MATH 485 - Quiz #2

Question 1.)

From → To	Sunny	Cloudy	Rainy
Sunny	0.33	0.67	0.00
Cloudy	0.33	0.00	0.67
Rainy	0.33	0.33	0.33

Weather → Behavior	Walk	Umbrella
Sunny	4/4 = 1.0	0/3 = 0.0
Cloudy	2/3 ≈ 0.67	1/3 ≈ 0.33
Rainy	1/3 ≈ 0.33	2/3 ≈ 0.67

Day	Observation	? →	Sunny	Cloudy	Rainy
		$V0(?)$	0.33	0.33	0.33
1	**Walk**	$P(W\|?)$	1.00	0.67	0.33
		$V1(?) = V0(?) * P(W\|?)$	0.33	0.22	0.11
2		$V1(S) * P(?\|S)$		0.22	0.00
		$V1(C) * P(?\|C)$		0.00	0.15
		$V1(R) * P(?\|R)$		0.04	0.04
	Umbrella	$P(U\|?)$	0.00	0.33	0.67
		$V2(?) = max(?) * P(U\|?)$	0.00	0.07	0.10
3		$V2(S) * P(?\|S)$	0.00	0.00	0.00
		$V2(C) * P(?\|C)$	0.02	0.00	0.05
		$V2(R) * P(?\|R)$	0.03	0.03	0.03
	Walk	$P(W\|?)$	1.00	0.67	0.33
		$V3(?) = max(?) * P(W\|?)$	0.03	0.02	0.02

Markov Decision Process (MDP)

Value Iteration Process with Policy Changes in MDP

We begin with a Markov Decision Process (MDP) where an agent decides whether to invest conservatively (C) or aggressively (A) in a financial portfolio. The objective is to find an optimal policy maximizing long-term rewards.

Step 1: Defining the MDP Components

States (S):

- Low Wealth (L)
- Medium Wealth (M)
- High Wealth (H)

Actions (A):

- Conservative (C)
- Aggressive (A)

Transition Probabilities:

Current State	Action	Next State Probabilities
Low (L)	C	80% Stay in L, 20% Move to M
Low (L)	A	60% Stay in L, 40% Move to M
Medium (M)	C	70% Stay in M, 30% Move to H
Medium (M)	A	50% Stay in M, 50% Move to H
High (H)	C	90% Stay in H, 10% Drop to M
High (H)	A	70% Stay in H, 30% Drop to M

Rewards:

- Low Wealth (L): -1
- Medium Wealth (M): 3
- High Wealth (H): 5

Discount Factor (γ): 0.9

Step 2: Value Iteration Updates

We initialize values: $V_0(L) = 0$, $V_0(M) = 0$, $V_0(H) = 0$.

In [34]:
```
v0_L = 0
v0_M = 0
v0_H = 0
```

Iteration 1

Using Bellman's equation:

> $$V_1(s) = \max_a \left[R(s) + \gamma \sum_{s'} P(s'|s,a) V_0(s') \right]$$

For **Low Wealth (L):**

> $$V_1(L) = \max\left[-1 + 0.9(0.8V_0(L) + 0.2V_0(M)), -1 + 0.9(0.6V_0(L) + 0.4V_0(M))\right]$$

For **Medium Wealth (M):**

> $$V_1(M) = \max\left[3 + 0.9(0.7V_0(M) + 0.3V_0(H)), 3 + 0.9(0.5V_0(M) + 0.5V_0(H))\right]$$

For **High Wealth (H):**

> $$V_1(H) = \max\left[5 + 0.9(0.9V_0(H) + 0.1V_0(M)), 5 + 0.9(0.7V_0(H) + 0.3V_0(M))\right]$$

In [35]:
```
# For Low Wealth (L)
vL_func = lambda l, m: (max(-1 + 0.9*(0.8*l + 0.2*m), -1 + 0.9*(0.6*l + 0.4*m)))
# For Medium Wealth (M)
vM_func = lambda m, h: (max(3 + 0.9*(0.7*m + 0.3*h), 3 + 0.9*(0.5*m + 0.5*h)))
# For High Walth (H)
vH_func = lambda h, m: (max(5 + 0.9*(0.9*h + 0.1*m), 5 + 0.9*(0.7*h + 0.3*m)))
```

Since $V_0(L) = V_0(M) = V_0(H) = 0$, the initial values are just the rewards.

> $$V_1(L) = -1, \quad V_1(M) = 3, \quad V_1(H) = 5$$

In [36]:
```
v1_L =  vL_func(v0_L, v0_M)
v1_M =  vM_func(v0_M, v0_H)
v1_H = vH_func(v0_H, v0_M)
print(f"V₁(L) = {v1_L},  V₁(M) = {v1_M},  V₁(H) = {v1_H},")
```

```
V₁(L) = -1.0,  V₁(M) = 3.0,  V₁(H) = 5.0,
```

Policy Evaluation after Iteration 1

For **Low Wealth (L):**

> $$Q(L, C) = -1 + 0.9(0.8(-1) + 0.2(3)) = -1.18$$

> $$Q(L, A) = -1 + 0.9(0.6(-1) + 0.4(3)) = -0.46$$

For **Medium Wealth (M):**

> $$Q(M, C) = 3 + 0.9(0.7(3) + 0.3(5)) = 6.24$$

> $$Q(M, A) = 3 + 0.9(0.5(3) + 0.5(5)) = 6.60$$

For **High Wealth (H):**

> $$Q(H, C) = 5 + 0.9(0.9(5) + 0.1(3)) = 9.32$$
>
> $

> $$Q(H, A) = 5 + 0.9(0.7(5) + 0.3(3)) = 8.96$$

In [37]:
```
# Defning funcitons
Q_LC = lambda l, m: -1 + 0.9*(0.8*l + 0.2*m)    # Q(L,C)
Q_LA = lambda l, m: -1 + 0.9*(0.6*l + 0.4*m)    # Q(L,A)
Q_MC = lambda m, h: 3 + 0.9*(0.7*m + 0.3*h)     # Q(M,C)
Q_MA = lambda m, h: 3 + 0.9*(0.5*m + 0.5*h)     # Q(M,A)
```

```python
Q_HC = lambda h, m: 5 + 0.9*(0.9*h + 0.1*m)     # Q(H,C)
Q_HA = lambda h, m: 5 + 0.9*(0.7*h + 0.3*m)     # Q(H,A)

Q_LC_1 = Q_LC(v1_L, v1_M)
Q_LA_1 = Q_LA(v1_L, v1_M)
Q_MC_1 = Q_MC(v1_M, v1_H)
Q_MA_1 = Q_MA(v1_M, v1_H)
Q_HC_1 = Q_HC(v1_H, v1_M)
Q_HA_1 = Q_HA(v1_H, v1_M)

print(f"Q(L,C) = {Q_LC_1:.2f}\tQ(L,A) = {Q_LA_1:.2f}")
print(f"Q(M,C) = {Q_MC_1:.2f}\tQ(M,A) = {Q_MA_1:.2f}")
print(f"Q(H,C) = {Q_HC_1:.2f}\tQ(H,A) = {Q_HA_1:.2f}")
```

```
Q(L,C) = -1.18  Q(L,A) = -0.46
Q(M,C) = 6.24   Q(M,A) = 6.60
Q(H,C) = 9.32   Q(H,A) = 8.96
```

Policy at Iteration 1:

- L → Aggressive (A)
- M → Aggressive (A)
- H → Conservative (C)

Iteration 2

Updating $V_2(s)$:

For **Low Wealth (L):**

> $$V_2(L) = \max\left[-1 + 0.9(0.8(-1) + 0.2(3)), -1 + 0.9(0.6(-1) + 0.4(3))\right] = -0.46$$

For **Medium Wealth (M):**

> $$V_2(M) = \max\left[3 + 0.9(0.7(3) + 0.3(5)), 3 + 0.9(0.5(3) + 0.5(5))\right] = 6.60$$

For **High Wealth (H):**

> $$V_2(H) = \max\left[5 + 0.9(0.9(5) + 0.1(3)), 5 + 0.9(0.7(5) + 0.3(3))\right] = 9.32$$

In [38]:
```python
v2_L = vL_func(v1_L, v1_M)
v2_M = vM_func(v1_M, v1_H)
v2_H = vH_func(v1_H, v1_M)
print(f"V₂(L) = {v2_L:.2f},  V₂(M) = {v2_M:.2f},  V₂(H) = {v2_H:.2f}")
```

```
V₂(L) = -0.46,  V₂(M) = 6.60,  V₂(H) = 9.32
```

Policy Evaluation after Iteration 2

For **Low Wealth (L):**

> $$Q(L, C) = -1 + 0.9(0.8(-0.46) + 0.2(6.6)) = -0.14$$

> $$Q(L, A) = -1 + 0.9(0.6(-0.46) + 0.4(6.6)) = 1.13$$

For **Medium Wealth (M):**

> $$Q(M, C) = 3 + 0.9(0.7(6.6) + 0.3(9.32)) = 9.67$$

> $$Q(M, A) = 3 + 0.9(0.5(6.6) + 0.5(9.32)) = 10.16$$

For **High Wealth (H):**

> $$Q(H, C) = 5 + 0.9(0.9(9.32) + 0.1(6.6)) = 13.14$$

$

$$Q(H,A) = 5 + 0.9(0.7(9.32) + 0.3(6.6)) = 12.65$$

In [39]:
```
Q_LC_2 = Q_LC(v2_L, v2_M)
Q_LA_2 = Q_LA(v2_L, v2_M)
Q_MC_2 = Q_MC(v2_M, v2_H)
Q_MA_2 = Q_MA(v2_M, v2_H)
Q_HC_2 = Q_HC(v2_H, v2_M)
Q_HA_2 = Q_HA(v2_H, v2_M)

print(f"Q(L,C) = {Q_LC_2:.2f}\tQ(L,A) = {Q_LA_2:.2f}")
print(f"Q(M,C) = {Q_MC_2:.2f}\tQ(M,A) = {Q_MA_2:.2f}")
print(f"Q(H,C) = {Q_HC_2:.2f}\tQ(H,A) = {Q_HA_2:.2f}")
```

```
Q(L,C) = -0.14  Q(L,A) = 1.13
Q(M,C) = 9.67   Q(M,A) = 10.16
Q(H,C) = 13.14  Q(H,A) = 12.65
```

Policy at Iteration 2:

- L → Aggressive (A)
- M → Aggressive (A)
- H → Conservative (C)

Iteration 3

Updating $V_3(s)$:

For **Low Wealth (L):**

$$V_3(L) = \max\left[-1 + 0.9(0.8(-0.46) + 0.2(6.6)), -1 + 0.9(0.6(-0.46) + 0.4(6.6))\right] = 1.13$$

For **Medium Wealth (M):**

$$V_3(M) = \max\left[3 + 0.9(0.7(6.6) + 0.3(9.32)), 3 + 0.9(0.5(6.6) + 0.5(9.32))\right] = 10.16$$

For **High Wealth (H):**

$$V_3(H) = \max\left[5 + 0.9(0.9(9.32) + 0.1(6.6)), 5 + 0.9(0.7(9.32) + 0.3(6.6))\right] = 13.14$$

In [40]:
```
v3_L = vL_func(v2_L, v2_M)
v3_M = vM_func(v2_M, v2_H)
v3_H = vH_func(v2_H, v2_M)
print(f"V₃(L) = {v3_L:.2f},  V₃(M) = {v3_M:.2f},  V₃(H) = {v3_H:.2f}")
```

```
V₃(L) = 1.13,  V₃(M) = 10.16,  V₃(H) = 13.14
```

Policy Change Analysis

From **Iteration 2 to Iteration 3**, let's check the action values to determine if the policy changed.

For **Low Wealth (L):**

$$Q(L,C) = -1 + 0.9(0.8(1.13) + 0.2(10.16)) = 1.64$$

$$Q(L,A) = -1 + 0.9(0.6(1.13) + 0.4(10.16)) = 3.27$$

For **Medium Wealth (M):**

$$Q(M,C) = 3 + 0.9(0.7(10.16) + 0.3(13.14)) = 12.95$$

$$Q(M,A) = 3 + 0.9(0.7(10.16) + 0.5(13.14)) = 13.49$$

For **High Wealth (H):**

> $$Q(H,C) = 5 + 0.9(0.9(13.14) + 0.1(10.16)) = 16.56$$
>
> $

> $$Q(H,A) = 5 + 0.9(0.7(13.14) + 0.3(10.16)) = 16.02$$

In [41]:
```
Q_LC_3 = Q_LC(v3_L, v3_M)
Q_LA_3 = Q_LA(v3_L, v3_M)
Q_MC_3 = Q_MC(v3_M, v3_H)
Q_MA_3 = Q_MA(v3_M, v3_H)
Q_HC_3 = Q_HC(v3_H, v3_M)
Q_HA_3 = Q_HA(v3_H, v3_M)

print(f"Q(L,C) = {Q_LC_3:.2f}\tQ(L,A) = {Q_LA_3:.2f}")
print(f"Q(M,C) = {Q_MC_3:.2f}\tQ(M,A) = {Q_MA_3:.2f}")
print(f"Q(H,C) = {Q_HC_3:.2f}\tQ(H,A) = {Q_HA_3:.2f}")
```

```
Q(L,C) = 1.64   Q(L,A) = 3.27
Q(M,C) = 12.95  Q(M,A) = 13.49
Q(H,C) = 16.56  Q(H,A) = 16.02
```

Compare $Q(L,A), Q(L,C)$ and $Q(H,C), Q(H,A)$, decide the policy updates:

- **Low Wealth (L)** → Aggressive (A)
- **Medium Wealth (M)** → Aggressive (A)
- **High Wealth (H)** → Conservative (C)

Summary: Policy Evolution Over Iterations

State	Iteration 1	Iteration 2	Iteration 3
Low	Aggressive	Aggressive	Aggressive
Medium	Aggressive	Aggressive	Aggressive
High	Conservative	Conservative	Conservative