

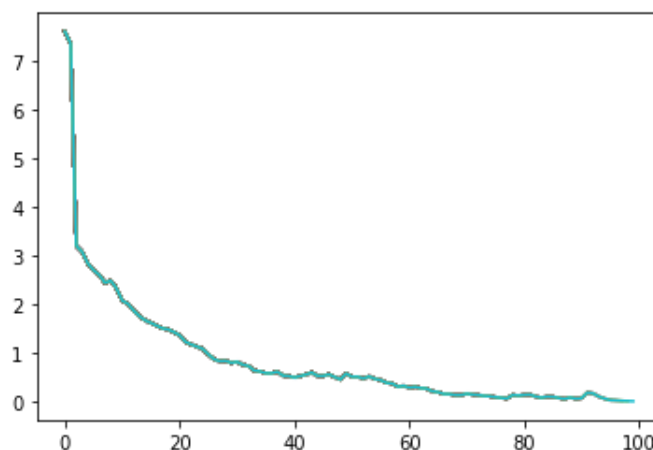
Mašinsko učenje – Domaći 5

Rešavanje poznatog MPO planiranjem

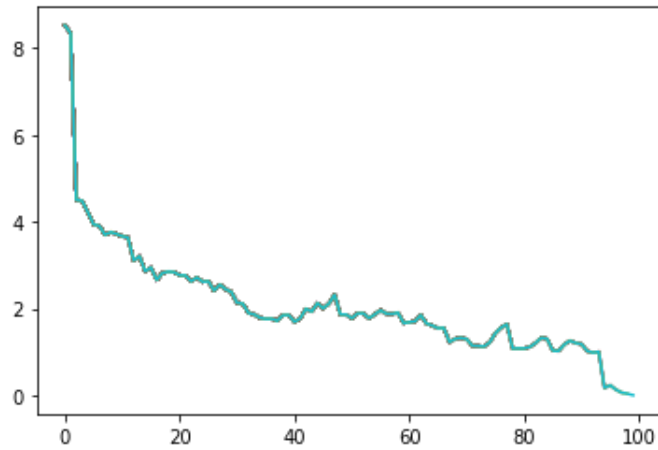
1. Sa povećanjem parametra gama se ubrzava konvergencija
2. V vrednosti postaju negativne i veoma male. Vrednosti koje vode na desno, a ne na gore vrlo brzo dobijaju velike negativne vrednosti, odnosno obezbeđuje se da se tom putanjom nikako ne ide

Rešavanje nepoznatog MPO problema kroz učenje podsticanjem

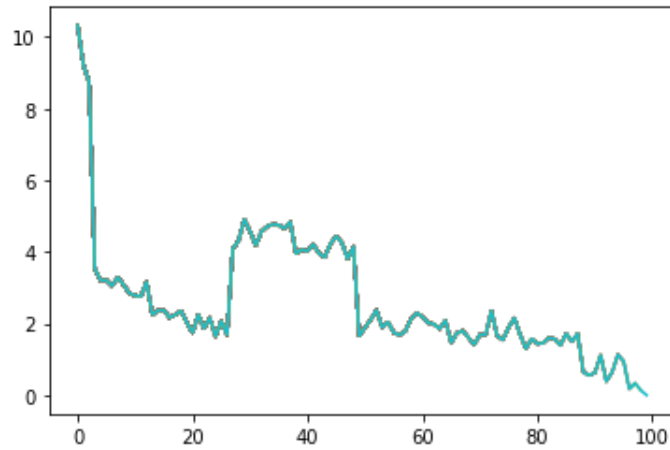
1. Funkciji se može zadati bool parametar da li se koristi ϵ - gramziva pretraga ili ne. Ideja je da se početno epsilon koje sam uzela da iznosi 0.5 dodatno smanjuje kako se prolazi kroz korake algoritma.
2. a) Ekperimentisanjem sa razlicitim stopama i poredjenjem *plot*-a njihove greske u odnosu na najbolje moguće Q vrednosti tokom iteracija, donet je zakljucak da je najbolja stopa 0.2



