AI For Math

簡偉恆、盧詠涵

輔仁大學數學系資訊數學組一年級

2025 年 6 月 8 日

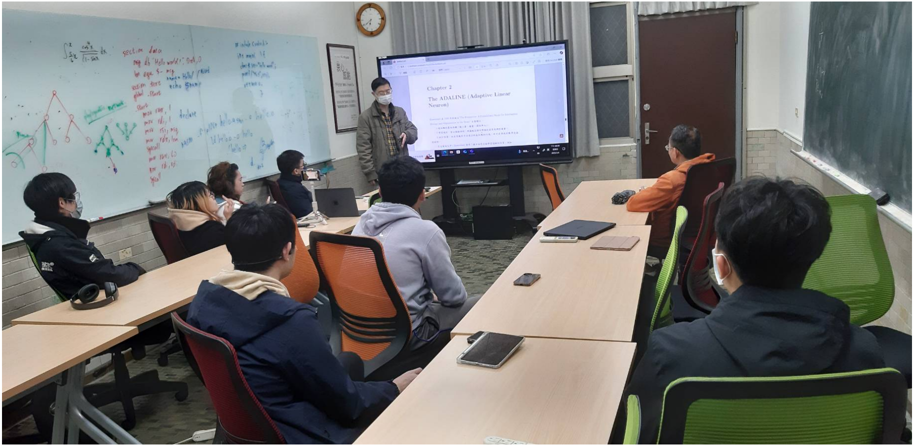
大綱

1. 社群活動簡介
2. Kolmogorov–Arnold Network 簡介
3. Kolmogorov–Arnold Network 實作
4. 記憶模型是否能降低 KAN 的誤差

113 年學年度第二學期 AI for Math 系列演講

日期	講者	講題
114/02/27	潘老師	感知機（The Perceptron）
114/03/06	潘老師	淺談 Adaptive Linear Neuron 和 Widrow-Hoff Learning
114/03/20	潘老師	The Basics of Multilayer Perceptron and Backpropagation
114/03/27	俞讚城教授	Introduction to Shannon Entropy and Cross Entropy
114/04/17	俞讚城教授	Introduction to Universal Approximation Theorems and Application in AI
114/05/08	嚴健彰教授	KAN: Kolmogorov-Arnold Networks

活動剪影

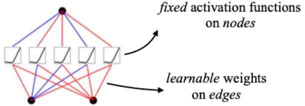
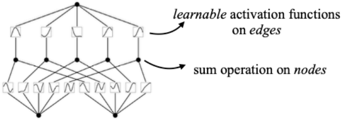
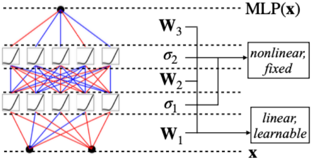
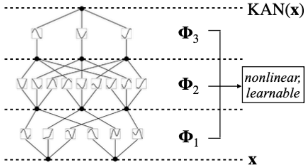


論文簡介

- 標題：KAN: Kolmogorov–Arnold Networks
- 作者：Ziming Liu 等人
- 主要特點：整合領域先驗知識與深度神經網路
- 應用領域：數學建模與科學計算
  https://arxiv.org/pdf/2404.19756

