1 RL: Open Discussion

Recall that in Q-learning, we continually update the values of each Q-state by learning through a series of episodes, ultimately converging upon the optimal policy.

(a) Assume that we are given our exploration probability $\epsilon = 1$. Are we still guaranteed to converge upon the optimal policy?

(b) Consider a variant of the Q-iteration algorithm that is changed such that instead of using the policy extracted from our current Q-values, we use a fixed policy instead, with exploration. If this fixed policy happens to be optimal, how does the performance of this algorithm compare to normal Q-iteration?

(c) How would we be able to tell if a given policy is optimal or not?

(d) **(Bonus)** Let's revisit the CandyGrab code from recitation 1 (https://www.cs.cmu.edu/~15381/recitations/rec1/candygrab.zip). What RL strategies does `AgentRL` employ? Does it evaluate states or Q-states?

2 Chain Rule

In class, we discussed the product rule, which states that $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$. Let's now try to extend this to understand the chain rule. By the chain rule, we can write any joint distribution as an incremental product of conditional distributions (in other words, the chain rule can be found by repeated application of the product rule).

(a) Apply the chain rule to express $P(A, B, C)$ in terms of $P(A), P(B|A)$, and $P(C|A, B)$.

(b) There are five other ways to express this probability. Write them out below.

(c) Finally, let's also express $P(A, B, C, D)$ (in any way) using the chain rule.

3 The Distributive Property

We'll start with something more basic. This example will serve as a good reference in trying to understand more complicated concepts.

(a) Compute $\sum_{x=1}^{3} \left(\sum_{y=4}^{6} xy \right)$

(b) Compute $\sum_{x=1}^{3} \left(x \sum_{y=4}^{6} y \right)$

(c) What do you notice about the values? Why is this the case?

4 Generic Functions

Consider two sets $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2\}$. Let us also define arbitrary functions $f(x)$, $g(y)$, and $h(x, y)$. The notation $\sum_{x \in X} f(x)$ means that we want to apply the function f to all elements of X and add all the results. For the X defined above, this means $f(x_1) + f(x_2) + f(x_3)$.

(a) Consider the sums $\sum_{x \in X} \sum_{y \in Y} h(x, y)$ and $\sum_{y \in Y} \sum_{x \in X} h(x, y)$. Are they the same? Can they be simplified? In general, is it valid to switch the sums in this manner?

(b) Now consider

$$\sum_{x \in X} \sum_{y \in Y} f(x)g(y)h(x, y).$$

Can we simplify the calculation of this sum?

5 Magnetic Factors

Observe that in a summation over x, any term that does not depend on x can be treated as a constant, and therefore moved out of the sum. This is to say that $\sum_{x \in X} cf(x) = c\left(\sum_{x \in X} f(x)\right)$, where c is any term that is constant *with respect to* x. This includes functions of y or any other variable.

Consider this analogy: pretend each Σ is a *magnet* that attracts only factors containing the variable it is iterating over. Factors can only pass through this magnet if they are independent of that variable. All factors will be pulled as far to the left as possible. Since some factors contain more than one variable, the ordering of the summations affects the result, and choosing an optimal ordering can greatly speed up computation.

With this in mind, simplify the following expressions so that computing them requires as few operations as possible.

(a) $\sum_{s \in S} \sum_{r \in R} \sum_{q \in Q} f(q)g(q,r)h(q,r,s)$

(b) $\sum_{a \in A} \sum_{b \in B} \sum_{c \in C} f(b)f(c)h(a,b,c)g(a,j)g(a,m)$

(c) $\sum_{d \in D} \sum_{k \in K} \sum_{c \in C} \sum_{x \in X} g(c,k)f(x)h(c,k,x)f(k)l(c,d,k,x)g(k,x)$

6 And Finally, Some Probability Again

Recall the example of *marginalization*, which means summing out variables from a joint distribution. Consider three binary random variables A, B, and C with domains $\{+a, -a\}$, $\{+b, -b\}$, and $\{+c, -c\}$, respectively. Remember that $P(A)$ refers to the table of probabilities of all the elements of the domain A.

(a) Express $P(A)$ in terms of the joint distribution $P(A, B, C)$. Your answers should contain summations.

(b) Express $P(A)$ in terms of $P(C)$, $P(B \mid C)$ and $P(A \mid B, C)$.

(c) Expand the sums from part (b) to show the two elements of $P(A)$ ($P(+a)$ and $P(-a)$) in terms of the individual probabilities (e.g. $P(+b)$, $P(-c)$ instead of $P(B)$ or $P(C)$).