Alfa = 0.2

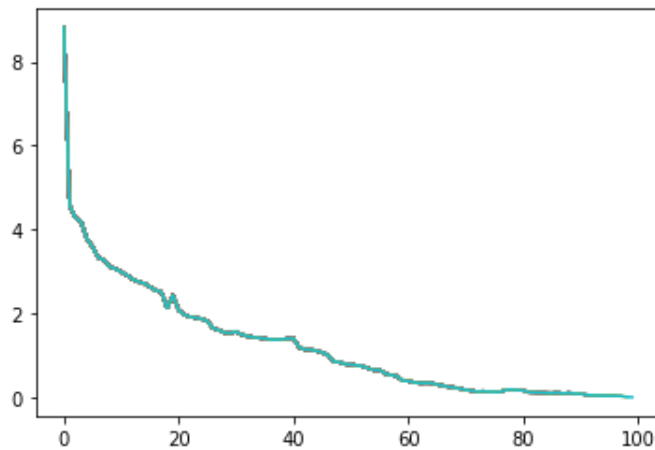


Alfa = 0.5

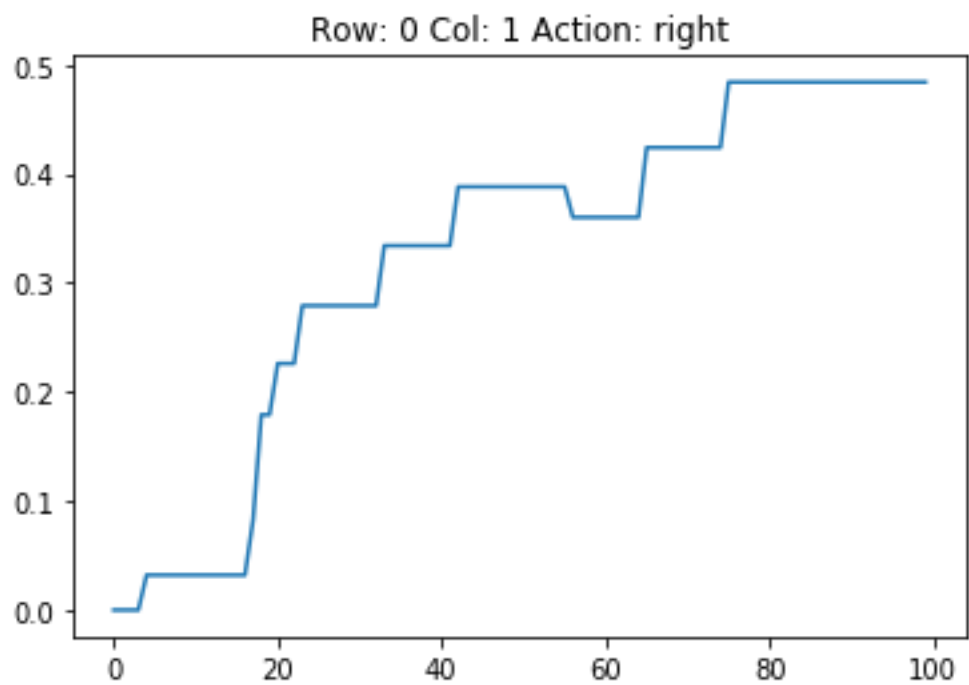
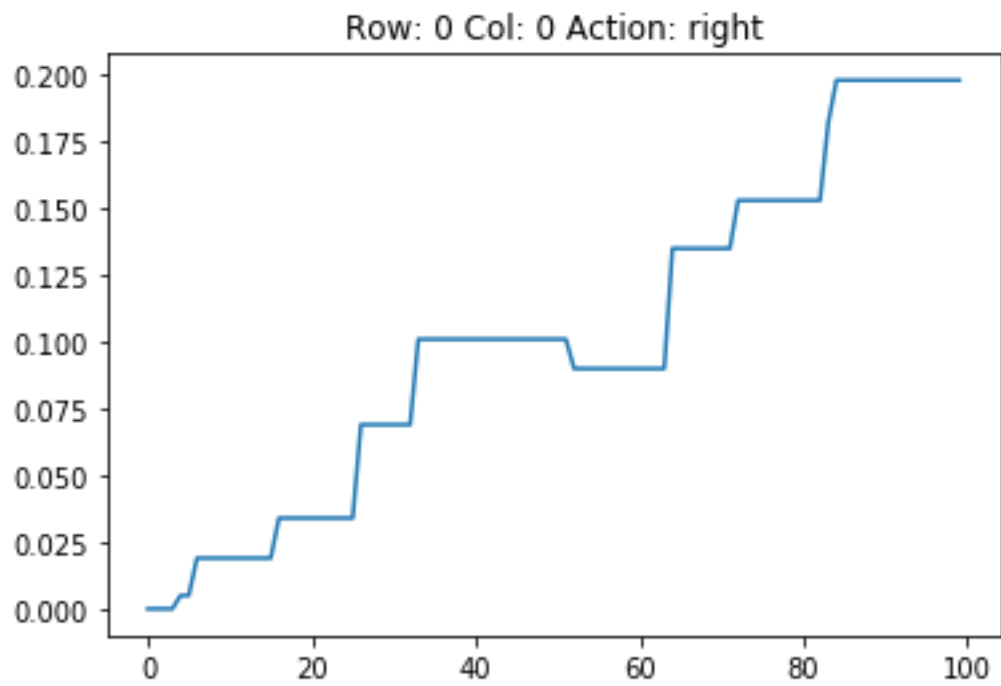


Alfa = 0.8

b) Adaptivna stopa – implementirana je adaptivna stopa predložena u tekstu domaćeg i dobijen je sledeći rezultat:



V – vrednosti za stanja u prvoj vrsti i njihove najbolje moguće vrednosti:



Row: 0 Col: 2 Action: right

