Mathematics for AI
Homework 3

Read these first:

i To achieve the full score, you need to write your solutions using LaTeX. If you choose to write your solutions on paper or in a word processing software (e.g., MS Word, LibreOffice), you can receive up to 90% of the score.

ii If writing on paper, you must use a scanner or a camera scanning app (e.g., CamScanner) to scan the document and submit it as a *single* PDF file. Ensure your answers are written neatly, organized, and legible on paper.

iii When using LaTeX, follow one of these two conventions:

(a) Represent scalars with italic letters (a, A), vectors with bold lowercase letters ($\mathbf{a}$, using `\mathbf{a}`), and matrices with bold uppercase letters ($\mathbf{A}$, using `\mathbf{A}`), or

(b) Represent scalars with italic letters (a, A), vectors with bold letters ($\mathbf{a}$, $\mathbf{A}$), and matrices with typewriter uppercase letters ($\mathtt{A}$, using `\mathtt{A}`).

iv Your LaTeXdocument must include a *title*, a *date*, and your name as the *author*.

v If writing on paper, submit a *single* PDF file; do not send multiple image files.

vi If using LaTeX, submit the *.tex* source file (along with any other required source files) in addition to the PDF file.

Here is a short tutorial on LaTeX: `https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes`

Questions

For each questions, you may use the results of the previous questions (but not the following questions).

Positive Definite Matrices

For all question in this section, by *positive definite* we mean *symmetric positive definite*.

1. Prove that a symmetric matrix is positive definite if and only if all its eigenvalues are positive. (Remember from the class that the eigen-decomposition of a symmetric matrix is in the form of $\mathtt{A} = \mathtt{V}\Lambda\mathtt{V}^{-1}\mathtt{V}\Lambda\mathtt{V}^T$.)

2. Show that the diagonal elements of a positive definite matrix are all positive definite.

3. Remember from the class that an operation $\langle\cdot,\cdot\rangle : \mathcal{V}\times\mathcal{V}\to\mathbb{R}$ defined on a vector space $\mathcal{V}$ is an *inner product* if

   (a) $\langle\mathbf{u},\mathbf{u}\rangle \geq 0$ for all $\mathbf{u}\in\mathcal{V}$,

   (b) $\langle\mathbf{u},\mathbf{u}\rangle = 0$ if and only if $\mathbf{u}=\mathbf{0}$,

   (c) $\langle\mathbf{u},\mathbf{v}\rangle = \langle\mathbf{v},\mathbf{u}\rangle$ for all $\mathbf{u},\mathbf{v}\in\mathcal{V}$,

   (d) $\langle\alpha\mathbf{u}+\beta\mathbf{v},\mathbf{w}\rangle = \alpha\langle\mathbf{u},\mathbf{w}\rangle+\beta\langle\mathbf{v},\mathbf{w}\rangle$ for all $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ and $\alpha,\beta\in\mathbb{R}$.

   Let $\mathtt{A}\in\mathbb{R}^{n\times n}$ be any *positive definite* matrix. Show that the operation $\langle\cdot,\cdot\rangle_{\mathtt{A}} : \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ defined by

   $$\langle\mathbf{u},\mathbf{v}\rangle_{\mathtt{A}} = \mathbf{u}^T\mathtt{A}\mathbf{v}$$

   is indeed an inner product.

Singular Value Decomposition

4. Let $\mathtt{A}$ be a nonsingular square matrix and $A = \mathtt{U}\Sigma\mathtt{V}^T$ be its (full) SVD. Prove that $\det(\mathtt{U})\det(\mathtt{V}) = \mathrm{sign}(\det(\mathtt{A}))$, that is $\det(\mathtt{U})\det(\mathtt{V}) = 1$ if $det(\mathtt{A}) > 0$ and $\det(\mathtt{U})\det(\mathtt{V}) = 1$ if $det(\mathtt{A}) < 0$.

5. Show that for a symmetric positive definite matrix the eigenvalue decomposition $\mathtt{A} = \mathtt{V}\Lambda\mathtt{V}^{-1} = \mathtt{V}\Lambda\mathtt{V}^T$ is the same as the singular value decomposition.

6. Find a way to obtain the SVD of a symmetric matrix from its eigenvalue decomposition $\mathtt{A} = \mathtt{V}\Lambda\mathtt{V}^T$. Notice that the diagonal elements of Λ might be negative.