Kolmogorov–Arnold Network 簡介

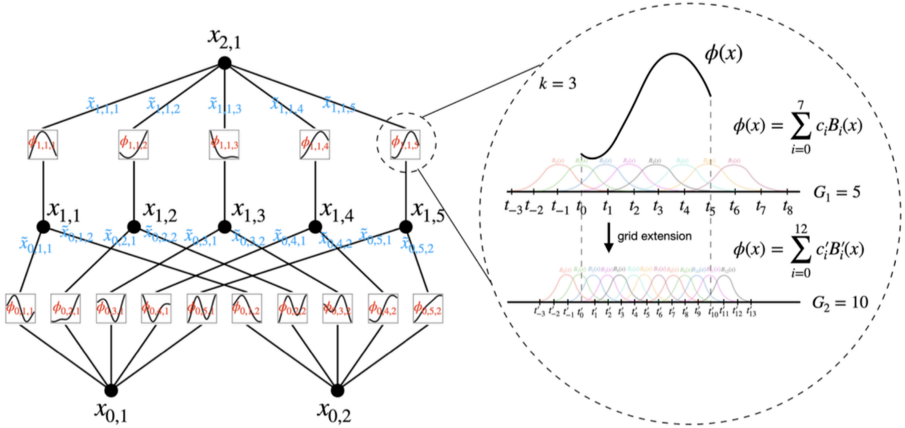
MLP vs. KAN

Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

Kolmogorov–Arnold Networks 架構與特點

- Kolmogorov–Arnold Network(KAN) 受 Kolmogorov-Arnold Representation theorem(KAT) 啟發
- 創新架構:
    - 可學習的一維激活函數位於邊上，取代傳統線性權重
    - 使用樣條函數（spline）參數化
    - 每個節點僅執行線性加總，不附加任何非線性激活函數

Kolmogorov–Arnold Networks 架構圖



定理 (Hilbert’s 13th problem)

Can the roots of the equation

$$x^7 + ax^3 + bx^2 + cx + 1 = 0$$

be represented as superpositions of continuous functions of two variables?

定理 (Kolmogorov 1956)

Any continuous function f of $n \in \mathbb{N}$ variables can be represented as a finite number of superpositions of functions of 3 variables. For instance, for $n = 4$ one has

$$f(x_1, x_2, x_3, x_4) \;=\; \sum_{i=1}^{4} g^i\big(u(x_1, x_2, x_3),\; v(x_1, x_2, x_3),\; x_4\big),$$

for some continuous functions $g^i,\; u,\; v : \mathbb{R}^3 \to \mathbb{R}$.

定理 (Kolmogorov–Arnold Representation Theorem)

For any continuous function $f(x_1, \ldots, x_n)$ defined on the $[0,1]^n$,

$$f(x_1, \ldots, x_n) \;=\; \sum_{q=1}^{2n+1} \Phi_q\Big(\sum_{p=1}^{n} \varphi_{q,p}(x_p)\Big)$$

where $\varphi_{q,p} : [0,1] \to \mathbb{R}$ are univariate continuous functions acting on each variable x_p and $\Phi_q : \mathbb{R} \to \mathbb{R}$ are univariate continuous functions acting on the sum of all $\varphi_{q,p}(x_p)$.

Kolmogorov–Arnold Networks 優勢

- 效能優勢:
  - 小規模 AI 任務中，參數量更少
  - 比 MLP 擁有更高精度
  - 更快的泛化縮放律
- 可解釋性:
  - 激活函數可視化
  - 可逐層稀疏修剪
  - 適用於（準）符號回歸與科學發現
- 應用優勢:
  - 可用於 PDE 求解（PINN 框架）
  - 連續學習中能有效避免遺忘現象
  - 結合樣條高精度與 MLP 組合結構

Kolmogorov–Arnold Networks 實作

```
[2,1,1] Test RMSE: 1.265029

=== [2,1,1] 測試點比較 ===
x=0.10, y=0.10  真實：1.375775  預測：3.680561  誤差：2.304786
x=0.50, y=0.50  真實：3.490343  預測：2.694129  誤差：0.796214
x=0.75, y=0.25  真實：2.158917  預測：2.694129  誤差：0.535212
```

▲ [2,1,1] 測試點比較

```
[2,5,1] Test RMSE: 1.445008

=== [2,5,1] 測試點比較 ===
x=0.10, y=0.10  真實：1.375775  預測：2.344764  誤差：0.968989
x=0.50, y=0.50  真實：3.490343  預測：3.075765  誤差：0.414578
x=0.75, y=0.25  真實：2.158917  預測：2.608403  誤差：0.449486
```

