Machine Learning theory (CS 6783)

Problem set 0

This assignment is not counted towards your grade and you don't need to submit it. Its meant to brush up some basics and get your hands wet.

Some facts/results you will need for this assignment :

1. Markov Inequality : For any non-negative integrable random variable X, and any $\epsilon > 0$,

$$P(X \geq \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}$$

2. Hoeffding inequality : Let $X_1, \ldots, X_n$ be n independent identically distributed (iid) random variables such that each $X_i \in [a, b]$ and let μ be the expected value of these random variables. Then,

$$P\left(\left|\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right)$$

3. Berntein inequality : Let $X_1, \ldots, X_n$ be n independent identically distributed (iid) random variables such that each $X_i \in [a, b]$ and let μ be the expected value of these random variables and let σ^2 be their variance. Then,

$$P\left(\left|\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2\sigma^2 + (b-a)\epsilon/3}\right)$$

4. Hoeffding-Azuma inequality : Let $(X_t)_{t \geq 0}$ be a martingale and for any $t \geq 2$, $|X_t - X_{t-1}| \leq c$ then, for any n,

$$P\left(|X_N - X_0| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2nc}\right)$$

Q1 Let $X_1, \ldots, X_n$ be n independent identically distributed (iid) random variables that are possibly unbounded having expected value μ. Also assume that for any $i \in [n]$, $|X_i| \leq c$. Use Markov inequality to provide a bound of form

$$P\left(\left|\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right| \geq \epsilon\right) \leq F(n, c, \epsilon)$$

What is the form of $F(n, c, \epsilon)$.

Hint :

(a) Write μ as $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} X'_t\right])$ where $X'_1, \ldots, X'_n$ are drawn iid from same distribution.

(b) Notice that $X_t - X'_t$ has same distribution as $\varepsilon_t(X_t - X'_t)$ where each ε_t is a Rademacher random variable (ie. $\{\pm 1\}$ valued random variable which is either $+1$ or -1 with equal probability)

(c) Use the fact that $\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t\right|\right] \leq \sqrt{\frac{2}{n}}$

Q2 Markov Vs Hoeffding Vs Bernstein

Let D be some distribution over the interval $[a, b]$ such that expectation of random variables X's drawn from D is μ and variance is σ^2. We want the following statement to hold :

For any $\delta > 0$ and $\epsilon > 0$, as long as $n > n(\epsilon, \delta)$, if $X_1, \ldots, X_n$ are drawn iid from D, with probability at least $1 - \delta$,

$$\left|\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right| \leq \epsilon$$

What is $n(\epsilon, \delta)$ implied by

(a) Markov bound (more specifically bound from Q1)

(b) Hoeffding inequality

(c) Bernstein inequality

Which of the three is better when

(a) σ^2 is much smaller compared to $(a - b)^2$ and δ is large (say $1/2$)

(b) σ^2 is much smaller compared to $(a - b)^2$ and δ is small

(c) σ^2 is large and δ is small

Basically get a feel for when each of these bounds are useful.

Q3 In this question we will learn to derive what is called Bounded difference inequality or McDiarmid's inequality using Hoeffding-Azuma inequality. The bounded difference inequality states that :

Theorem 1. *Consider independent $\mathcal{X}$ valued random variables $X_1, \ldots, X_n$. Let $F : \mathcal{X}^n \mapsto \mathbb{R}$ be any function such that for any $x_1, \ldots, x_n, x'_1, \ldots, x'_n$ and any $i \in [n]$,*

$$|F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n)| \leq c$$

(the above is called the bounded difference property). Then we have that

$$P\left(|F(X_1, \ldots, X_n) - \mathbb{E}\left[F\right]| \geq \epsilon\right) \leq 2 \exp\left(-\frac{\epsilon^2}{2nc^2}\right)$$

Prove the above theorem using Hoeffding Azuma bound.
Hint :

(a) Define the right martingale sequence using conditional expectations of the function

(b) Use the bounded difference inequality to show that the premise of the Hoeffding-Azuma bound holds.