universität freiburg

Machine Learning for Stochastics

Summer semester 2025

Lecturer: Prof. Dr. Thorsten Schmidt
Assistance: Simone Pavarana, MSc.

Tutorial 1

Exercise 1 (3 Points)**.**
Consider a stationary Markov Decision Model with planning horizon N satisfying assumptions (A_N) and (SA_N). Let $g,\hat{g} \in \mathbb{M}(E)$ be two terminal measurable reward functions, and define

$$J_n := \mathcal{T}^n g, \qquad \hat{J}_n := \mathcal{T}^n \hat{g},$$

where the Bellman operator $\mathcal{T}$ is given by

$$(\mathcal{T}v)(x) := \max_{a\in A}\left\{ r(x,a) + \beta \int_E v(y)\, Q(dy \mid x,a)\right\}.$$

Show the following:

(a) The Bellman operator $\mathcal{T}$ is a contraction on $(\mathbb{M}(E), \|\cdot\|_\infty)$ with modulus β.

(b) For all $k = 1,\dots,N$, it holds that

$$J_N(x) - \hat{J}_k(x) \le \beta^N \sup_{x\in E}(g(x) - \hat{g}(x)) + \sup_{x\in E}(\hat{J}_k(x) - \hat{J}_{k-1}(x)) \sum_{j=1}^{N-k} \beta^j.$$

(c) For all $k = 1,\dots,N$, it holds that

$$J_N(x) - \hat{J}_k(x) \ge \beta^N \inf_{x\in E}(g(x) - \hat{g}(x)) + \inf_{x\in E}(\hat{J}_k(x) - \hat{J}_{k-1}(x)) \sum_{j=1}^{N-k} \beta^j.$$

Hint: For point (a), you may use without proof the inequality

$$|\max_x f(x) - \max_x g(x)| \le \max_x |f(x) - g(x)|,$$

from which it follows that the max operator is Lipschitz continuous with constant 1.

Exercise 2 (3 Points)**.**
Consider the following variant of the red-and-black card game. A standard deck of 52 cards is shuffled and turned over one card at a time. Among these cards, there is exactly one ace of spades. The player observes the cards as they are revealed and must say “stop” at some point, indicating that they believe the *next* card (i.e., the one that has not yet been uncovered) is the ace of spades.
What strategy maximizes the probability of correctly identifying the position of the ace of spades? Justify your answer by formulating the problem as a Markov decision process.

Exercise 3 (3+1 Bonus Point).

Imagine you enter a casino and are allowed to play N times the same game. The probability of winning a single game is $p \in (0,1)$, and the outcomes of the games are independent. You begin with an initial wealth $x > 0$ and are allowed to stake any amount in the interval $[0,x]$. If you win a game, you receive twice your stake; if you lose, the stake is forfeited. The objective is to maximize the expected wealth $\mathbb{E}_x^\pi[X_N]$ after N rounds.

(a) Formulate this problem as a Markov Decision Model.

(b) **(Bonus)** Find an upper bounding function and prove that condition (SA_N) can be satisfied.

(c) Determine an optimal strategy in the three cases: $p < \frac{1}{2}$, $p = \frac{1}{2}$, and $p > \frac{1}{2}$.

(d) How does the problem change if the objective becomes maximizing $\mathbb{E}_x^\pi[U(X_N)]$ for a strictly increasing and strictly concave utility function $U\colon \mathbb{R}_+ \to \mathbb{R}$?