

# 1 \*Off-policy Methods with Approximation

## 1.1 Exercise 11.

### 1.1.1 Q

### 1.1.2 A

## 1.2 Exercise 11.3 (programming)

### 1.2.1 Q

Apply one-step semi-gradient Q-learning to Baird's counterexample and show empirically that its weights diverge.

### 1.2.2 A

This is a programming exercise. For the relevant code please see [the repo](#).

