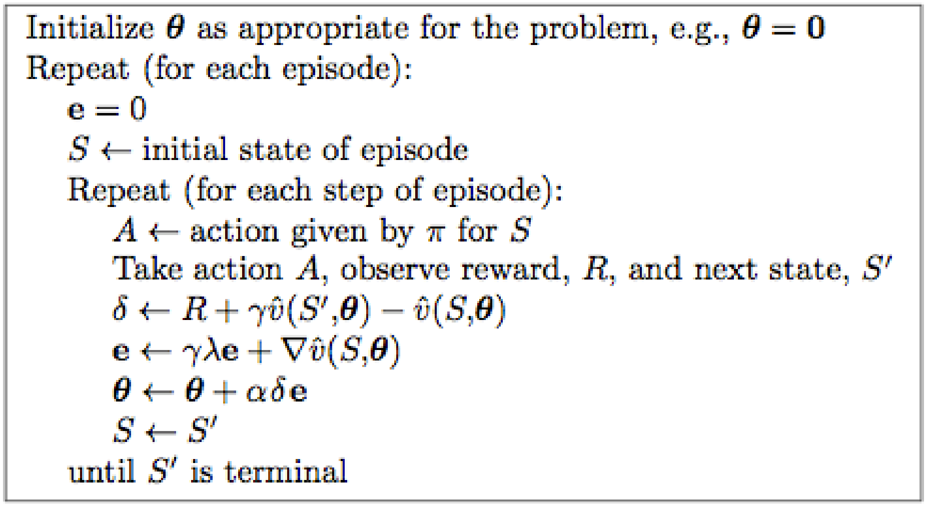
Foundations of machine Learning II

Project: Mountain Car

Flavie Vampouille
MSc in DS&BA
ESSEC Business School & CentraleSupélec

Mountain car learning using Linear Gradient Descent

I use the algorithm given in the Barton and Sutton book in chapter 9:



```
Initialize $\boldsymbol{\theta}$ as appropriate for the problem, e.g., $\boldsymbol{\theta} = \mathbf{0}$
Repeat (for each episode):
    $\mathbf{e} = 0$
    $S \leftarrow$ initial state of episode
    Repeat (for each step of episode):
        $A \leftarrow$ action given by $\pi$ for $S$
        Take action $A$, observe reward, $R$, and next state, $S'$
        $\delta \leftarrow R + \gamma\hat{v}(S',\boldsymbol{\theta}) - \hat{v}(S,\boldsymbol{\theta})$
        $\mathbf{e} \leftarrow \gamma\lambda\mathbf{e} + \nabla\hat{v}(S,\boldsymbol{\theta})$
        $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha\delta\mathbf{e}$
        $S \leftarrow S'$
    until $S'$ is terminal
```

With

- π : ε greedy method
- θ = W
- $v(S,\theta) = Q((x, v_x), a) = \sum_{i,j} W_{i,j}^{a} \, \phi_{i,j}$
- $\nabla v = \nabla_W Q = \phi$

I implement it in four ways:

- SARSA(0)
- Q(0)
- SARSA(λ)
- Q(λ)

I first run the SARSA(0) and Q(0) algorithms in which I set the number of episodes and step to test.learn(200, 5000). Neither of these agents converge with only 200 episodes but due to big computation time I don't run them with more episodes.

Instead I implement the SARSA(λ) and Q(λ) agents and I first run them with a number of episodes and step set to test.learn(200, 5000). Once again it doesn't converge. However this time I try to run them on more episodes and set test.learn(200, 5000) but it doesn't converge either, the reward being between -20 and -150.

SARSA(0)

In the staterSarsa(0).py code I use the above code without the λ part.

Watkins' Q(0)

The starterQ(0).py code is basically the same that SARSA(0) but taking the max of the future state for updating δ parameter.

SARSA(λ)

The starterSARSA(lambda).py code is the implementation of the Barton and Sutton algorithm above.

Q(λ)

The starterSARSA(lambda).py code is the implementation of the Barton and Sutton algorithm above taking max(Q) updating δ parameter.