EE 516: Mathematical Foundations of Machine Learning **Winter 2023**

Homework 1
Due: January 20, 2023, 11:59PM PT

Student Name: *Instructor Name: John Lipor*

Note: For all questions asked in this course, unless the question is simply asking about definitions, please justify all answers. Stating the correct answer without justification will earn no credit. For example, if you write down a string of equations, each successive equation should be accompanied by some justification such as "by orthogonality" or "which is a property of the determinant." **Include** your code whenever you are asked to write some. This is easily done by attaching the pdf using the `pdfpages` package (see link on course website).

Problem 1 (5 pts, 5 pts, 1 pt)

Let A be an $m \times n$ matrix defined such that $A_{i,j} = i$.

(a) Express A mathematically as the outer product of two appropriately-defined vectors.

(b) Write (on paper) a one-line MATLAB or PYTHON expression for expressing A when $m = 3$ and $n = 2$.

(c) Test your expression and include your code and the output.

Problem 2 (10 pts)

Using the property that $\det(A) = \det(A^T)$ and that $A, B \in \mathbb{R}^{n\times n} \implies \det(AB) = \det(A)\det(B)$, solve the following problem:

If A is orthogonal, what possible values are there for $\det(A)$?

Problem 3 (10 pts, 10 pts, 5 pts, 5 pts)

(a) For $x, y \in \mathbb{R}^n$, prove that $\det(I - xy^T) = 1 - y^T x$. Use the properties below, and **do not** use Sylvester's determinant identity/the matrix determinant lemma.

(b) Express $\det(\lambda I - xy^T)$ in terms of λ, $y^T x$, and n.

(c) What are the solutions of the equation $\det(\lambda I - xy^T) = 0$? Note that these are exactly the eigenvalues of the matrix xy^T.

(d) When are the eigenvalues of the matrix xy^T all equal to 0 even when the matrix is not equal to zero?

Hint: Use the following properties.

- If $A, B \in \mathbb{R}^{n\times n}$ then $\det(AB) = \det(A)\det(B)$.
- If $A \in \mathbb{R}^{n\times n}$ is invertible then $\det(A^{-1}) = \frac{1}{\det(A)}$.
- If $A \in \mathbb{R}^{n\times n}$ is invertible and $D \in \mathbb{R}^{m\times m}$, then $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B)$.
- If $A \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{m\times m}$ is invertible, then $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(D)\det(A - BD^{-1}C)$.

Problem 4 (5 pts, 10 pts, 7 pts, 3 pts)

For $A \in \mathbb{R}^{n\times n}$, the *trace* of A, denoted by $\mathrm{tr}(A)$, is defined as the sum of its diagonal elements: $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$.

(a) Show that the trace is a linear function, i.e., if $A, B \in \mathbb{R}^{n\times n}$ and $\alpha, \beta \in \mathbb{R}$, then $\mathrm{tr}(\alpha A + \beta B) = \alpha\mathrm{tr}(A) + \beta\mathrm{tr}(B)$.

(b) Show that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ when both AB and BA are square, even though in general $AB \neq BA$.

(c) Let $S \in \mathbb{R}^{n\times n}$ be *skew-symmetric*, i.e., $S^T = -S$. Show that $\mathrm{tr}(S) = 0$.

(d) Either prove the converse of part (c)—that all matrices with trace zero are skew symmetric—or provide a counterexample.

Problem 5 (7 pts)

Let $U_1, U_2, \ldots, U_k \in \mathbb{R}^{n\times n}$ be unitary matrices. Show that the product $U_1 U_2 \ldots U_k$ is a unitary matrix (formal proof not required—simply state your reasoning).

Problem 6 (3 pts, 3 pts, 1 pt)

(a) Determine the eigenvalues (show your hand calculations) of $A = \begin{bmatrix} 6 & 16 \\ -1 & -4 \end{bmatrix}$.

(b) For the matrix A above, compute the determinant and trace of A. Compute the product of the eigenvalues of A and the sum of the eigenvalues. How do these compare to the determinant and trace?

(c) Check your answers using the `eig` or `linalg.eig` command and include your output result.

Problem 7 (5 pts)

Rewrite the expression

$$y = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j$$

as a matrix-vector operation in terms of the $n \times n$ matrix A and the $n \times 1$ vector x whose ith entry is x_i. Check your answer in MATLAB or PYTHON by trying some random examples (no need to include code or output).

Problem 8 (2 pts, 5 pts, 2 pts)

Discrete time convolution (a DSP term) of an input signal $x[n]$ and a filter $h[n]$ is defined in textbooks as:

$$y = h * x = x * h \implies y[m] = \sum_{k=-\infty}^{\infty} h[k]x[m-k] = \sum_{n=-\infty}^{\infty} x[n]h[m-n],$$

where $*$ denotes convolution, not ordinary multiplication. In practice, we often use finite-length signals and filters, and not all textbooks define this case clearly. The convolution of an input signal $(x[0], \ldots, x[N-1])$ with a finite impulse response (FIR) filter $(h[0], \ldots, h[K-1])$ becomes:

$$y[m] = \sum_{k=0}^{\min(K-1,m)} h[k]x[m-k] = \sum_{n=0}^{\min(N-1,m)} x[n]h[m-n], \qquad m = 0, \ldots, M-1.$$

This is exactly the type of convolution used in "convolutional neural networks" that are a hot topic in machine learning.

(a) Determine M, the length of the (possibly) non-zero part of y, in terms of N and K.

(b) We will use the notation y, h, x to denote the finite vectors of length M, K and N, respectively. Determine (write down) the matrix H such that $y = h * x = Hx$.

(c) Implement a function called `convolution` that takes h, x as inputs and outputs $y = h * x$ as a matrix operation (WITHOUT calling `conv` or a similar function). Compare the output of your function to the built-in `conv` or `np.convolve` function. Turn in your code demonstrating that your code works (or doesn't).

Problem 9 (5 pts)

The Kronecker product of two matrices $A \in \mathbb{R}^{m_1 \times n_1}, B \in \mathbb{R}^{m_2 \times n_2}$ is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n_1}B \\ & & \ddots & \\ a_{m_1 1}B & a_{m_1 2}B & \dots & a_{m_1 n_1}B \end{bmatrix} \in \mathbb{R}^{(m_1 m_2) \times (n_1 n_2)}.$$

Let $\text{vec}(C)$ be the "vectorize" operation that returns as its output a $mn \times 1$ vector formed by stacking the n columns of $C \in \mathbb{R}^{m \times n}$ on top of each other.

For arbitrary vectors $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, prove that

$$\text{vec}(xy^T) = y \otimes x \in \mathbb{R}^{mn}.$$

Verify in MATLAB using $\text{vec}(C) =$ `C(:)` or PYTHON using `reshape` or `flatten`.

Problem 10 (5 pts)

Note: The purpose of this problem is to help you gauge how much you learn throughout this course and to provide motivation for the time we will spend on heavily theoretical topics. You are not expected to understand the below paper in great detail. You may benefit from guides on reading a research paper, such as the ones here and here.

Spend at most one hour reading the paper, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation" by Aharon, Elad, and Bruckstein. Comment on your ability to understand the following:

- the problem the authors are trying to solve
- why this problem is interesting or important
- the mathematical formulation of the problem
- the mathematical/algorithmic solution to the problem
- why the mathematical/algorithmic solution works.

Do not summarize the problem, why it is important, etc. Instead, comment on *your ability to understand these things* given your current background. The goal of this course is to provide you with the tools needed to understand papers like the one above in enough detail to implement them and explain them to others.