Advanced Topics in Machine Learning 2024

Warmup for the Course

We assume that students joining this course have previously taken our "Machine Learning A" (ML-A) course, or its precursor the "Machine Learning" course. In this self-preparation document, we ask some questions that will be relevant for this course and help you prepare for the technical components of this course.

1 Hoeffding's Inequality

We assume that the students are familiar with Markov's, Chebyshev's, and Hoeffding's inequalities. If you did not encounter them in the past, you can learn them from Yevgeny's lecture notes `https://drive.google.com/file/d/1netJc1pmtW9m2yOD_JGF_LdW1N87TB7u/view?usp=sharing`.

1.1 Illustration of Markov's, Chebyshev's, and Hoeffding's inequalities

Make 1,000,000 repetitions of the experiment of drawing 20 i.i.d. Bernoulli random variables $X_1, \ldots, X_{20}$ (20 coins) with bias $\frac{1}{2}$ and answer the following questions.

1. Plot the empirical frequency of observing $\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha$ for $\alpha \in \{0.5, 0.55, 0.6, \ldots, 0.95, 1\}$.
2. Explain why the above granularity of α is sufficient. I.e., why, for example, taking $\alpha = 0.51$ will not provide any extra information about the experiment.
3. In the same figure plot the Markov's bound[1] on $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$.
4. In the same figure plot the Chebyshev's bound[2] on $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$. (You may have a problem calculating the bound for some values of α. In that case and whenever the bound exceeds 1, replace it with the trivial bound of 1, because we know that probabilities are always bounded by 1.)
5. In the same figure plot the Hoeffding's bound[3] on $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$.
6. Compare the four plots.
7. For $\alpha = 1$ and $\alpha = 0.95$ calculate the exact probability $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$. (No need to add this one to the plot.)

Repeat the question with $X_1, \ldots, X_{20}$ with bias 0.1 (i.e., $\mathbb{E}[X_1] = 0.1$) and $\alpha \in \{0.1, 0.15, \ldots, 1\}$.

Discuss the results.

[1] Markov's bound is the right hand side of Markov's inequality.
[2] Chebyshev's bound is the right hand side of Chebyshev's inequality.
[3] Hoeffding's bound is the right hand side of Hoeffding's inequality.

1.2 The Role of Independence

Design an example of identically distributed, but *dependent* Bernoulli random variables $X_1, \ldots, X_n$ (i.e., $X_i \in \{0,1\}$), such that

$$\mathbb{P}\left(\left|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \frac{1}{2}\right) = 1,$$

where $\mu = \mathbb{E}\left[X_i\right]$.

Note that in this case $\frac{1}{n}\sum_{i=1}^{n} X_i$ does not converge to μ as n goes to infinity. The example shows that independence is crucial for convergence of empirical means to the expected values.

2 Union Bound

We assume that you know how to use the union bound. For example, for $i \in \{1, \ldots, k\}$ let $S_i = \frac{1}{n}\sum_{j=1}^{n} X_{i,j}$ be an average of n independent Bernoulli random variables $X_{i,j}$ with bias p. Let $\varepsilon \in (0,1)$. Bound the following probability: $\mathbb{P}(\exists i : |S_i - p| \geq \varepsilon)$.

3 Normal and Laplace distributions

In the course we will be using normal distributions (also known as Gaussian distributions) and Laplace distributions as a tool for differential privacy. In both cases we will be using distributions with mean zero, $\texttt{Lap}(0, b)$ and $\mathcal{N}(0, \sigma^2)$, but we vary the scale parameters b and σ.

1. What is the variance of a Laplace random variable with scale parameter b? (The intended answer involves computing an integral.)

2. For a normal distribution, how should the scale parameter σ be chosen if we want the variance to be the same as that of a Laplace random variable with scale parameter b?

3. For a (measurable) set $S \subseteq \mathbf{R}$ and $X \sim \texttt{Lap}(0, b)$ show that

   $$\exp(-1/b)\Pr[1 + X \in S] \leq \Pr[X \in S] \leq \exp(1/b)\Pr[1 + X \in S] \enspace .$$

   This means that X and $1 + X$ are, in a sense, hard to distinguish if b is sufficiently large.

4. For a (measurable) set $S \subseteq \mathbf{R}$ and $X \sim \mathcal{N}(0, \sigma^2)$ show that we do *not* in general have $\Pr[X \in S] \leq f(\sigma)\Pr[1 + X \in S]$ for any function f. That is, for a given value of $f(\sigma)$ give an example of a set S for which the inequalities do not hold.