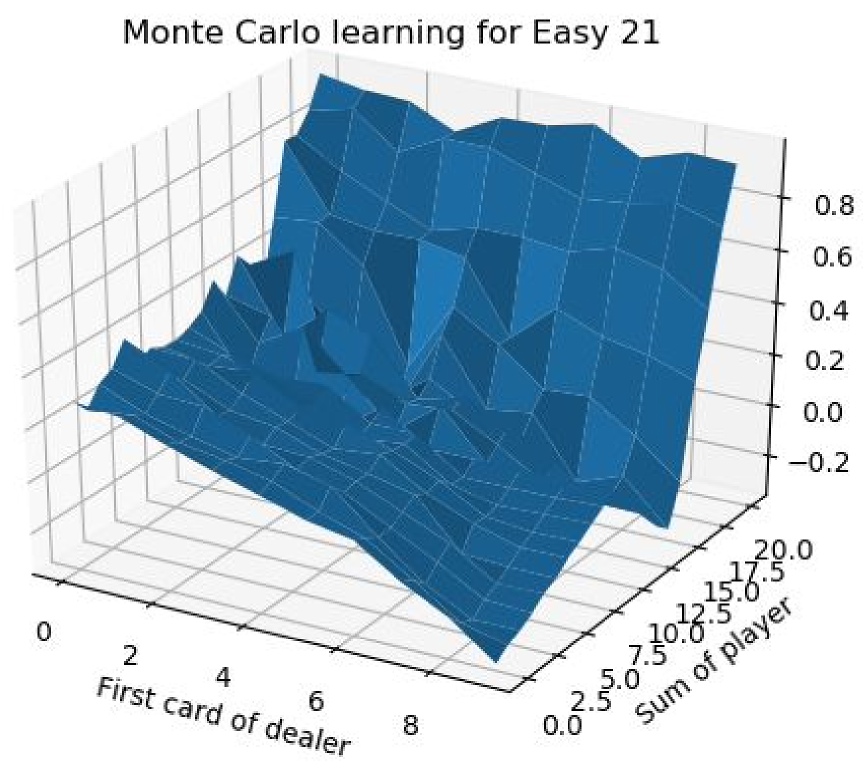
Marín McGinley
mim17@hi.is

Reiknigreind

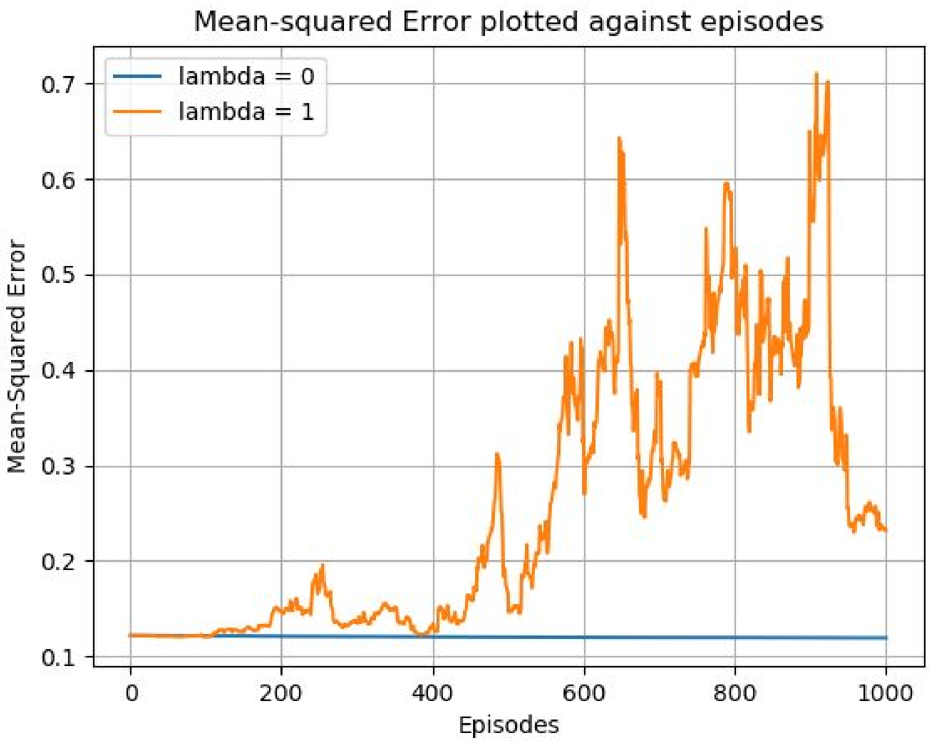
Reikniverkefni 1

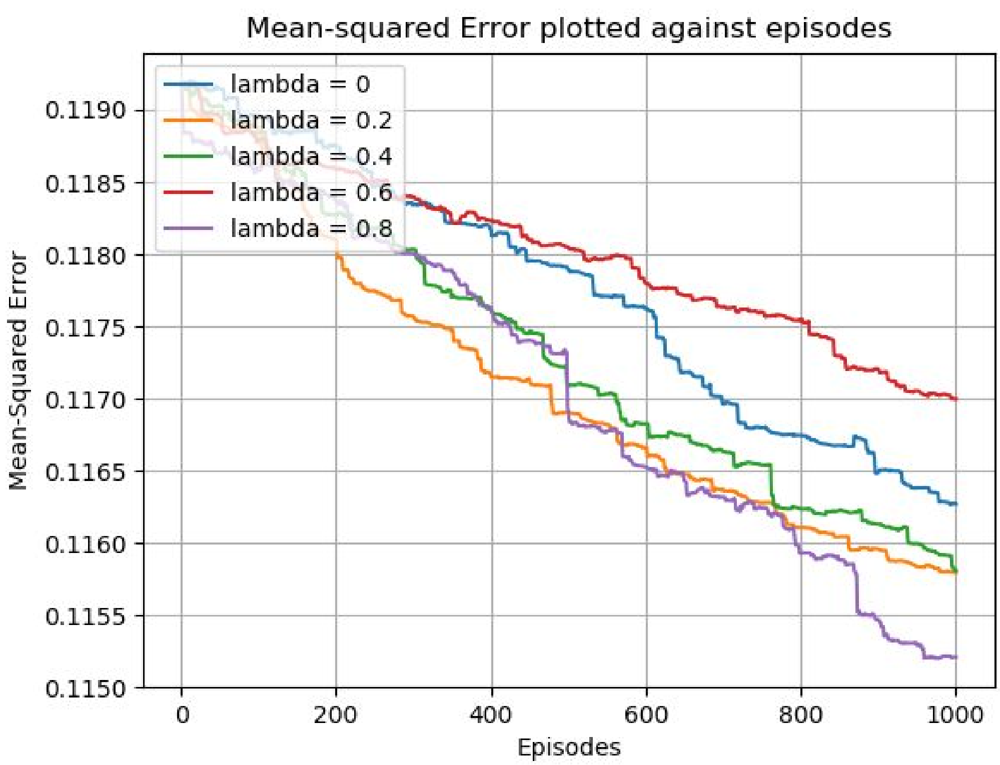
Here is the optimal value function $V^*(s) = max_a Q^*(s, a)$ that I plotted from learning the game Easy 21 with Monte-Carlo Control:



I did have some problems with this project that I was unable to solve or explain. For calculating a policy function Q for TD Learning I found that while $\lambda < 1$, Q did behave as expected but for $\lambda = 1$, Q showed some behaviour that I have no explanation for. Because the plotting of the learning curve of the mean-squared error against episode number for $\lambda = 1$ is so widely erratic and different from the learning curve of $\lambda = 0$ I decided to include two pictures, one of the two learning curves together and one of the learning curves for $\lambda = 0,\ 0.2,\ 0.4,\ 0.6,\ 0.8$. This way the latter is more obvious and easier to interpret for $\lambda = 0$. It also shows that this unexplained behaviour only occurs when $\lambda > 0.9$.

Marín McGinley
mim17@hi.is





Marín McGinley
mim17@hi.is

For the plot of the Mean-Squared Error plotted against lambda I again had the same problem as before. Because the MSE for $\lambda = 1$ was of such a different scale it didn't allow for the finer details of the MSE for $\lambda < 1$ to shine through. Therefore I include here two pictures. First of the Mean-Squared Error plotted against lambda for $0 \leq \lambda \leq 1$, and the second against lambda for $0 \leq \lambda < 1$.