Advanced Machine Learning - HW1

Yunting Chiu

18 September, 2021

Exercise 1 (5 points)

Show that the eigenvalues of a symmetric positive matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ are all positive. (Hint: Recall that the eigenvalues of a symmetric matrix are real.)

If A is a real n-by-n symmetric matrix, then $A = A^T$.
Let λ be a (real) eigenvalue of A and V be the corresponding real eigenvector. Now, we have:

$$AV = \lambda V \tag{1.1}$$

Then we multiply by V^T on both sides (V^T is a transpose of V), the following equation would be:
Note: The outcome of $V^T * V$ would be 1-by-1 vector.

$$V^T AV = \lambda V^T V = \lambda ||V||^2 \tag{1.2}$$

Because A is positive definite so the left hand side is positive, and V is a non-zero eigenvector. Also, the length $||V||^2$ must be a positive, we can derive the eigenvalues of a symmetric positive matrix λ is positive. It follows that every eigenvalue λ of A is real.

$$\lambda = \frac{V^T AV}{||V||^2} > 0 \tag{1.3}$$

Exercise 2 (5 points)

Show that the determinant of an orthogonal matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is ± 1. Next, for the rotation matrix $\mathbf{R} \in \mathbb{R}^{2 \times 2}$, show that the determinant equals 1.

Prove A:

- Let I_n denotes a n-by-n identity matrix, and A a n-by-n orthogonal matrix. According to the definition of an orthogonal matrix, we know that: $AA^T = I = A^T A$.
- Also, we reviewed this from the class: $det(A \cdot B) = det(A) \cdot det(B)$ and $det(A) = det(A^T)$.
- Now we have $\det(I) = 1 = \det(AA^T) = \det(A)\det(A^T) = \det(A)\det(A) = [\det(A)]^2$
- $[\det(A)]^2$ so $\det(A) = \sqrt{1} = \pm 1$

Prove R:

By definition, a rotation matrix in n-dimensions is a n-by-n special orthogonal matrix. Therefore, it is a subset of the orthogonal group. Following the same proof above, we can know that the determinant of a rotation matrix is ± 1.

Exercise 3 (5 points)

Show that $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ have the same eigenvalues.

We know that λ is a eigenvalue of A, and V is a corresponding real eigenvector. To be precise, A is a n-by-n matrix, V is a non-zero n-by-1 vector, and λ is a scalar (real or complex number).

$$AV = \lambda V \tag{3.1}$$

Let's use $A^T A$ be a matrix as a whole to replace A, the equation would be:

$$A^T AV = \lambda V \tag{3.2}$$

Then, we multiply by A on both sides, the equation would be:

$$AA^T AV = \lambda AV \tag{3.3}$$

Because A is a n-by-n matrix, V is a non-zero n-by-1 vector so the result of $A * V$ is a n-by-1 value. In equation 3.2, $A^T AV$ is equal to λV; In equation 3.3, $AA^T AV$ is equal to λAV. Therefore, we can find out there is a same eigenvalue λ between $A^T A$ and AA^T.

Exercise 4 (5 points)

Show that the negative entropy function $f(x) = x \log x$ is convex for all $x > 0$. (Hint: If we know that a function is twice differentiable, that is, the Hessian exists for all values in the domain of x, then the function is convex if and only if the Hessian is positive semi-definite.)

The Hessian matrix, also known as the Hessian, is a square matrix of **second-order** partial derivatives of a scalar-valued function, or scalar field.

The negative entropy function is:

$$f(x) = -x \; logx \tag{4.1}$$

We take the first derivative:

$$f'(x) = -1 + log(x) \tag{4.2}$$

We take the second derivative:

$$f''(x) = \frac{1}{x} \tag{4.3}$$

Because $x > 0$, then $\frac{1}{x} > 0$. The second-order condition can only be used for twice-differentiable functions. If $\nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}) > \mathbf{0}$, we can conclude that it is strictly convex.

Exercise 5 (5 points)

Show that the least-squares objective function $f(\mathbf{x}) = ||\mathbf{Ax} - \mathbf{b}||_2^2$ is convex for any invertible matrix $\mathbf{A}$.

Where $A \in R^{k \times n}$, $b \in R^k$, $x \in R^n$, and $|| \cdot ||_2$ denotes the Euclidean norm.

For the least-square objective function, we take the first derivative:

$$\nabla f(x) = 2A^T(Ax - b) \tag{5.1}$$

Then, we take the second derivative, and it would always convex.

$$\nabla^2 f(x) = 2A^T A \tag{5.2}$$

The Hessian is $A^T A$, which is positive semi-definite. It follows that f is strict convexity.

Exercise 6 (Extra credit)(5points)

Show that the spectral norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is its largest singular value σ_1. This is

$$\max_{\mathbf{x}} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||} = \sigma_1$$

The norm $||A||_2$ is always called the *spectral norm* so the spectral norm of a matrix $A \in R^{m \times n}$ is denoted as

$$||A||_2 = \max_{x \neq 0} \frac{||Ax||_2}{||x||_2} = \max_{u \in R^n : ||u||_2 = 1} ||Au||_2 \tag{6.1}$$

To proof, we first square the above equation:

$$||A||_2^2 = \max_{x \neq 0} \frac{||Ax||_2^2}{||x||_2^2} = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x} = \sqrt{\lambda max(A^T A)} = \lambda_1 (A^T A) = \sigma_1 (A)^2 \tag{6.2}$$

Where λ is the eigenvalues of $(A^T A)$ so we have

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^* A)} = \sigma_{\max}(A) \tag{6.3}$$

Because singular value is defined as the square roots of non-negative eigenvalues of the self-adjoint operator $A^T A$.

References

- https://yutsumura.com/positive-definite-real-symmetric-matrix-and-its-eigenvalues/
- https://lpsa.swarthmore.edu/MtrxVibe/EigMat/MatrixEigen.html
- https://www.wikihow.com/Find-Eigenvalues-and-Eigenvectors
- https://www.brainm.com/software/pubs/math/Rotation_matrix.pdf
- https://math.stackexchange.com/questions/483339/proof-of-convexity-of-linear-least-squares
- http://web.cs.ucla.edu/~patricia.xiao/files/ECE_236B_Course_Notes.pdf
- https://www.cis.upenn.edu/~cis515/cis515-12-sl4.pdf
- https://math.stackexchange.com/questions/586663/why-does-the-spectral-norm-equal-the-largest-singular-value
- https://math.stackexchange.com/questions/586663/why-does-the-spectral-norm-equal-the-largest-singular-value
- https://en.wikipedia.org/wiki/Singular_value