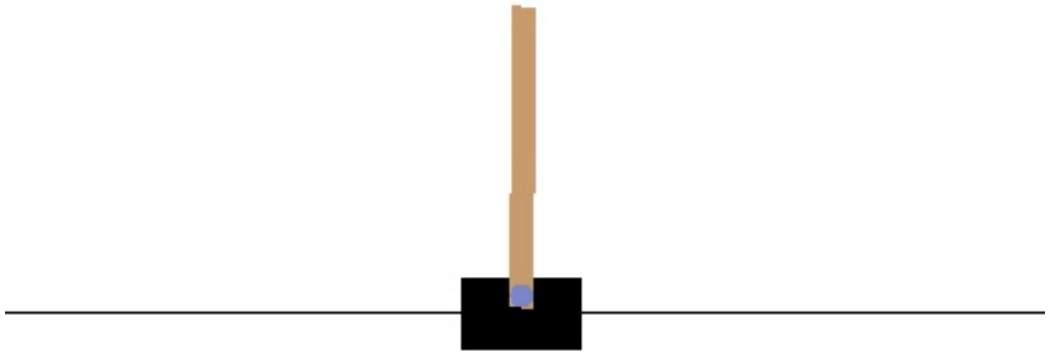
Reinforcement Learning

First environment: Cartpole



First environment: Cartpole

Observation

Type: Box(4)

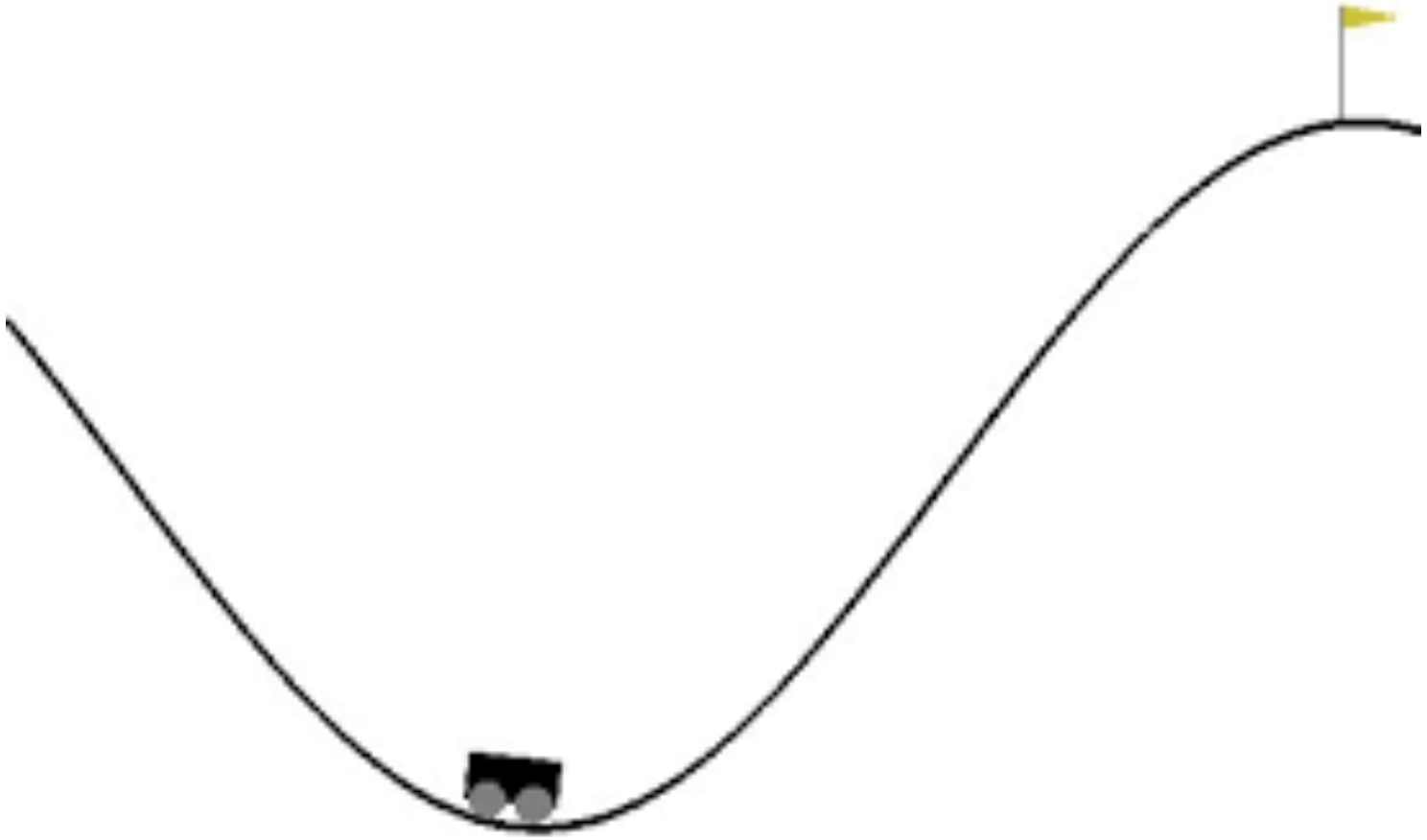
Num	Observation	Min	Max
0	Cart Position	-2.4	2.4
1	Cart Velocity	-Inf	Inf
2	Pole Angle	~ -41.8°	~ 41.8°
3	Pole Velocity At Tip	-Inf	Inf

