

HW1-1: 網格地圖開發

1. 目標：開發一個大小為 $n \times n$ 的網格地圖，允許用戶指定維度 'n' (範圍從 5 到 9)，使用 Flask 網頁應用程式。
2. 功能要求：
 - 用戶可以透過滑鼠點擊指定起始單元格 (點擊後顯示為綠色) 和結束單元格 (點擊後顯示為紅色)。
 - 用戶可以設定 $n-2$ 個障礙物，透過滑鼠點擊將這些單元格變為灰色。

Prompt:

Prompt HW1-1：網格地圖互動式設定

請幫我用 Python 搭配 Flask 架設一個網頁應用，完成以下功能：

🎯 功能目標

建立一個可互動的網格地圖 (Gridworld) 設定介面，讓使用者可以：

1. 輸入一個整數 'n'，生成 ' $n \times n$ ' 的網格 (限制 5~9)
2. 點擊格子依序指定：
 - 一個起點 (顯示為綠色)
 - 一個終點 (顯示為紅色)
 - 最多 $n-2$ 個障礙物 (顯示為灰色)

☑ 操作邏輯要求

- 點擊第一次設定為起點
- 點擊第二次設定為終點
- 後續點擊可設定障礙物 (最多 $n-2$ 個)
- 點擊已設定的格子可取消該狀態
- 每次重新生成網格時要清除所有設定

🖥️ 技術需求

- Flask 後端
- HTML + CSS + JavaScript (含 jQuery)
- 使用 'Flask render_template()' 傳送 HTML
- 資料更新可用 '/update_grid' POST API 傳送 JSON 格子資訊

Result:

HW1：網格地圖 + 策略與價值估計

輸入網格大小（5-9）：

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

HW1-2: 政策表示與評估

1. 隨機政策顯示：

- 為格子中的每個格子顯示一個隨機行動作為政策，行動可以用箭頭表示，指示：
 - 上，下，左，右（每個格子對應一個行動）。

2. 政策評估：

- 實現政策評估，計算並顯示每個狀態（格子）的價值函數 $V(s)$ 。
- 在格子中顯示價值評估結果，每個格子的值表示在當前政策下從該狀態開始的期望回報。

Prompt:

Prompt HW1-2：策略顯示與價值函數估計

請幫我在 HW1-1 的基礎上，加入策略與價值估計的功能（策略評估）。

🎯 功能目標

- 根據目前使用者設定的網格地圖（包含起點、終點、障礙物）
- 自動隨機產生一份策略（每格一個動作 $\uparrow \downarrow \leftarrow \rightarrow$ ）
- 執行策略評估，估算每個狀態 $V(s)$

☑ 策略邏輯要求

- 每格選擇一個合法的方向（不能走進障礙或牆）
- 終點與障礙格不產生策略（顯示空白）

☑ 策略評估邏輯

- 折扣因子 $\gamma = 0.9$
- 每步獎勵 $R = -1$
- 終點價值固定為 $V(s) = 0$

其中 s 為根據當前策略選擇的下一個狀態

🖨 顯示需求

- 於網頁下方顯示：
 - **Policy Matrix****（每格箭頭）與
 - **Value Matrix****（每格估算後的 V 值，小數顯示至小數第二位）
- 起點、終點、障礙物顏色需與上方網格同步標示

Result:

HW1：網格地圖 + 策略與價值估計

輸入網格大小 (5-9) :

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

策略與價值矩陣

Policy Matrix

←	→	↓	←	←
↑	←	←		→
↑	←		→	↑
←		↑	↓	↑
↓	↓	↑	→	

Value Matrix

-9.95	-9.95	-9.95	-9.95	-9.95
-9.95	-9.95	-9.95		-9.95
-9.95	-9.95		-9.95	-9.95
-9.95		-9.95	-1.90	-9.95
-9.95	-9.95	-9.95	-1.00	0.00