Assignment 1: Math for Machine Learning

Due October 14, 2022 at 23:59

This assignment is to be done individually.

Important Note: The university policy on academic dishonesty (cheating) will be taken very seriously in this course. You may not provide or use any solution, in whole or in part, to or by another student.

You are encouraged to discuss the concepts involved in the questions with other students. If you are in doubt as to what constitutes acceptable discussion, please ask! Further, please take advantage of office hours offered by the instructor and the TA if you are having difficulties with this assignment.

DO NOT:

- Give/receive code or proofs to/from other students
- Use Google to find solutions for assignment

DO:

- Meet with other students to discuss assignment (it is best not to take any notes during such meetings, and to re-work assignment on your own)
- Use online resources (e.g. Wikipedia) to understand the concepts needed to solve the assignment.

Submitting Your Assignment

- The assignment must be submitted online on Gradescope. (No need to submit to Canvas!)
- You must submit a report in PDF format. You may typeset your assignment in LaTeX or Word, or submit neatly handwritten and scanned solutions. We will not be able to give credit to solutions that are not legible.
- Please indicate on Gradescope the area of your submission that corresponds to each question part.

---

1 SVD and Eigendecomposition

1.1 SVD and Eigendecomposition Basics

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$.

a) Compute $A^\top A$.

b) Show that the following is the eigendecomposition of $A^\top A$:

$$\begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \tag{1}$$

Hints: you can refer to the characteristics of eigendecomposition components of a matrix.

c) Find the singular value matrix for A. That is, if $A = U\Sigma V^\top$ is the SVD of A, find Σ.

1.2 Geometric Interpretation of SVD

Consider a SVD of a matrix B as follows:

$$B = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^\top \tag{2}$$

a) A matrix $R \in \mathbb{R}^{2\times 2}$ is a **2D rotation matrix** if it has the following form:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3}$$

where $\theta \in \mathbb{R}$. Geometrically speaking, $R_\theta \vec{v}$ rotates $\vec{v}$ counterclockwise by angle θ, for any $\vec{v} \in \mathbb{R}^2$, as shown in Figure 1.

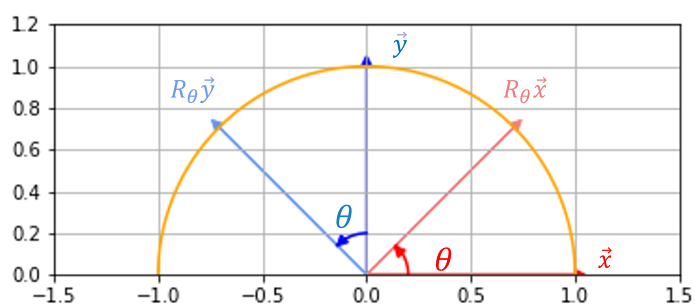

Figure 1: In this case, $\vec{x} = (1, 0)$ and $\vec{y} = (0, 1)$ are both rotated by $\theta = \frac{\pi}{4}$.

Show that $U = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$ and $V^\top = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^\top$ are both rotation matrices, and find their corresponding rotational angles θ_U and $\theta_{V^\top}$.

b) Give an explanation for all geometric transformations performed by the SVD of B. In what order are the transformations performed?

2 Convexity and Linear Algebra

2.1 Taylor Expansions

Given $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, consider a nonlinear function $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ as follows:

$$f(\vec{x}) = 2x_1^2 + x_2^2 + x_3^2 + 2x_2x_3 \tag{4}$$

a) Compute the **Gradient** and **Hessian matrix** of f.

b) Find the **second order Taylor Expansion** at the point $(x_1 = 0, x_2 = 0, x_3 = 0)$.

c) State whether f is **Convex** or **Strictly Convex**. Prove your claim.

2.2 Matrix Rank and Inverse

Prove that if $A \in \mathbb{R}^{m \times n}$, with $m \geq n$, is full rank, then $A^\top A$ is invertible via the following steps:

a) Prove that $A\vec{x} = \vec{0}$ if and only if $\vec{x} = \vec{0}$.

b) Prove that $A^\top A$ is positive definite.

c) Prove that any symmetric positive definite matrix is always invertible by using Eigendecomposition to construct the inverse.

2.3 The Normal Equations

Let $\vec{x} \in \mathbb{R}^n$, $\vec{b} \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$.

a) Suppose $m = n$ and A is full rank. Determine the vector $\vec{x}$ that minimizes

$$\|A\vec{x} - \vec{b}\|_2^2. \tag{5}$$

b) Suppose A is not full rank and we are given A and $\vec{b}$. Show that there always exists $\vec{x}$ such that

$$A^\top A\vec{x} = A^\top \vec{b}. \tag{6}$$

You may use the following fundamental property in linear algebra:

Let $R(A) \subset \mathbb{R}^m$ denote the range, or column space, of a matrix $A \in \mathbb{R}^{m\times n}$ (the span of the columns of A), and let $N(A) \subset \mathbb{R}^n$ denote the nullspace of A; that is, the span of the set of vectors $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{0}$.

Then, any vector $\vec{b} \in \mathbb{R}^m$ can be written as $\vec{b} = \vec{b}_1 + \vec{b}_2$, where $\vec{b}_1 \in R(A)$ and $\vec{b}_2 \in N(A^\top)$.

3 Probability

3.1 Conditional Bayes' Rule

a) Let X, Y, Z be random variables. Let x, y, z be the realized values of the random variables X, Y, Z. Use the definitions of conditional probabilities to prove that

$$p(x, y|z) = p(x|y, z)p(y|z). \tag{7}$$

b) Use part a) to show that

$$p(y|x, z) = \frac{p(x|y, z)p(y|z)}{p(x|z)}. \tag{8}$$

3.2 Multivariate Normal Distributions

a) Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be normally distributed with mean vector $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$. Show that $p(\vec{x}) = p(x_1)p(x_2)$. Make sure to state what the distributions of x_1 and x_2 are and why.

b) Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be normally distributed with mean vector $\mu = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$.

- Compute the means and variances of x_2 and $x_1|x_2 = 3$.
- Compare $\mathbb{E}[x_2]$ with $\mathbb{E}[x_1|x_2 = 3]$, and provide an intuitive geometric explanation of the difference.

Hint: Consider the geometric interpretation of normal distributions, marginal distributions, and conditional distributions. Also, $\Sigma = U\Lambda U^{-1}$, where

$$U \approx \begin{bmatrix} -0.92 & 0.38 \\ 0.38 & 0.92 \end{bmatrix}, \quad \Lambda \approx \begin{bmatrix} 0.79 & 0 \\ 0 & 2.20 \end{bmatrix}. \tag{9}$$