Actions

Type: Discrete(2)

Num	Action
0	Push cart to the left
1	Push cart to the right

Second environment: MountainCar



Second environment: MountainCar

Observation

Type: Box(2)

Num	Observation	Min	Max
0	position	-1.2	0.6
1	velocity	-0.07	0.07

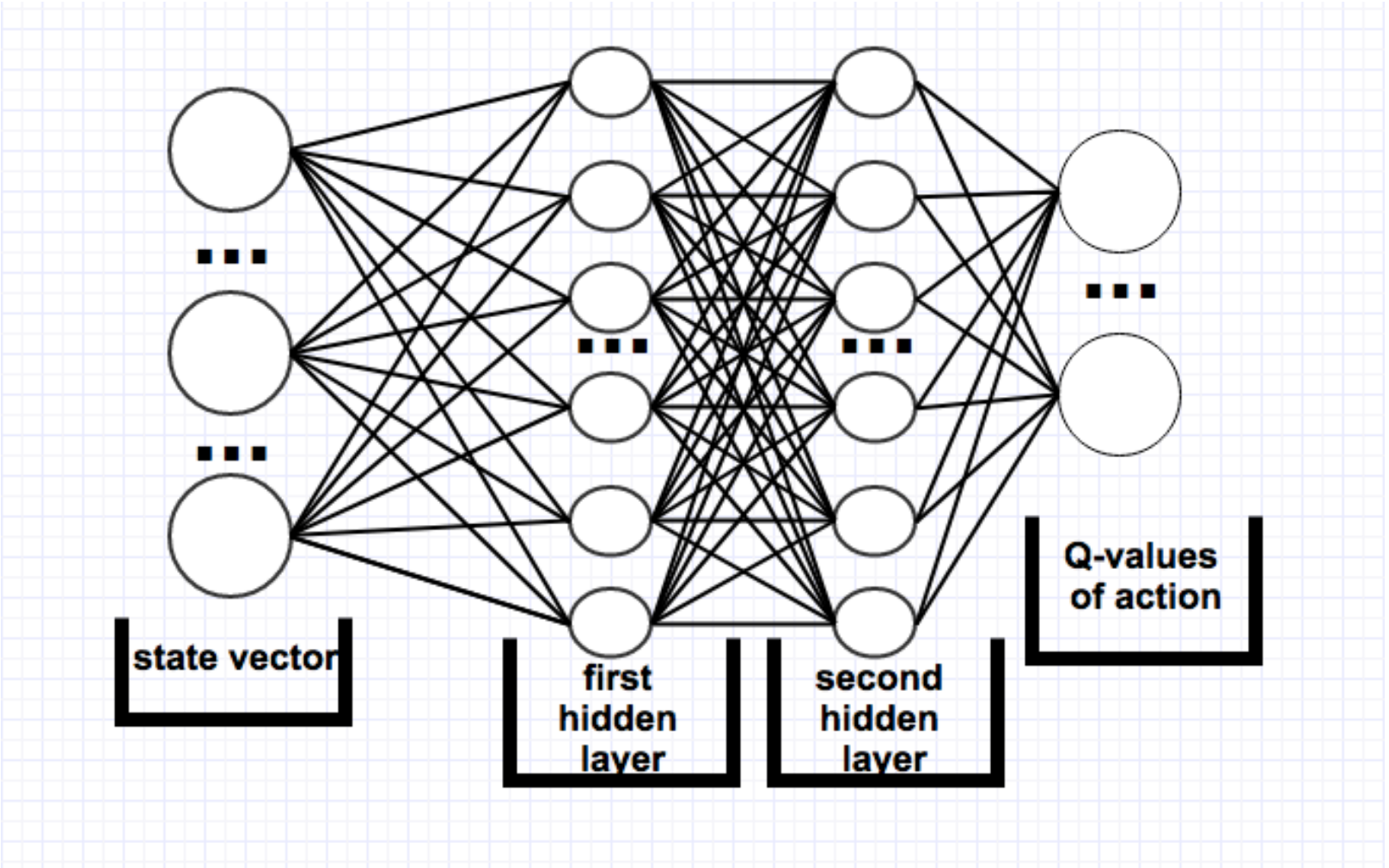
Actions

Type: Discrete(3)

Num	Observation
0	push left
1	no push
2	push right

First model

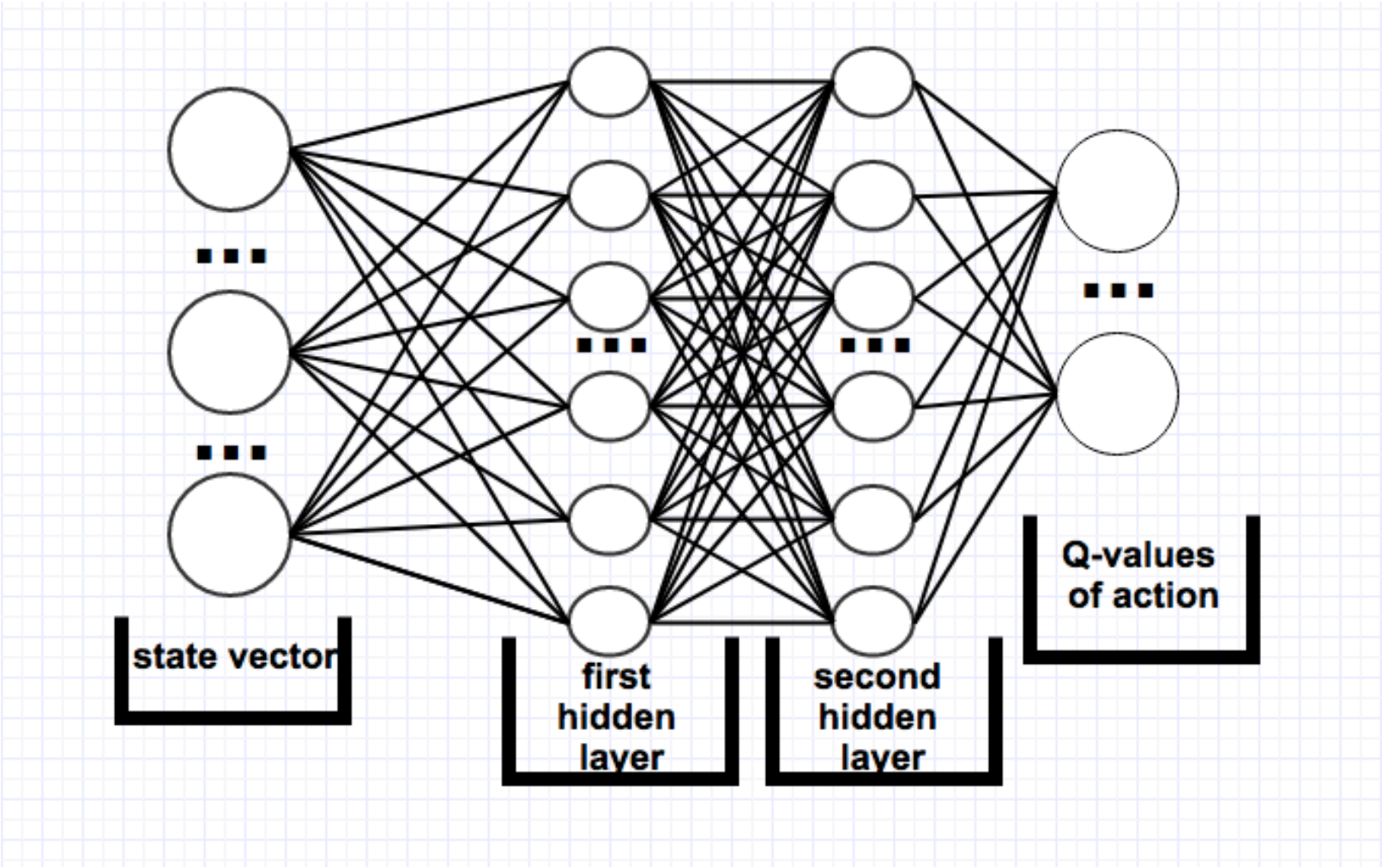
First hidden layer – 24 neurons (relu activation)
Second hidden layer – 24 neurons(relu activation)



