

Game Plan:

Each player has:

- a location attribute in the form of a coordinate (x,y)
- whether it has the ball in the form of a boolean
- every second, each player decides how s/he should act

Player actions include

- intercept
- run
 - expressed as a vector, with the magnitude being the speed and the direction being the direction
- pass
- shoot

The ball has:

- a location attribute in the form of a coordinate (x,y)
- the speed and direction it's going in the form of a vector

Training Plan:

- going near the ball = reward
- intercepting the ball = bigger reward
- having the ball taken from you = punishment

Have a transition function that takes an action, a state (everything happening at the moment; a snapshot in time), a specific player this is intended for, and returns a new action for this specific player

We plan on doing policy iteration to train the agents. Not sure how though.

We should have a while loop that keeps calling the transition function, which requires a reward function.

$$Q_i(s, a) \leftarrow \sum_{s'} P(s'|s, a) [R(s, a, s') + \gamma \max_{a'} Q_{i-1}(s', a')]$$

Each player should have his own policy. Then there's a separate reward function.

Important: 1v1 should be our first milestone, maybe even only one player only first!!

Right now, we're considering only have two players – one player on each team. When there's only one player on each team, the only actions are run and shoot. And the probability of reaching a new state is always 100% for now.

$$p_i \begin{cases} \text{has ball} \begin{cases} \text{run} \\ \text{shoot} \end{cases} \\ \text{doesn't have ball} \begin{cases} \text{run} \\ \text{intercept} \end{cases} \end{cases}$$

Low reward for shooting from afar, even if it gets in. High reward for getting it in a short distance.

```
R(s, a, s'){
  match a with
  | ball_run -> return whatever
  | ball_shoot -> return whatever2
  | no_ball_run -> return whatever
  ...
}
```

$ball_shoot \begin{cases} \text{near} = \text{high reward} \\ \text{far} = \text{low reward} \end{cases}$

$ball_run \begin{cases} \text{far} = \text{high reward} \\ \text{near} = \text{low reward} \end{cases}$

$no_ball_run \begin{cases} \text{far from ball} = \text{high reward} \\ \text{near ball} = \text{low reward} \end{cases}$

$intercept \begin{cases} \text{successful} = \text{high reward} \\ \text{unsuccessful} = 0 \text{ reward} \end{cases}$