7. Consider a matrix $\mathtt{A}\in\mathbb{R}^{m\times n}$ and two orthogonal matrices $\mathtt{P}\in\mathbb{R}^{m\times m}$ and $\mathtt{Q}\in\mathbb{R}^{n\times n}$. Show that the singular values of $\mathtt{PAQ}$ is the same as the singular values of $\mathtt{A}$.

Matrix inner product

8. Perhaps the simplest way to define an inner product between a pair of matrices $\mathtt{A}, \mathtt{B} \in \mathbb{R}^{m \times n}$ is $\langle \mathtt{A}, \mathtt{B} \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}$. This is the same as vectorizing the matrices and taking their dot product, and is sometimes called the *Frobenius Inner Product*.

   (a) Prove that real matrices $\langle \mathtt{A}, \mathtt{B} \rangle = \mathrm{trace}(\mathtt{A}^T \mathtt{B}) = \mathrm{trace}(\mathtt{B}^T \mathtt{A}) = \mathrm{trace}(\mathtt{A}\mathtt{B}^T)$, where $\mathrm{trace}(\mathtt{S}) = \sum_i S_{ii}$ gives the sum of the diagonal elements of a square matrix $\mathtt{S}$.

   (b) Prove that $\langle \mathtt{AB}, \mathtt{C} \rangle = \langle \mathtt{B}, \mathtt{A}^T \mathtt{C} \rangle = \langle \mathtt{A}, \mathtt{C}\mathtt{B}^T \rangle$ Hint: $(\mathtt{AB})^T = \mathtt{B}^T \mathtt{A}^T$.

   Note: Same results hold for complex matrices by replacing the transpose operation with conjugate transpose: $\langle \mathtt{AB}, \mathtt{C} \rangle = \langle \mathtt{B}, \mathtt{A}^* \mathtt{C} \rangle = \langle \mathtt{A}, \mathtt{C}\mathtt{B}^* \rangle$.

Matrix Norms

9. Show that the squred Frobenius norm is the same as the Frobenius inner product of a matrix by itself, that is $\|\mathtt{A}\|_F^2 = \langle \mathtt{A}, \mathtt{A} \rangle$.

10. A matrix norm is called *Unitarily Invariant* if $\|\mathtt{A}\| = \|\mathtt{UAV}\|$ for any orthogonal matrices $\mathtt{U}$ and $\mathtt{V}$ of compatible size. Using the above and the properties of matrix inner product prove that the Frobenius norm is unitarily invariant. Notice that for orthogonal matrices we have $\mathtt{U}^T \mathtt{U} = \mathtt{U}\mathtt{U}^T = \mathtt{I}$. (A more general definition that also works for complex matrices is when $\mathtt{U}$ and $\mathtt{V}$ are unitary, that is $\mathtt{U}^* \mathtt{U} = \mathtt{U}\mathtt{U}^* = \mathtt{I}$).

11. Use Question 7 to prove that the *spectral norm* and *nuclear norm* are also unitarily invariant.

Adjoint

Consider two inner product spaces $\mathcal{U}$ and $\mathcal{V}$. A mapping $f^* \colon \mathcal{V} \to \mathcal{U}$ is called the *adjoint* of the linear map $f \colon \mathcal{U} \to \mathcal{V}$ if

$$\langle \mathbf{y}, f(\mathbf{x}) \rangle = \langle f^*(\mathbf{y}), \mathbf{x} \rangle \,,$$

for all $\mathbf{x} \in \mathcal{U}$ and $\mathbf{y} \in \mathcal{V}$.

12. Show that for the linear map $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = \mathtt{A}\mathbf{x}$ with $\mathtt{A} \in \mathbb{R}^{m \times n}$ the adjoint is defined by $f^*(\mathbf{y}) = \mathtt{A}^T \mathbf{y}$.

13. Show that the $\mathrm{diag}(\cdot)$ and $\mathrm{Diag}(\cdot)$ operations defined below are adjoints of each other (with respect to the ordinary dot product defined in previous assignments).

    The operations diag() and $\mathrm{Diag}(\cdot)$ are defined as follows:

    - $\mathrm{diag}(\mathtt{A})$ creates a vector $\in \mathbb{R}^n$ from the diagonal elements of the matrix $\mathtt{A} \in \mathbb{R}^{m \times n}$, and

- Diag($\mathbf{x}$) creates an $n \times n$ diagonal matrix whose diagonal elements are the entries of $\mathbf{x} \in \mathbb{R}^n$.

Notice that both these operations are linear.