# Optimization for Machine Learning HW 1

Due: 9/13/2022

This homework is optional. You may do as much of it as you like.

1. This question provides practice thinking about random variables.

   (a) Is there a random variable $X$ supported on the positive integers that has finite mean $\mathbb{E}[X]$ but infinite second moment $\mathbb{E}[X^2]$? If so, explicitly state a probability mass function for such an $X$ and prove that it has the desired properties. If not, prove that no such distribution random variable exists.

   **Solution:**

   (b) If $X$ is a random variable satisfying $\mathbb{E}[\|X - \mathbb{E}[X]\|^n] \le \sigma^n$ for some $n > 0$, show that for any $\delta > 0$, $P[\|X - \mathbb{E}[X]\| \le \frac{\sigma}{\delta^{1/n}}] \ge 1 - \delta$. This statement is often written instead as "with probability at least $1 - \delta$, $\|X - \mathbb{E}[X]\| \le \frac{\sigma}{\delta^{1/n}}$".

   **Solution:**

   (c) Suppose that $X$ is a random variable such that for all real numbers (not just integers!) $n > 0$, $\mathbb{E}[\|X - \mathbb{E}[X]\|^n]^{1/n} \le \sigma\sqrt{n}$. Show that with probability at least $1 - \delta$, $\|X - \mathbb{E}[X]\| \le \sigma\sqrt{2\exp(1)\log(1/\delta)}$. Distributions satisfying this property are called *subgaussian*. The Normal distribution is an example of a distribution satisfying this kind of property.

   **Solution:**

2. This question provides practice in some linear algebra ideas.

   (a) For any matrix $M$, the *operator norm* of $M$ is $\|M\|_{\text{op}} = \sup_{\|v\|=1} \|Mv\|$. Prove that the operator norm satisfies the triangle inequality: for all matrices $A$ and $B$ of the same dimensions, $\|A+B\|_{\text{op}} \le \|A\|_{\text{op}} + \|B\|_{\text{op}}$.

   **Solution:**

   (b) Most of the matrices we will discuss in this class are *symmetric* matrices. The *real spectral theorem* states that any symmetric matrix $M \in \mathbb{R}^{d\times d}$ has an orthonormal basis of eigenvectors. That is, there exists $v_1, \dots, v_d$ such that each $v_i$ has norm 1, $\langle v_i, v_j\rangle = 0$ for all $i \ne j$, and $Mv_i = \lambda_i v_i$ for some real numbers $\lambda_1, \dots, \lambda_d$. Prove that $\|M\|_{\text{op}} = \max_i |\lambda_i|$ for any symmetric matrix $M$.

   **Solution:**

3. Suppose you are working for a online store and you need to predict the probability that a person who visits your homepage will buy something. You have a dataset $z_1, \dots, z_N$ where $z_i$ is 1 if the $i$th visitor

to the homepage bought something, and 0 otherwise. You assume that each $z_i$ is an independent and indentically distributred random variable, so your task is just to learn $p = P[z = 1] = \mathbb{E}[z]$. You decide to use the simple estimate $\hat{p} = \frac{1}{N}\sum_{t=1}^{N} z_i$. Show that for any $N$ and any $p$, $\mathbb{E}[|\hat{p} - p|] \leq \frac{\sqrt{p(1-p)}}{\sqrt{N}}$.

**Solution:**