Machine, Data and Learning (CS7.301)

Spring 2022, IIIT Hyderabad
04 Apr, Monday (Lecture 16)

Taught by Prof. Vikram Pudi

Utility Theory

Decision theory is comprised mainly of probability theory (which deals with chance) and utility theory (which deals with outcomes). The fundamental idea is that we weight each outcome's utility by its probability.
An agent is called *rational* iff it maximises expected utility by its actions.

Risk

The slope of the utility function tends to be continuously decreasing (diminishing returns). Thus we may somtimes refuse to play what appears to be a monetarily fair bet. For example, if the choice was between a 50-50 chance to lose or gain \$1000, and not losing or gaining anything, we would pick the second option. Theoretically,

$$\begin{aligned} U(x+c) - U(x) &< U(x) - U(x-c) \\ U(x+c) + U(x-c) &< U(x) + U(x) \\ \frac{U(x+c) + U(x-c)}{2} &< U(x) \end{aligned} .$$

This is a *risk-averse* philosophy.

Correspondiongly, one might be *risk-neutral* (in which case the utility function is a straight line) or *risk-seeking* (in which case the utility function's slope continuously increases).

Note that sometimes the utility function depends on more than one feature of the outcomes (*e.g.*, the decision of whether or not to buy a house would depend on its location, cost, age, etc.).

Markov Decision Processes

A Markov decision process is defined as a tuple $\langle S, A, P, R \rangle$, where S represents a set of states, A represents a set of actions, $P : S \times A \times S \to [0, 1]$ a transition

function, and $R : S \times A \to \mathbb{R}$ is a reward function. We need to choose a sequence of actions to maximise reward.

The solution to this problem is to determine a *policy*, rather than a plan (as there is uncertainty at every step). A policy is a complete mapping from states to actions; by the Markov assumption, the action to be taken is independent of the history of actions leading up the current state.

An MDP might be a finite-horizon (in which the process ends after a fixed time) or an infinite-horizon MDP.