REINFORCEMENT LEARNING
ASSIGNMENT 2

Name: Vaibhav Poddar
Roll No: 16CS10051

The algorithm used is REINFORCE

function REINFORCE
Initialise θ arbitrarily
for each episode $\{s_1, a_1, r_2, ..., s_{T-1}, a_{T-1}, r_T\} \sim \pi_\theta$ **do**
for $t = 1$ to $T - 1$ **do**
$\theta \leftarrow \theta + \alpha \nabla_\theta \log \pi_\theta(s_t, a_t) v_t$
end for
end for
return θ
end function

Here we learn the parameters of the artificial neural network such that given a vector representation of any state we can predict the action probability distribution for the corresponding state.

Network Architecture:

37 (input) ⇒
256 (hidden1) ⇒ Relu ⇒
256 (hidden2) ⇒ Relu ⇒
4 ⇒ softmax (output)

- It contains 2 hidden layers of Neuron with Relu activation Function
- First hidden Layer contains 256 neurons, each neuron in this layer has 37 inputs.
- Second hidden Layer contains 256 neurons, each neuron in this layer has 256 inputs.
- Third layer contains 4 neurons, each neuron in this layer has 256 inputs.
- Output of the third layer is given to Sofmax function so that we can get values in between [0,1]

Results: