

1.Unanswered Questions

1. 為什麼要用 sigmoid 函數？Sigmoid 函數有什麼特性及好處？

Cramer, J.S. (2002). The Origins of Logistic Regression. Tinbergen Institute Discussion Paper, TI 2002-005/4.[02119.pdf](#)

回顧並詳細介紹了邏輯迴歸和 Sigmoid 函數在統計學中的應用和歷史發展。

2.

$$\text{Cost} \left(W^{[2]}, W^{[3]}, W^{[4]}, b^{[2]}, b^{[3]}, b^{[4]} \right) = \frac{1}{10} \sum_{i=1}^{10} \left(\frac{1}{2} \|y(x^{[i]}) - F(x^{[i]})\|_2^2 \right)^{\text{平方}}$$
$$\|v\|_2 = \sqrt{v_1^2 + \dots + v_n^2}$$

Here, the factor $\frac{1}{2}$ is included for convenience; it simplifies matters when we start differentiating. We emphasize that Cost is a function of the weights and biases—the

微分的用意是什麼？是為了找cost的最小值嗎

對。加上 $1/2$ 的作用就是抵消微分時產生的 2 係數，使最終的梯度表達式更簡潔，避免了在每一次梯度計算中都帶上一個額外的 2 係數。這純粹是數學上的便利性，並不會改變 Cost 函數的最小值位置。

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to Linear Regression Analysis. Wiley.迴歸分析的標準教材。迴歸分析的核心就是最小化殘差平方和 (Sum of Squared Errors, SSE)，也就是 Cost 函數的平方項部分。 $1/2$ 僅是為了微分方便而引入的係數。 [Introduction to Linear Regression Analysis](#)

3. Lemma 3.1 和 Lemma 3.2 證明了淺層 tanh 神經網路可以任意精度地逼近任意次數的單變數多項式 $f_p(y) = y^p$ 這種使用 $W^{[k]}$ 範數來控制導數誤差的逼近，對於深度學習模型具有什麼重要的理論意義？

文獻片段中使用 $W^{[k]}$ 範數來限制 0 到 k 階導數的誤差，這證明了神經網路不僅能學習函數本身，還能學習函數的形狀（即導數），這對於科學計算和物理模型非常重要。

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks, 3(5): 551 – 560. [Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks - ScienceDirect](#) 這篇論文將 UAT 擴展到逼近函數的導數。它證明了使用光滑激活函數（如 tanh）的神經網路不僅可以逼近函數 f ，還可以逼近其導數 $D^\alpha f$ 。

4. 若想用機器學習來預測傑出操盤手的行為，我們需要哪些資料樣本？

Modeling Traders' Behavior with Deep Learning and Machine Learning Methods:

Evidence from BIST 100 Index ([PDF](#)) Modeling Traders' Behavior with Deep Learning and Machine Learning Methods: Evidence from BIST 100 Index

5.

我上次作業有用到Bootstrap Aggregating，其中使用隨機子空間法，令原始空間維度為 p ，則子空間維度通常為 $\lfloor \sqrt{p} \rfloor$ ，我不太明白這樣選擇子空間維度的用意，有什麼特性或好處嗎？

Breiman, L. (2001). _Random Forests_. Machine Learning, 45(1): 5 – 32. [Random Forests | Machine Learning](#) 這篇論文是隨機森林的正式介紹。作者 Leo Breiman 提出了隨機森林的核心思想 (Bagging + 隨機子空間)，並在論文中通過經驗性測試，建議將分類問題的特徵子集大小設為 $\text{int}(\sqrt{p})$ ，以達到最佳的準確性和去相關性。

6. 如果數據點的分布不是類似橢圓，使用 GDA 的效果可能會很差，有沒有可能先將數據點映射成類似橢圓的分布再使用 GDA？

Box, G. E. P., & Cox, D. R. (1964). _An Analysis of Transformations_. Journal of the Royal Statistical Society. Series B (Methodological), 26(2): 211 – 252. [An Analysis of Transformations](#) 提出了 Box-Cox 變換，這是一種確定最佳幕次變換 x^{λ} (其中 λ 為參數) 的技術，其目標是使變換後的數據最接近常態分佈，這可以作為 GDA/QDA 的預處理步驟。

7.

Q: 能說 $p_{\text{data}}(x)$ 靠近 $\begin{cases} 1 & \Rightarrow \text{像} \\ 0 & \Rightarrow \text{不像} \end{cases}$ 嗎？ PDF $0 \leq p(x) \leq 1$ ？

A: No! $p(x)$ 有可能大於 1.

