CS229 Assignment 0

A Khan

May 2025

This document contains my solutions to the homework problems from the Stanford Machine Learning course CS229 (2018 set).

1 Solution to Problem 1: Gradients and Hessians

(a) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and b is a vector.
The gradient is:

$$\nabla f(x) = Ax + b$$

(since $x^T = x$ for vectors, and $b^T = b$ for column vectors).

(b) Let $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable.
The gradient is:

$$\nabla f(x) = g'\left(h(x)\right) \cdot \nabla h(x)$$

(c) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, as before.
The gradient is:

$$\nabla f(x) = Ax + b$$

The Hessian is:

$$\nabla^2 f(x) = A$$

(d) Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and a is a vector.

The gradient is:

$$\nabla f(x) = g'\left(a^T x\right) \cdot a$$

The Hessian is:

$$\nabla^2 f(x) = g''\left(a^T x\right) \cdot aa^T$$

2 Solution to Problem 2: Positive Definite Matrices

(a) Let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^T$ is positive semi-definite.

Solution:
$A = zz^T$ is an $n \times n$ matrix. For any $x \in \mathbb{R}^n$, consider:

$$x^T Ax = x^T(zz^T)x = (x^T z)(z^T x) = (z^T x)^2 \geq 0$$

Since $z^T x$ is a real number, its square is always non-negative. Therefore, A is positive semi-definite.

(b) Let $z \in \mathbb{R}^n$ be a *non-zero* n-vector. Let $A = zz^T$. What is the null-space of A? What is the rank of A?

Solution:
The null-space of A consists of all $x \in \mathbb{R}^n$ such that $Ax = 0$:

$$zz^T x = 0 \implies z^T x = 0$$

So, the null-space is the set of all vectors orthogonal (perpendicular) to z.
The rank of A is 1, because A can be written as the outer product of z with itself, and thus all columns of A are linearly dependent (multiples of z).

(c) Let $A \in \mathbb{R}^{n\times n}$ be positive semi-definite and $B \in \mathbb{R}^{m\times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so, prove it. If not, give a counterexample with explicit A, B.

Let $x \in \mathbb{R}^m$ be any vector. We consider:

$$x^T BAB^T x = (B^T x)^T A(B^T x)$$

Let $y = B^T x \in \mathbb{R}^n$. Then,

$$x^T BAB^T x = y^T Ay$$

Since A is positive semi-definite, we know $y^T Ay \geq 0$ for any $y \in \mathbb{R}^n$. Therefore,

$$x^T BAB^T x \geq 0 \quad \forall x \in \mathbb{R}^m$$

So BAB^T is always positive semi-definite if A is positive semi-definite.

Expanded Calculation (element-wise): Suppose B and A are given as:

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$$B^T = \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{m1} \\ b_{12} & b_{22} & \cdots & b_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1n} & b_{2n} & \cdots & b_{mn} \end{bmatrix}$$

Multiplying out BAB^T gives an $m \times m$ matrix, where the (i, j)-th entry is:

$$[BAB^T]_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} b_{ik} a_{kl} b_{jl}$$

If A is positive semi-definite, and b_{ik}, b_{jl} are real numbers, the sum above consists of positive or zero values (if A is diagonal with non-negative elements, each term is a squared element times a non-negative value).

Conclusion: Each element is a sum of such products, so BAB^T is positive semi-definite.

3 Problem 3: Eigenvectors, Eigenvalues, and the Spectral Theorem

(a) Suppose that the matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable, that is, $A = T\Lambda T^{-1}$ for some invertible matrix $T \in \mathbb{R}^{n \times n}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Let $t^{(i)}$ denote the i-th column of T. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalue/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$.

Solution:

Given $A = T\Lambda T^{-1}$ and $T = [t^{(1)} \; t^{(2)} \; \cdots \; t^{(n)}]$, consider $t^{(i)}$ (the i-th column of T):

$$At^{(i)} = T\Lambda T^{-1} t^{(i)}$$

But $T^{-1} t^{(i)} = e^{(i)}$, the i-th standard basis vector:

$$At^{(i)} = T\Lambda e^{(i)}$$

Since Λ is diagonal,

$$\Lambda e^{(i)} = \lambda_i e^{(i)}$$

So,

$$At^{(i)} = T(\lambda_i e^{(i)}) = \lambda_i T e^{(i)} = \lambda_i t^{(i)}$$

Conclusion: $t^{(i)}$ is an eigenvector of A with eigenvalue λ_i.

(b) Let A be symmetric, so $A = U\Lambda U^T$ for an orthogonal matrix U, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Show that if $U = [u^{(1)} \cdots u^{(n)}]$ is orthogonal, then $u^{(i)}$ is an eigenvector of A and $Au^{(i)} = \lambda_i u^{(i)}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

Solution:

Since U is orthogonal, $U^T = U^{-1}$. Let $u^{(i)}$ be the i-th column of U.

$$Au^{(i)} = U\Lambda U^T u^{(i)}$$

But $U^T u^{(i)} = e^{(i)}$, so

$$Au^{(i)} = U\Lambda e^{(i)}$$

Since $\Lambda e^{(i)} = \lambda_i e^{(i)}$,

$$Au^{(i)} = U(\lambda_i e^{(i)}) = \lambda_i U e^{(i)} = \lambda_i u^{(i)}$$

Conclusion: $u^{(i)}$ is an eigenvector of A with eigenvalue λ_i.

(c) Show that if A is PSD, then $\lambda_i(A) \geq 0$ for each i.

Solution:

Recall that A is positive semi-definite (PSD) if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$.

Let $u^{(i)}$ be an eigenvector of A with eigenvalue λ_i, and assume $u^{(i)}$ is normalized ($\|u^{(i)}\| = 1$).

Set $x = u^{(i)}$ in the PSD condition:

$$(u^{(i)})^T Au^{(i)} \geq 0$$

By the eigenvalue equation,

$$(u^{(i)})^T Au^{(i)} = (u^{(i)})^T(\lambda_i u^{(i)}) = \lambda_i (u^{(i)})^T u^{(i)} = \lambda_i \|u^{(i)}\|^2 = \lambda_i \cdot 1 = \lambda_i$$

Therefore,

$$\lambda_i \geq 0$$

This shows that all eigenvalues of a PSD matrix are non-negative.

4 Conclusion

Through this assignment, I gained a deeper understanding of some fundamental concepts in linear algebra and their relevance to machine learning. I explored the structure and properties of quadratic forms, gradients, and Hessians, which are foundational for optimization problems. The analysis of positive semi-definite (PSD) matrices improved my grasp of matrix theory, including how rank, null space, and definiteness are determined by matrix structure.

Working with eigenvalues and eigenvectors reinforced the importance of spectral decomposition and diagonalization in simplifying complex linear transformations. I saw how these concepts are used to analyze the behavior of matrices, especially symmetric and PSD matrices, which frequently appear in machine learning algorithms.

Overall, this assignment helped clarify the theoretical underpinnings of algorithms used in data science and machine learning, and strengthened my ability to reason about matrices, vector spaces, and their associated properties.