

Fitzwilliam Maths Circle

Topic: Mathematics for AI #2

An agent must choose actions, not just make predictions. In these problems, choose the strategy with the largest *expected total score*. Assume the stated probabilities are correct. Starred exercises are optional extensions. Recall that

$$\mathbb{E}[R] = \sum_r r p(R = r), \quad p(C | T) = \frac{p(T | C)p(C)}{p(T)}.$$

Exercise 1. When should it answer? An agent believes its answer is correct with probability $3/4$. Submitting a correct answer earns 1 point; submitting a wrong answer loses 4 points. Abstaining earns 0.

- (a) Find the expected score from submitting. Should it submit or abstain?
- (b) Replace $3/4$ by $p \in [0, 1]$. For which values of p should it submit? What happens at the boundary?
- (c) Can randomising between the two actions give a better expected score than both?

Exercise 2. A fallible checker. Keep the prior probability $p(C) = 3/4$, where C means that the answer is correct. A checker examines the answer without changing it and returns *pass* (+) or *fail* (−), with

$$p(+ | C) = \frac{4}{5}, \quad p(+ | \overline{C}) = \frac{1}{5}.$$

- (a) Find $p(+)$, $p(C | +)$ and $p(C | -)$.
- (b) Using the scores in Exercise 1, should the agent submit or abstain after each possible result? Ignore the cost of checking for now.

Exercise 3. The price of information. Using the checker in Exercise 2 costs $c \geq 0$ points, paid even if the agent subsequently abstains.

- (a) Find the expected score from checking and then following your rule from Exercise 2, calculated *before* seeing the result. What is the largest c for which checking is at least as good as deciding without a check?
- (b) A perfect checker reveals whether the original answer is correct, but cannot repair it. What is the largest fee worth paying for this checker? Why is this an upper bound on the value of any checker of this answer?

Exercise 4. Learning by doing. A new task uses different scores: 1 for success, 0 for failure. Tool A succeeds independently with probability $3/5$ on each use. Tool B has a fixed, unknown type: it either always succeeds or always fails, with equal probability. Outcomes from A tell us nothing about B . Each task uses one tool once; its outcome is observed before the next choice.

- (a) With just one task remaining, which tool should the agent use?
- (b) With two tasks remaining, find the best strategy starting with each tool, and compare their expected total scores.
- (c) Would the answer change if B instead succeeded independently with probability $1/2$ on every use, independently of A ?

Exercise 5. Can two unhelpful checks help? Return to the answer/abstain scores of Exercise 1, but now suppose $p(C) = 1/2$. A different checker has

$$p(+ | C) = \frac{3}{4}, \quad p(+ | \overline{C}) = \frac{1}{4}.$$

Two checks are *conditionally independent given correctness*: their result probabilities multiply once we fix whether C or $\overline{C}$ holds. Neither check changes the answer.

- With only one free check available, show that the agent still abstains whichever result it sees.
- With two free checks available, after which pairs of results should it submit? Find the expected score before either result is observed.
- Now each check costs $c \geq 0$ points. Consider buying the first check, buying the second only if the first passes, and then submitting only when both pass. For which c does this strategy beat abstaining immediately? Explain why evaluating only the first check in isolation would miss this opportunity.

Exercise 6*. How far ahead should it look? Extend Exercise 4 to $N \geq 1$ tasks. Tool A still succeeds independently with probability $3/5$, but now B is the always-successful type with probability q , where $0 < q < 3/5$, and otherwise always fails.

- Compare the expected total scores of using A throughout and of trying B first, then using B after success or A after failure for all remaining tasks. Find the condition on q and N under which trying B is better.
- For $q = 2/5$, what is the smallest N for which trying B is better?
- Compare trying B first with using A for $0 \leq k < N$ tasks before trying B , following the same success/failure rule afterwards. Can delaying the trial improve the expected total score?

Exercise 7*. Why free information cannot hurt. Let s range over finitely many possible states of the world and a over finitely many available actions. Taking action a in state s gives score $R(a, s)$. A finite-valued signal T can be observed without changing the state, scores or available actions. Write $p(s)$ for the prior and $p(s | t)$ for the posterior after observing $T = t$. Define

$$V_0 = \max_a \sum_s p(s) R(a, s), \quad V_1 = \sum_t p(t) \max_a \sum_s p(s | t) R(a, s),$$

where the sum over t includes only outcomes with positive probability.

- Prove that $V_1 \geq V_0$. *Hint:* fix an action attaining V_0 and use $\sum_t p(t) p(s | t) = p(s)$.
- Must the inequality be strict whenever the signal changes the agent's beliefs? Find a counterexample among the earlier exercises.
- If observing T costs c points, what condition determines whether buying this information is at least as good as acting without it?

Sources: Christopher M. Bishop and Hugh Bishop, *Deep Learning: Foundations and Concepts*, Sections 2.1, 2.2.2 and 4.2 (probability, expectations and decision theory), and Exercise 2.11 (total expectation).