Actor-Critic Methods

**Actor-Critic Methods**: A hybrid architecture combining value-based and Policy-Based methods that helps to stabilize the training by reducing the variance using:

1. An Actor that controls how our agent behaves. (Policy-Based method)
2. A Critic that measures how good the taken action is. (Value-Based method)

**The Actor-Critic Process**:

1. At each timestep, t, we get the current state from the environment and pass it as input through our Actor and Critic.
2. Our Policy takes the state and outputs an action.
3. The Critic takes that action also as input and computes the value of taking that action at that state: the Q-value.
4. The action performed in the environment outputs a new state and a reward.
5. The Actor updates its policy parameters using the Q value.
6. Thanks to its updated parameters, the Actor produces the next action to take at given the new state.
7. The Critic then updates its value parameters.

**Advantage Actor-Critic (A2C)**: The Advantage function calculates the relative advantage of an action compared to the others possible at a state: how taking that action at a state is better compared to the average value of the state.

In other words, this function calculates the extra reward we get if we take this action at that state compared to the mean reward we get at that state.

The extra reward is what’s beyond the expected value of that state.

* If A(s,a) > 0: our gradient is pushed in that direction.
* If A(s,a) < 0: (our action does worse than the average value of that state), our gradient is pushed in the opposite direction.