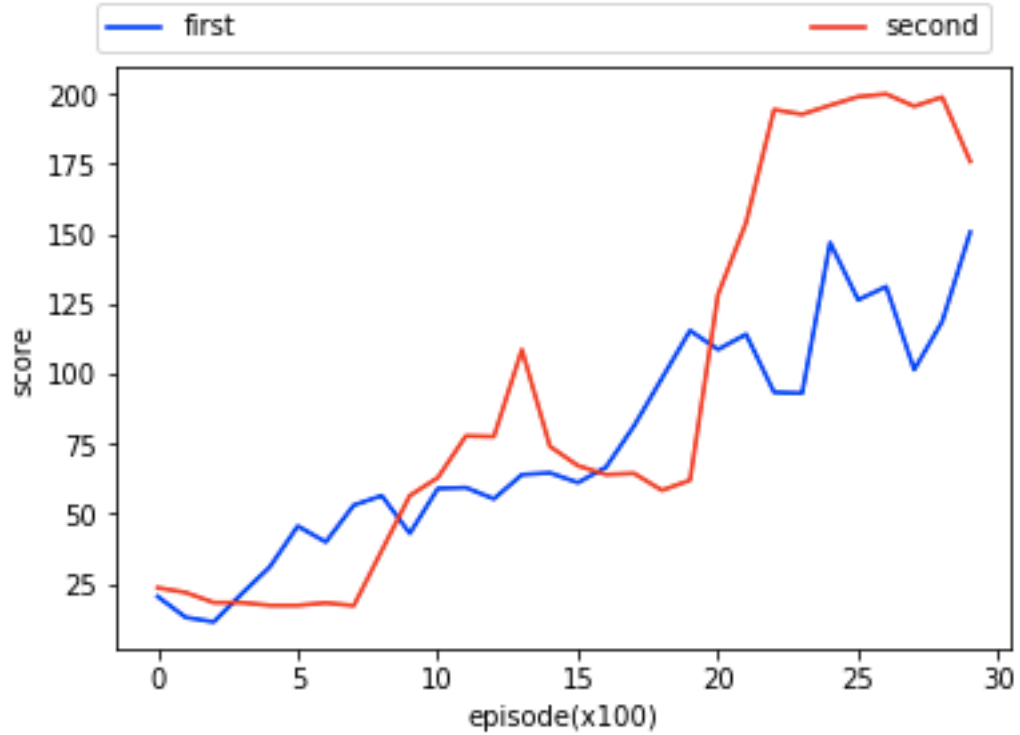
Second Model

Two same models to increase training stability: target model and basic model
Fisrt we train on basic model and every n iterations update target model

First hidden layer – 32 neurons (relu activation)
Second hidden layer – 16 neurons(relu activation)



Results (cartpole)



Results (mountain car)