CS234 Problem Session

Week 7: Feb 24

1) [CA Session] Useful Probability Bounds

In this problem, we will derive bounds to answer questions of the form: given a random variable Z with expectation $\mathbb{E}[Z]$, how likely is Z to be close to its expectation?

(a) First, we will prove Markov's inequality: Let $Z \geq 0$ be a non-negative random variable. Prove that for all $t \geq 0$,

$$\mathbb{P}(Z \geq t) \leq \frac{\mathbb{E}[Z]}{t}$$

(b) Next, we will prove Chebyshev's inequality. Let Z be any random variable with $Var(Z) < \infty$. Prove that for all $t \geq 0$,

$$\mathbb{P}(Z \geq \mathbb{E}[Z] + t \text{ or } Z \leq \mathbb{E}[Z] - t) \leq \frac{Var(Z)}{t^2}$$

(c) It can be useful to derive tighter bounds through exponentially decreasing functions. Let us define the moment generating function for a random variable Z as

$$M_Z(\lambda) = \mathbb{E}[exp(\lambda Z)]$$

We will now prove the Chernoff bound. Let Z be a random variable. Prove that for any $t \geq 0$,

$$\mathbb{P}(Z \geq \mathbb{E}[Z] + t) \leq min_{\lambda \geq 0} \mathbb{E}[e^{\lambda(Z - \mathbb{E}[Z])}]e^{-\lambda t} = min_{\lambda \geq 0} M_{Z - \mathbb{E}[Z]}(\lambda) e^{-\lambda t}$$

2) [Breakout Rooms] KL Divergence

The Kullback-Leibler (KL) divergence is defined is a measure of how different a probability distribution is from a second reference probability distribution. For discrete probability distributions P and Q defined over the same probability space X, the KL divergence is defined as

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log(\frac{P(x)}{Q(x)})$$

Show that the KL divergence is guaranteed to be non-negative.

3) [Breakout Rooms] Probably Approximately Correct

Let $A(\alpha, \beta)$ be a hypothetical reinforcement learning algorithm, parametrized in terms of α and β such that for any $\alpha > \beta > 1$, it selects action a for state s satisfying $|Q(s,a) - V^*(s)| \leq \frac{\beta}{\alpha}$ in all but $N = \frac{|S||A|\alpha\beta}{1-\gamma}$ steps with probability at least $1 - \frac{1}{\beta^2}$.

Per the definition of *Probably Approximately Correct Reinforcement Learning*, express N as a function of $|S|$, $|A|$, δ, ϵ and γ. What is the resulting N? Is algorithm A probably approximately correct? Briefly justify.