MATH FOUNDATION OF DATA SCIENCE HOMEWORK 1

$>>$**NAME:** $<<$

Instructions: Please print the homework on paper, finish it, and submit it on time as indicated on Moodle.

1 Linear Space Fundamentals

Q1 [5 points] Let X be a linear space. Prove that the set of all linear combinations of its elements $x_1, \ldots, x_j$ is a subspace, and that it is the smallest subspace of X containing $x_1, \ldots, x_j$. This is called the subspace spanned by $x_1, \ldots, x_j$.
Solution:

Q2 [5 points] Prove that the following two linear spaces over the same field K are isomorphic:
(i) Set of all row vectors: $(a_1, a_2, \cdots, a_n)$ where $a_j \in K$; addition, multiplication defined componentwise. This space is denoted as K^n.
(ii) Set of all functions with values in K, defined on an arbitrary set S of n distinct points.
Solution:

Q3 [10 points] Take X equal to the space of all polynomials $p(s)$ with complex coefficients of degree less than n, and take $U = \mathbb{C}^n$. We choose $s_1, \cdots, s_n$ as n distinct complex numbers, and define the linear mapping $T : X \mapsto U$ by

$$Tp = (p(s_1), \cdots, p(s_n)).$$

Prove that the range of T is all of U; that is, the values of p at $s_1, \cdots, s_n$ can be prescribed arbitrarily.
Solution:

Q4 [10 points] Let X be a linear space, and let A, B be two linear mappings from X to X. Prove that if $AB = I$, then $BA = I$.
Solution:

Q5 [10 points] Suppose T and S are linear maps of a finite dimensional vector space into itself. Show that the rank of ST is less than or equal the rank of S. Show that the dimension of the nullspace of ST is less than or equal the sum of the dimensions of the nullspaces of S and of T.
Solution:

Q6 [10 points] Write

$$y = [1, -2, 5]$$

as a linear combination of

$$\begin{aligned} x_1 &= [1, 1, 1], \\ x_2 &= [1, 2, 3], \\ x_3 &= [2, -1, 1]. \end{aligned}$$

Solution:

2 Matrix Decomposition

Q7 [10 points] The spectral norm of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as:

$$\|A\|_2 := \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

where $\|v\|_2 = \sqrt{v^\top v}$ is the Euclidean norm for a vector v. Prove that $\|A\|_2$ is the largest singular value of A.
Solution:

Q8 [10 points] The Fibonacci sequence $f_0, f_1, \cdots$ is defined by the recurrence:

$$f_{n+1} = f_n + f_{n-1},$$

with the starting $f_0 = 0, f_1 = 1$. Rewrite the recurrence in matrix-vector form:

$$\begin{bmatrix} f_n \\ f_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} f_{n-1} \\ f_n \end{bmatrix}.$$

We deduce recursively that

$$\begin{bmatrix} f_n \\ f_{n+1} \end{bmatrix} = A^n \begin{bmatrix} f_0 \\ f_1 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

(1)[5 points] Find the eigendecomposition $A = U \Lambda U^\top$.
(2)[5 points] Using $A^n = U \Lambda^n U^\top$, derive a closed-form solution for f_n.
Solution:

Q9 [10 points] Let $H, M \in \mathbb{R}^{n\times n}$ be two symmetric real matrices. Suppose M is positive, that is, $\langle x, Mx\rangle > 0$ for any $x \neq 0$. The generalized Rayleigh quotient is defined as:

$$R_{H,M}(x) = \frac{\langle x, Hx\rangle}{\langle x, Mx\rangle}.$$

Consider the same minimum problem as for the original Rayleigh quotient: Minimize $R_{H,M}(x)$, that is:

$$\min_{x\in\mathbb{R}^n, x\neq 0} \frac{\langle x, Hx\rangle}{\langle x, Mx\rangle}.$$

(a) Show that the minimum problem has a nonzero solution f. (b) Show that a solution f of the minimum problem satisfies the equation

$$Hf = bMf$$

where the scalar b is the value of the minimum problem.
Solution:

Q10 [10 points] Given a dataset $\{x_i\}_{i=1}^n$ with $x_i \in \mathbb{R}^D$. Recall that PCA is to find orthonormal basis vectors $w_1, \cdots, w_d$ to minimize the reconstruction error:

$$\begin{aligned} \min_{w_1,\cdots,w_d \in \mathbb{R}^D} \quad & \frac{1}{n}\sum_{i=1}^n \|x_i - \hat{x}_i\|^2 \\ \text{s.t.} \quad & \hat{x}_i = \sum_{j=1}^d \alpha_{ij} w_j, \alpha_{ij} = w_j^\top x_i, \forall i, j \\ & w_i^\top w_j = 0, w_i^\top w_i = 1, \forall i \neq j. \end{aligned}$$

Prove that this optimization problem is equivalent to the low-rank optimization problem:

$$\begin{aligned} \min_{W \in \mathbb{R}^{D \times d}} \quad & \|X - XWW^\top\|_F^2 \\ \text{s.t.} \quad & W^\top W = I \end{aligned}$$

where $X \in \mathbb{R}^{D \times n}$ has x_i in the i-th row, and $\|A\|_F^2 = \sum_{ij} A_{ij}^2$ is the Frobenius norm of a matrix A.
Solution:

3 Vector Calculus

Q11 [10 points] Read Chapter 5 Vector Calculus of the book *Mathematics for Machine Learning* by Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong on its website https://mml-book.github.io.
(1) Let $f : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^m \to \mathbb{R}^k$ be two functions and suppose they are smooth enough so that their gradients are formed by the partial derivatives. Assume that we have proved the chain rule for any i, j:

$$\frac{\partial g_i(f(x))}{\partial x_j} = \frac{\partial g_i(f(x))}{\partial f(x)} \cdot \frac{\partial f(x)}{\partial x_j}.$$

Use these to prove the chain rule:

$$\frac{\partial g(f(x))}{\partial x} = \frac{\partial g(f)}{\partial f} \cdot \frac{\partial f(x)}{\partial x}.$$

(2) Calculate to show that

$$\begin{aligned}
&\frac{\partial a^\top x}{\partial x} = a^\top \\
&\frac{\partial x^\top Bx}{\partial x} = x^\top (B + B^\top) \\
&\frac{\partial}{\partial s}(a - As)^\top W(a - As) = -2(a - As)^\top WA \quad \text{for symmetric } W
\end{aligned}$$

Solution: