QUESTION 4 CAI ZESHUO 213733

(a)

1. Robot colour / sticker design – doesn't affect fruit collection, battery, obstacles, or safety, so it won't change future reward.
2. Time/date (e.g., "Monday", "12:30pm") – by itself it doesn't help the robot decide where to move or avoid mud/obstacles, unless the task explicitly depends on time (which is not stated). Keeping it out makes the state simpler.

(b)

Reward fruit collected, punish battery waste, time, and accidents:

$$R = +a(\Delta W) - b(\Delta B) - c(\Delta t) - d(\text{collision/stall})$$

(c)

- Fruit weight in bin: give positive reward when weight increases ($+\Delta W$).
- Battery consumption: give negative reward so it saves energy ($-\Delta B$).
- Collision/stall: give big penalty because unsafe and delays work.
- Time to finish: small penalty per step so it completes faster ($-\Delta t$).

(d)

Episodic, because one "run" starts when the robot begins collecting and ends when it reaches a target amount, bin full, battery low, or returns to drop-off. Then it resets and starts a new run.

(e)

In the plantation, fruit locations and mud/obstacles change, so the robot must explore sometimes to discover better routes/actions and update $Q(s, a)$. But it must also exploit often to use the best-known action to collect fruits efficiently. With ε-greedy, it usually chooses the

best Qaction, but with probability εit tries other actions so it doesn't get stuck with a bad strategy.