CS234 Problem Session Solutions

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 $t \geq 0$,

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

Solution First, notice that $\mathbb{P}(Z \geq t) = \mathbb{E}[\mathbf{1}\{Z \geq t\}]$, and that if $Z \geq t$, then it must be that $\frac{Z}{t} \geq \mathbf{1}\{Z \geq t\}$. Otherwise, if $Z < t$, then we have that $\frac{Z}{t} \geq 0 = \mathbf{1}\{Z \geq t\}$. Thus, we have that
$\mathbb{P}(Z \geq t) = \mathbb{E}[\mathbf{1}\{Z \geq t\}] \leq \mathbb{E}[\frac{Z}{t}] = \frac{\mathbb{E}[Z]}{t}$, as required.

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

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

Solution This result follows from Markov's inequality. Notice that if $Z \geq \mathbb{E}[Z] + t$, then it is also true that $(Z - \mathbb{E}[Z])^2 \geq t^2$. Similarly, if $Z \leq \mathbb{E}[Z] - t$, then we have $(Z - \mathbb{E}[Z])^2 \geq t^2$. Hence, by Markov's inequality, we have that

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

(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 $t \geq 0$,

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

Solution We will again prove this using Markov's inequality. For any $\lambda > 0$, we see that $Z \geq \mathbb{E}[Z] + t$ if and only if $e^{\lambda Z} \geq e^{\lambda(\mathbb{E}[Z]+t)}$. Rearranging, we have $e^{\lambda(Z-\mathbb{E}[Z])} \geq e^{\lambda t}$. Now, we can apply Markov's inequality and see that

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

Notice that this bound certainly holds if $\lambda = 0$. Further, we have proven this for an arbitrary non-negative λ, so we can minimize the bound with respect to λ to achieve the tightest bound.

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.

Solution This can be proven in a few ways. We will prove this using Jensen's inequality. We will show that $-D_{KL}(P||Q) \leq 0$.

$$\begin{aligned}
-D_{KL}(P||Q) &= -\sum_{x \in X} P(x) \log(\frac{P(x)}{Q(x)}) \\
&= \sum_{x \in X} P(x) \log(\frac{Q(x)}{P(x)}) \\
&\leq \log \sum_{x \in X} P(x) \frac{Q(x)}{P(x)} \text{ by Jensen's inequality since log is concave.} \\
&= \log \sum_{x \in X} Q(x) \\
&= \log(1) \\
&= 0
\end{aligned}$$

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 $|S|$, $|A|, \delta, \epsilon$ and γ. What is the resulting N? Is algorithm A probably approximately correct? Briefly justify.

Solution We want to achieve the bound that $|Q(s, a) - V^*(s)| \leq \epsilon$ with probability $1 - \delta$. So let $\frac{\beta}{\alpha} = \epsilon$ and $1 - \frac{1}{\beta^2} = 1 - \delta$, which gives $\alpha = \frac{1}{\epsilon\sqrt{\delta}}$ and $\beta = \frac{1}{\sqrt{\delta}}$.

Substituting, $N = \frac{|S||A|\alpha\beta}{1-\gamma} = \frac{|S||A|}{\epsilon\delta(1-\gamma)}$.

Since N is a polynomial function of $|S|, |A|, \frac{1}{\epsilon}$ and $\frac{1}{\delta}$ and achieves the ϵ, δ bounds above, then A is PAC.