$L(x_{\hat{i}}, \hat{\varphi}^i, \hat{\psi}^i, \lambda)$ learns $\hat{\varphi}^i$ and $\hat{\psi}^i$

over X and Y with initial guess $x_{\hat{i}}$.

If $L(x_{\hat{i}}, \hat{\varphi}^i, \hat{\psi}^i, \lambda) < L(x_0, \hat{\varphi}, \hat{\psi}, \lambda)$

update initial guess x_0 to $x_{\hat{i}}$

and repeat this refinement until $L(x_{\hat{i}}, \hat{\varphi}^i, \hat{\psi}^i, \lambda)$

is sufficiently small for purpose or stops.

On the final iteration, use $A[x_{\hat{i}}, \hat{\varphi}^i, \hat{\psi}^i, \lambda]$

as the predictor function.

A HPNN generalization:

Analogous to Liao, S.J., 'Beyond Perturbation – Introduction to the Homotopy Analysis Method', Chapman and Hall / CRC, 2003, we introduce artificial degrees of freedom to the homotopy deformations, enlarging the solution space:

$$H_j : \{x_1, \dots, x_n\} \longrightarrow \mathbb{R}, \qquad j = 0, 1, \dots, k \qquad k \geq 2$$

$$x_i \longmapsto (1 - P_\alpha(p_j)) \, [L_\varphi(x_i) - L_\varphi(x_0)]$$
$$+ Q_\beta(p_j) \, [L_\varphi(x_i) + N_\psi(x_i) - y_i]$$

where learnable deformation FFNs, with input $\{p_j\}_{j=0}^{k}$, P_α and Q_β with parameters α and β respectively, satisfy $P_\alpha(0) = 0 = Q_\beta(0)$ and $P_\alpha(1) = 1 = Q_\beta(1)$, and $p_j \triangleq \frac{j}{k}$, $j = 0, \dots, k$.

Given training data $\{(x_i, y_i)\}_{i=0}^{n}$ we learn $\varphi, \psi, \alpha, \beta$ via homotopy deformations $H_j(x_i) \approx 0$ and so predictor $A(x) = L_{\hat{\varphi}}(x) + N_{\hat{\psi}}(x)$ for $x \in \mathcal{X} \supset \{x_0, \dots, x_n\}$.

A learning strategy for generalized HPNN:

The homotopy equations $H_j(x_i) = 0$

enable learning $\varphi, \psi, \alpha, \beta$ via SGD

with cost function $l(\varphi, \psi, \alpha, \beta)$.

① Set $P(p_j) = p_j = Q(p_j)$, learn $\mathring{\varphi}, \mathring{\psi}$

② Given $\mathring{\varphi}, \mathring{\psi}$, learn $\mathring{\alpha}, \mathring{\beta}$

③ Given $\mathring{\varphi}, \mathring{\psi}, \mathring{\alpha}, \mathring{\beta}$, learn optimal $\hat{\varphi}, \hat{\psi}, \hat{\alpha}, \hat{\beta}$

Yielding predictor $A = L_{\hat{\varphi}} + N_{\hat{\psi}}$

References

Ji-Huan He, 'Homotopy perturbation technique'
Comput. Methods Appl. Mech. Engrg., 178 (1999), pp. 257-262

Ji-Huan He, 'A coupling method of a homotopy technique and a perturbation technique for non-linear problems'
International Journal of Non-Linear Mechanics 35 (2000) 37-43