Mathematics of Machine Learning
Homework 7

Instructors: Afonso S. Bandeira & Nikita Zhivotovskiy
Course Coordinator: Pedro Abdalla Teixeira

April 16, 2021

Try to solve the questions before looking to the answers. Every item must be proved rigorously. Starred problems are harder.

Problem 1

Let X be a random variable. Prove the following classical results

(a) Markov's inequality: For every $a > 0$, prove that

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}|X|}{a}$$

(b) Chebyshev's inequality: For every $a > 0$ and $p \geq 1$, prove that

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}|X|^p}{a^p}.$$

(Hint: Use that $\{a \in \mathbb{R} : |X| \geq a\} = \{a \in \mathbb{R} : |X|^p \geq a^p\}$.)

(c) Integral identity/Layer cake representation: For a nonnegative integrable random variable Y, prove that

$$\mathbb{E}Y = \int_0^\infty \mathbb{P}(Y \geq t)dt.$$

(d) Generalize the result above for an integrable random variable X that may take negative values. (Hint: Split the random variable in a sum of two nonnegative random variables)

Problem 2

In this exercise we denote $Var(X)$ by the variance of the random variable X. Consider the random variables $X_1, \dots, X_n$ and the random sum $Z := \sum_{i=1}^n X_i$.

(a) If $X_1, \dots, X_n$ are independent, prove that $Var(Z) = \sum_{i=1}^n Var(X_i)$.

(b) Is independence a necessary condition for the formula above?

(c) Let $\mathbb{E}_i$ denote the conditional expectation operator, conditioned on $X_1, \dots, X_i$ with the convention that $\mathbb{E}_0$ is the expectation with respect to all $X_1, \dots, X_n$. Check that

$$Z - \mathbb{E}Z = \sum_{i=1}^n \Delta_i,$$

where $\Delta_i := \mathbb{E}_i Z - \mathbb{E}_{i-1} Z$.

(d) Prove that $Var(Z) = \sum_{i=1}^n \mathbb{E}\Delta_i^2$

Problem 3

Consider a binary classification problem where $\mathcal{X} = \mathbb{N}$ and the class $\mathcal{F}$ consists of all classifiers that are equal to one on exactly one integer.

(a) For the class $\mathcal{F}$, construct a sample compression scheme of size one.

(b) Using a direct computation, prove that $\mathcal{F}$ is PAC learnable with the sample complexity $n(\varepsilon, \delta) = \lceil \frac{1}{\varepsilon} \log \frac{1}{\delta} \rceil$.

(Hint: It might be convenient to use a reconstruction function ρ that is not restricted to $\mathcal{F}$.)