Problem Set 1

Name:

Due: Midnight EST on January 19th, 2022

The goal of this problem set is to establish familiarity with some of the theoretical and computational tools that will be useful for the course. The theoretical and coding exercises below review topics from linear algebra, analysis, and probability. We use (T) to denote theory exercises and (C) to denote coding exercises. We lastly use the flag **Required** to denote problems that are required for all students.

Students should do all the required problems and at least 2 of the remaining problems (for a total of 6 problems).

Linear Algebra

Problem 1 (T). Let $A \in \mathbb{R}^{n \times n}$ be a diagonalizable matrix with eigenvalues $\{\lambda_i\}_{i=1}^n$.
(a) If $\{\lambda_i\}_{i=1}^n \subset (-1, 1]$, compute the eigenvalues of $(I + A)^{-1}$, where I is the identity matrix.
(b) Compute the eigenvalues of A^k for fixed $k \in \mathbb{Z}_+$.

Problem 2 (T).
(a) Let $\Sigma \in \mathbb{R}^{n \times n}$ be a diagonal matrix with diagonal entries $\{\sigma_i\}_{i=1}^n \in \mathbb{R}$. Compute $\Sigma^\dagger$.
Hint: Remember that σ_i can equal 0 for some i.
(b) Let $A = U\Sigma V^T \in \mathbb{R}^{n \times n}$ by the Singular Value Decomposition (SVD). Verify that $A^\dagger = V\Sigma^\dagger U^T$.
Hint: Check that $V\Sigma^\dagger U^T$ satisfies the properties of the pseudoinverse from Definition 6 of Lecture 1.

Problem 3 (C, Required). Generate a square matrix $A \in \mathbb{R}^{5000 \times 5000}$ and a vector $b \in \mathbb{R}^{5000 \times 1}$ with entries drawn i.i.d from a standard normal distribution. Compare the runtime of solving $Ax = b$ using the numpy `solve` function against that of first computing A^{-1} with the numpy `inv` function and then computing $x = A^{-1}b$.
Remark: For timing, one can use the `time.time()` function in Python. This exercise is to reinforce that the `solve` function should be used in place of `inv` when solving square matrix linear systems.

Problem 4 (C). Generate a square matrix $A \in \mathbb{R}^{100 \times 100}$ with entries drawn i.i.d. from a standard normal distribution. Compute U, Σ, V^T via the numpy `svd` function. Compute $A^\dagger$ via the numpy `pinv` function and verify that it matches $V\Sigma^\dagger U^T$.

Analysis

Problem 5 (T, Required). For $X \in \mathbb{R}^{d \times n}$, $w \in \mathbb{R}^{1 \times d}$, $y \in \mathbb{R}^{1 \times n}$, let $\mathcal{L}(w) = \|wX - y\|_2^2$. Compute $\nabla\mathcal{L}(w)$.
Hint: Try computing the partial derivatives $\frac{\partial L}{\partial w_i}$ and group the terms into a vector. Are there any obvious patterns that emerge?
Remarks: This is a required exercise since this computation will appear in Lecture 2.

Problem 6 (T, Required). In Lecture 1, we briefly introduced the notion of norms on function spaces. In particular, we presented the example of the L^2 norm of a function $f : \mathbb{R} \to \mathbb{R}$, which in general is given by:

$$\|f\|_{L^2}^2 = \int_{-\infty}^{\infty} |f(x)|^2 dx$$

In this course, we will commonly use another norm, which we call the $L^2(\mu)$ norm, that is defined as follows. For $f : \mathbb{R} \to \mathbb{R}$:

$$\|f\|_{L^2(\mu)}^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |f(x)|^2 e^{-\frac{x^2}{2}} dx$$

Compute $\|f\|_{L^2(\mu)}$ when $f(x) = \max(x, 0)$ and $\|g\|_{L^2(\mu)}$ when $g(x) = \begin{cases} 1 & \text{if } x \geq 0. \\ 0 & \text{if } x < 0. \end{cases}$.

Hint: Recall, that the probability density function of the standard normal Gaussian is given by:

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Use the fact that the integral of a density must be 1 to compute $\|g\|_{L^2(\mu)}$. Similarly, use the expectation of the Gaussian to compute $\|f\|_{L^2(\mu)}$.

Remarks. The computations in this problem will re-appear in the derivation of normalizing factors for the NNGP and NTK later in this course.

Problem 7 (C). Generate vectors $a, b \in \mathbb{R}^{500}$ with entries drawn i.i.d standard normal distribution.

(a) Compute $\|a\|_2^2$ via the numpy `norm` function and via your own implementation. Check that the two match.

(b) Compute the mean squared error (MSE) between vectors a, b given by $\frac{1}{500}\|a - b\|_2^2$.

Probability

Problem 8 (T, Required). Compute the following integral:

$$\int_{-\infty}^{\infty} e^{-x^2+4x} dx$$

Hint: This problem is in the probability section for a reason. Complete the square and see if a known integral appears.

Remarks: This technique will be useful in computing the NNGP and NTK later on in this course.

Problem 9 (T). Let $v, w \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, I_{n\times n})$. Note $v, w \in \mathbb{R}^n$.

(a) Compute $\mathbb{E}_w[ww^T]$ and $\mathbb{E}_w[\|w\|_2^2]$.

(b) Compute $\mathbb{E}_{(w,v)}[\langle v, w\rangle]$, where $\langle\cdot,\cdot\rangle$ is the standard dot product on vectors in $\mathbb{R}^n$.

Problem 10 (C). Generate $A \in \mathbb{R}^{n\times n}$ with entries drawn from an i.i.d. standard normal distribution.

(a) For $n \in \{10, 100, 1000\}$, compute $\frac{1}{n}AA^T$. What matrix does $\frac{1}{n}AA^T$ approach for large n?

(b) Fix $n = 10$. Sample m such matrices $\{A_i\}_{i=1}^m$ for $m \in \{10, 100, 1000\}$ and compute $\frac{1}{m}\sum_{i=1}^m A_i^2$. To what matrix does this sum converge?