Section overview

Reinforcement Learning 101

Ground work needed for Reinforcement Learning

We understand: Control is sequential decision making. Optimal control is sequential decision making so as to minimise a cost or maximise a reward ("$reward = -cost$").
Reinforcement Learning involves learning an optimal control of an "a priori" unknown system, unknown environment with unknown rewards – only experience will teach us.

Probability theory refresher

Example (London weather)

$A \in Rain, Sun$ and $B \in Windy, Calm$ $p(A\ B)$ is shorthand for probability that events A and B occur simultaneously.

$$P = \begin{pmatrix} 0.1 & 0.4 & Rain \\ 0.25 & 0.25 & Sun \\ Windy & Calm & \end{pmatrix} \tag{1}$$

All possible combination of event probabilities must sum to 1: $\sum_A \sum_B P(A\ B) = 1$.

1. What is the probability of rain $P('Rain')$? (marginalisation)
2. What is the probability of raining if it is a windy day $P('Rain'|'Windy')$?(conditioning)

Basic probability theory

Definition (Bayes Theorem)

$p(AB) = p(A|B) \times p(B) = p(B|A) \times p(A)$

Example (Clinical screening)

$p(AB)$ is the joint probability of blood test outcomes and a person's disease state.
$P(A) = P(\text{Person has disease}) = 1\%$ of population
$P(B) = P(\text{Person has positive blood test}) = 10\%$ of population
$P(B|A) = P(\text{blood test is positive given that person is ill}) = 70\%$
$P(A|B) = P(\text{person is ill given that blood test was positive}) = ?$

Probabilistic Reasoning as Extension of Basic Logic: Boole vs Bayes

- Bayesian inference is a simple mathematical theory which characterises plausible reasoning in the presence of uncertainties. Think about learning as the reduction of uncertainty about what we want to know.
- However, classic Boolean logic excludes certain forms of "plausible reasoning" (real-world):
  We observe that A is false. We find B becomes less plausible... although no conclusion can be drawn from classical logic. We observe that B is true. It seems A becomes more plausible.
- We use this form of reasoning daily: Our friend is late. Was she H1 abducted by aliens, H2 abducted by kidnappers or H3 delayed by traffic. How do we conclude H3 is the most plausible answer?

Hint: "Probability theory is in fact only common sense reduced to calculus" Pierre Simon Laplace (1749-1827).

Probabilistic Reasoning: Plausibilities

"For plausible reasoning It is necessary to extend the discrete true and false values of truth to continuous plausibilities."[1]
E.T. Jaynes (1922-1998) identified 3 mathematical criteria which must apply to all plausibilities:

1. The degrees of plausibility are represented by real numbers.
2. These numbers must be based on the rules of common sense.
3. Consistency:
   - a. Consistency or non-contradiction: when the same result can be reached through different means, the same plausibility value must be found in all cases.
   - b. Honesty: All available data must be taken into account.
   - c. Reproducibility: Equal levels of knowledge must have the same degree of plausibility.

[1] E. T. Jaynes in "Probability theory: The logic of science" (2003) Cambridge University Press.

Probabilistic Reasoning: Probabilities

- The Cox-Jaynes's theorem proves these plausibilities to be sufficient to define the universal mathematical rules which apply to plausibility p, up to an arbitrary monotonic function. Crucially, these rules **are** the rules of probability.
- If consistent rules for processing plausibility are the result of the evolution of the human brain (or any another autonomous agent), then they must approximate the standard rules of probability. Otherwise one can be systematically exploited to be at a loss ("Dutch Book").
- Important: these probabilities are no longer interpreted as the relative frequency of events (Frequentist interpretation), but rather as measurements of the degree of subjective knowledge or belief (Bayesian interpretation).

Bayesian talk

In the context of Bayesian or statistical machine learning we often want to learn unknown parameters of a model, therefore we want to determine the probability of the parameters given the data we observed, i.e. the posterior probability:

$$p(Parameters|Data) = \frac{p(Data|Parameters)p(Parameters)}{p(Data)}$$

we call $p(Data|Parameters)$ the likelihood or evidence and $p(Parameters)$ our prior belief over the possible parameter values.

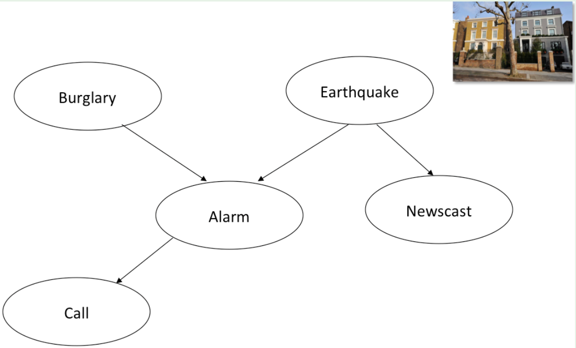
Graphical Model 1

Example (Mr Holmes's conundrum)

- Mr. Holmes is on holiday.
- He receives a call from his neighbor that the alarm of his house went off.
- He thinks that somebody broke into his house.
- Afterwards he hears an announcement on radio that a small earthquake just happened
- Since the alarm has been going off during an earthquake,
- He concludes it is more likely that earthquake causes the alarm (explaining away the burglary).

Graphical Model 2

Example (Mr Holmes's conundrum)



Structure of dependencies between random variables visualised as directed edges for conditional probabilities ($P\left[target|source\right]$) from source to target nodes and nodes with pure source nodes reflecting prior probabilities ($P\left[source\right]$). The Probabilistic Graphical model is a graphical representation of factorised joint probability distribution.

Benefits of Being Bayesian

- We can make inferences based on uncertain information: "Is the barking dog dangerous?"
- Probabilities used to represent degrees of belief
  Strength of a belief is given a value between 0 and 1
  e.g. our belief in proposition A ("dog is going to bite") being true is $P(A) = 0.95$
- Principled decision making: Is
  $P(A) \times \text{Cost of } A \;>\; P(\neg A) \times \text{Cost of } \neg A$?
  $\Rightarrow$ We can select optimal actions based on probabilistic inference.

Bayesian Decision Theory

- Make optimal decisions a^* by maximizing an expected utility

$$a^* \in \arg\max_a \mathbb{E}\left[r(a)\right] = \arg\max_a \sum_{j=1}^{M} r(s_j, a)p(s_j)$$

a : decision/action

s : information about environment/state indexed from 1 to M.

Examples

- Reinforcement Learning (computer science, neuroscience & psychology)
- Optimal control theory (engineering, robotics)
- Bayesian sequential decision theory (statistics, cryptography)