Why? $p(x)$ 是機率密度，不是機率

For example, the continuous uniform distribution on the interval $[0, \frac{1}{2}]$

has probability density function $p(x) = \begin{cases} 2 & \text{for } 0 \leq x \leq \frac{1}{2} \\ 0 & \text{o.w.} \end{cases}$

(check $\int_{-\infty}^{\infty} p(x) dx = \int_0^{\frac{1}{2}} 2 dx = 1$)

[Probability density function - Wikipedia](#) 的 Further details 有提到

8. 布朗運動在任何時間間隔內的軌跡是無限的嗎？這在物理上如何解釋？

Karatzas, I., & Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus. Springer. (第 2 章) [Brownian Motion and Stochastic Calculus | SpringerLink](#) 這是隨機分析領域的權威教材。它在嚴格的數學框架下證明了布朗運動的軌跡幾乎必然是處處不可微分的，並且具有無限變差，從而證實了軌跡長度是無限的。Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik, 322(8): 549 – 560. [Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen](#) 愛因斯坦的這篇開創性論文首次從理論上解釋了布朗運動，將其與分子動能論聯繫起來。這是布朗運動物理學研究的起點。

9. 在「模仿特定大師風格的短篇續寫」這個簡化模型中，如何量化所謂的風格、情感深度，從而讓機器理解什麼是「好」

Christiano, P. F., Leike, J., Brown, T., Chan, K., Beattie, M., Hubis, N., ... & Amodei, D. (2017). Deep reinforcement learning from human preferences. Advances in Neural Information Processing Systems (NeurIPS), 30. [Deep Reinforcement Learning from Human Preferences](#) 這篇論文奠定了 RLHF 的理論基礎。它展示了如何通過訓練一個獎勵模型來學習人類對複雜任務（如機器人運動或文字生成）的偏好排序，並以此指導深度強化學習模型。這是量化主觀「好」的根本方法。

2.Solvable Model Problem in the final project

我們將模擬一個簡化的鋁-銅-鎂 (Al-Cu-Mg) 三元合金系統。

- 輸入 (Input): 兩個變數 x_1 (銅 Cu 的百分比), x_2 (鎂 Mg 的百分比)。鋁 (Al) 的含量則為 $1 - x_1 - x_2$ 。
- 任務目標：預測該合金的「硬度 (Hardness)」，並找出能達到最大硬度的成分比例。
- 資料形式：數值型資料。

1. 虛擬物理環境構建 (The Synthetic Ground Truth)

在現實世界中，獲取合金數據需要熔煉與測試（耗時昂貴）。但在教學用的「模型問題」中，我們必須創造一個**「上帝視角函數 (Oracle Function)」**來產生訓練數據並驗證結果。

我們設計的 `ground_truth_hardness` 函數並非隨意亂寫，而是為了模擬真實材料科學中的三種物理機制：

$$y_{hardness} = f_{base} + f_{solution} + f_{precipitate} + f_{defect}$$

- f_{base} (基底硬度):

純鋁原本的硬度。這是一個常數項 (Bias)。

- $f_{solution}$ (固溶強化, Solid Solution Strengthening):

- 物理意義：當銅 (Cu) 或鎂 (Mg) 原子溶入鋁晶格中，會造成晶格扭曲，阻礙差排 (Dislocation) 移動，使材料變硬。

- 數學模擬：這通常大致呈線性關係。

- 公式： $w_1 \cdot x_{Cu} + w_2 \cdot x_{Mg}$

- $f_{precipitate}$ (析出強化, Precipitation Hardening): 這是鋁合金最強化的關鍵。

- 物理意義：只有在特定的成分比例 (例如 Al₂CuMg 相) 下，金屬原子會聚集成微小的顆粒，產生強大的強化效果。如果比例不對，這些顆粒就不會形成。

- 數學模擬：這是一個非線性、局部隆起的峰值。我們使用高斯函數 (Gaussian Function) 來模擬這個「甜蜜點 (Sweet Spot)」。

- 公式： $A \cdot \exp\left(-\frac{(x_{Cu} - \mu_1)^2}{\sigma_1^2} - \frac{(x_{Mg} - \mu_2)^2}{\sigma_2^2}\right)$

- f_{defect} (過量缺陷)：
 - 物理意義：如果添加太多雜質，可能會形成粗大的化合物，反而讓材料變脆或性質下降。
 - 數學模擬：一個高階的負向懲罰項。

為什麼這樣設計？

這強迫模型不能只學會「加越多越好 (線性關係)」，而必須學會「找到中間的最佳平衡點 (非線性關係)」。

2. 機器學習模型：多層感知機 (MLP Regressor)

我們選擇使用 Scikit-Learn 的 MLPRegressor，這是一個標準的前饋神經網路 (Feedforward Neural Network)。以下是其架構與選擇理由的詳細分析：

A. 模型架構 (Architecture)

這個模型可以被視為一個函數逼近器 $f_\theta(x)$ ，其結構如下：

1. 輸入層 (Input Layer)：

- 神經元數量：2 個 (x_1 : Cu%, x_2 : Mg%)。
- 處理：在進入網路前，數據經過了 StandardScaler (標準化)，將數值調整到平均值為 0、標準差為 1 的分佈。這是因為 Cu 的數值很小 (0.04)，而硬度數值很大 (100)，如果不標準化，梯度下降 (Gradient Descent) 會極難收斂。

2. 隱藏層 (Hidden Layers)：

- 第一層：64 個神經元。
- 第二層：32 個神經元。
- 激活函數 (Activation Function)：ReLU (Rectified Linear Unit)，即 $f(z) = \max(0, z)$ 。
- 關鍵作用：如果沒有激活函數，無論有多少層，神經網路都只是一個線性模型。引入 ReLU 讓模型能夠「彎曲」決策邊界，從而擬合我們上述設計的那個「高斯峰值」。

3. 輸出層 (Output Layer)：

- 神經元數量：1 個 (預測的硬度值 y)。
- 激活函數：Identity (線性輸出)，因為這是一個回歸問題，我們需要輸出一種連續的實數值，而非機率。