Assignment 11

1.

Question from assignment 1:

Beside the sigmoid function, is there any other commonly used activation functions? What are their respective limitation, pro and cons?

Answer:

- ReLU and its variants are widely used in hidden layers due to efficiency and reduced vanishing gradient issues.
- tanh is sometimes preferred for small networks or zero-centered outputs.
- Softmax is specialized for output layers in classification tasks.
- Each activation function has trade-offs between gradient behavior, output range, and computational cost.

Question from assignment 10:

1. How should the model be rewritten if a Poisson jump process is added to the SDE (Stochastic Differential Equation)?

2. In high-dimensional systems, how does the correlation of

the diffusion matrix affect stochastic dynamics?

Answer:

1. 1. 如果在 SDE 中加入 Poisson 跳躍過程 (Jump process)，模型該如何改寫？

當你在標準的擴散過程（由布朗運動 $\mathrm{d}W_t$驅動）中加入跳躍（由泊松過程 $\mathrm{d}N_t$ 驅動）時，該模型被稱為 **跳躍-擴散過程 (Jump-Diffusion Process)**。

模型需要改寫為包含一個 **跳躍項 (Jump term)**。

數學表達式：

標準的 SDE 為：

$$\mathrm{d}X_t = \mu(X_t, \mathrm{t})\mathrm{dt} + \sigma(X_t, t)dW_t$$

加入 Poisson 跳躍後，方程式改寫為：

$$\mathrm{d}X_t = \mu(X_t, \mathrm{t})\mathrm{dt} + \sigma(X_t, t)dW_t + Y_t dN_t$$

各項含義：

- N_t **(Poisson Process)**:這是一個計數過程，參數為強度λ。

    - 在極短時間 dt 內，發生跳躍的機率為 $\lambda \mathrm{dt}$ (即 $\mathrm{dN}_t = 1$)。
    - 沒有跳躍的機率為 $1 - \lambda \mathrm{dt}$ (即 $\mathrm{dN}_t = 0$)。
- Y_t **(Jump Size)**: 代表跳躍的幅度。這通常是一個隨機變數，服從某個機率分佈 (例如常態分佈或雙指數分佈)。
    - 當跳躍發生時 ($\mathrm{dN}_t = 1$)，系統狀態 X_t 會瞬間增加Y_t。

應用範例 (Merton 模型)：

在金融數學中，若股票價格除了波動外還可能發生崩盤，我們會寫成：

$$\frac{\mathrm{d}S_t}{S_t} = (\mu - \lambda k)dt + \sigma dW_t + (e^J - 1)dN_t$$

這描述了資產價格在連續路徑上偶爾會發生不連續的劇烈變化。

---

2. 在高維系統中，擴散矩陣的相關性如何影響隨機動態？

在高維系統 (例如$X_t \in \mathbb{R}^n$) 中，SDE 寫作：

$$\mathrm{d}X_t = \mu(X_t)dt + \sum(X_t)dW_t$$

其中 W_t 是多維布朗運動。擴散矩陣 (Diffusion Matrix) 通常指協方差結構 $\mathrm{D}=\sum\sum^T$。其相關性 (Correlation) 體現在矩陣的 非對角線元素 (Off-diagonal elements)。

相關性對隨機動態的影響主要體現在以下三點：

A. 狀態變數的耦合 (Coupling of Trajectories)

- **獨立系統 (無相關)**：如果擴散矩陣是對角的 (Diagonal)，各個維度的雜訊互不干擾，X_1的隨機波動不會影響X_2。
- **相關系統**：如果存在相關性，一個維度的隨機衝擊 (Shock) 會立即傳播到其他維度。
    - **正相關**：變數傾向於「同漲同跌」。例如在生態系統中，捕食者和獵物若受相同環境雜訊影響，種群數量可能同步波動。
    - **負相關**：變數傾向於反向變動。

B. 機率分佈的幾何形狀 (Geometry of Distribution)

- 相關性改變了系統狀態在空間中擴散的形狀 (Anisotropy)。
- 如果不相關，機率密度函數 (PDF) 的等高線可能是正圓形（或超球體）。

- 如果有高度相關性，PDF 會被拉伸成細長的 **橢球體 (Hyper-ellipsoid)**。這意味著系統極高機率會沿著特定的「特徵方向」演化，而在其他方向受到限制。

C. 系統性風險與極端事件 (Systemic Risk & Rare Events)

- 在高維系統中，相關性決定了 **聯合極端事件 (Joint Extreme Events)** 發生的機率。
- 例如在金融風險管理中，如果資產間的擴散矩陣高度相關，當市場受到衝擊時，所有資產會同時下跌（系統性風險），導致比獨立系統更嚴重的損失。
- 在物理擴散中，相關性可能影響粒子逃離某個區域的 **首達時間 (First Passage Time)**。

2.

在 20 年後我希望 AI 擁有能夠讓遊戲中的每個 NPC 結合智能，擁有獨立思考以及互相互動的能力。

我們先考慮一個問題，假設出現一個新的物品，而只有其中一個人知道了這件物品長甚麼樣，而其他人並不知道也從未見過這種物品，其他人必須透過詢問並猜測來得知此物品的大略長相，例如在所有 AI NPC 中，每個人都不認識貓這種生物的情況下，只有一

個人看到了他的長相(先假設只看到一次貓就消失了)，隨後因沒有任何 NPC 看過這生物，在其它人想要了解這個新生物的外貌便只能先隨意猜測一種生物或特徵例如跟狗長得很像，在詢問發現答案錯誤之後並了解大概與貓差多少後不斷透過修改特徵(耳朵長度、眼睛位置、鬍鬚有無等)，來成功學習到貓的各項特徵分布。這能夠賦予 NPC 社會中一種知識傳播的能力，如同人類社會中，有人說看到了外星人，我們因為沒有實際見過無法得知其面貌，只好透過不斷詢問來了解他所看到的”外星人”長相，也許不全然相同但也大差不差。

為了達到這種效果，我先將問題簡化成猜測一個由 RGB 色盤組成的顏色，這裡以紅色舉例是[255, 0, 0]。

為了帶入先前的假設為遊戲內 NPC，因此在模型中有兩個角色:一個是導師，負責告知顏色是否一樣以及與正確答案的誤差為多少，以 RGB 色盤來說可以是三種顏色個別差平方和開根號，對應程式就是所謂的 Loss function。另一個是學生，負責透過反覆詢問來猜測出導師所說的是哪種顏色，再以程式來說便是簡易神經網路，透過梯度下降法等數學方法不斷計算與微調，來從錯誤中收斂出正確的答案。

這個簡易模型在現行技術中很明顯是可行的，隨著研究進步可

以將單個顏色逐漸改為 2*2 像素、4*4 像素等直到能夠判斷出所謂的特徵，再者，若能結合大型語言系統，理想上便能達成擁有獨立思考與真實互動的智能 AI。