▲ [2,5,1] 測試點比較

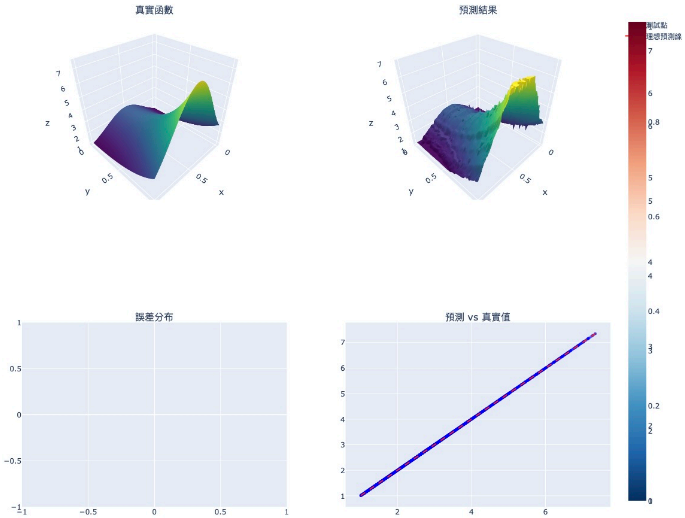
Kolmogorov–Arnold Networks 實作

```
[2,1,1] Test RMSE: 0.002997

=== [2,1,1] 測試點比較 ===
x=0.10, y=0.10  真實：1.375775  預測：1.376149  誤差：0.000375
x=0.50, y=0.50  真實：3.490343  預測：3.486988  誤差：0.003355
x=0.75, y=0.25  真實：2.158917  預測：2.158463  誤差：0.000454
```

▲ [2,1,1] 測試點比較 2

Kolmogorov–Arnold Networks 實作



▲ [2,1,1] 測試點比較 2-視覺化

Kolmogorov–Arnold Networks 實作

```
[2,1,1] Test RMSE: 0.003584

=== [2,1,1] 測試點比較 ===
x=0.10, y=0.10  真實：1.375775  預測：1.376706  誤差：0.000932
x=0.50, y=0.50  真實：3.490343  預測：3.489892  誤差：0.000452
x=0.75, y=0.25  真實：2.158917  預測：2.158352  誤差：0.000565
```

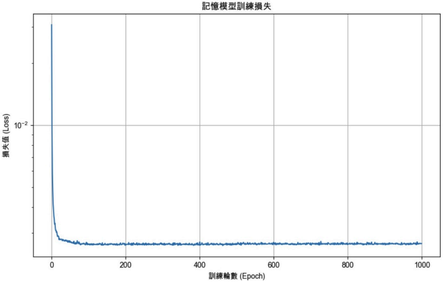
```
[2,1,1] Test RMSE: 1.112584

=== [2,1,1] 測試點比較 ===
x=0.10, y=0.10  真實：1.375775  預測：2.449288  誤差：1.073513
x=0.50, y=0.50  真實：3.490343  預測：2.449288  誤差：1.041055
x=0.75, y=0.25  真實：2.158917  預測：2.449288  誤差：0.290371
```

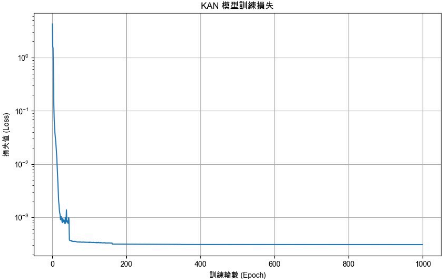
```
[2,1,1] Test RMSE: 0.002997

=== [2,1,1] 測試點比較 ===
x=0.10, y=0.10  真實：1.375775  預測：1.376149  誤差：0.000375
x=0.50, y=0.50  真實：3.490343  預測：3.486988  誤差：0.003355
x=0.75, y=0.25  真實：2.158917  預測：2.158463  誤差：0.000454
```

記憶模型是否能降低 KAN 的誤差

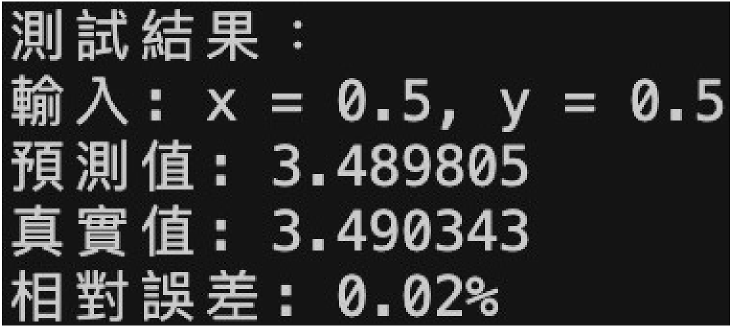


▲ 記憶功能訓練結果



▲ KAN 訓練結果

記憶模型能不能降低 KAN 的誤差



▲ 增加記憶功能後的結果數據

謝謝聆聽