CS229 problem set 0

James Chuang

January 19, 2017

Contents

1. Gradients and Hessians . . . 1
2. Positive definite matrices . . . 2
3. Eigenvectors, eigenvalues, and the spectral theorem . . . 2

My solutions to the CS229 problem set #0 from fall 2016. The official solutions are available there as well.

1. Gradients and Hessians

Recall that a matrix $A \in \mathbb{R}^{n\times n}$ is *symmetric* if $A^T = A$, i.e. $A_{ij} = A_{ji} \,\forall\, i, j$. Also recall the gradient $\nabla_x f(x)$ of a function $f : \mathbb{R}^n \mapsto \mathbb{R}$, which is the n-vector of partial derivatives

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \quad \text{where } x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The Hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R} \mapsto \mathbb{R}$ is the $n \times n$ symmetric matrix of second partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}.$$

a. Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla_x f(x)$?

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Ax + b^T x \\ \nabla_x f(x) &= \frac{1}{2}2Ax + b && \text{see linalg review 4.3} \\ \nabla_x f(x) &= Ax + b \end{aligned}$$

b. Let $f(x) = g(h(x))$, where $g : \mathbb{R} \mapsto \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable. What is $\nabla_x f(x)$?

$$\begin{aligned} f(x) &= g(h(x)) \\ \nabla_x f(x) &= g'(h(x))\nabla_x f(x) && \text{by the chain rule} \end{aligned}$$

c. Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is symmetric and $b \in \mathbb{R}^n$ is a vector. What is $\nabla_x^2 f(x)$?

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Ax + b^T x \\ \nabla_x^2 f(x) &= \frac{1}{2}2A && \text{see linalg review 4.3} \\ \nabla_x^2 f(x) &= A \end{aligned}$$

d. Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \mapsto \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$? (*Hint*: your expression for $\nabla^2 f(x)$ may have as few as 11 symbols, including $'$ and parentheses.)

$$\begin{aligned} f(x) &= g(a^T x) \\ \nabla_x f(x) &= g'(a^T x) a && \text{chain rule} \\ \nabla_x^2 f(x) &= g''(a^T x) a a^T && \text{chain rule again} \end{aligned}$$

2. Positive definite matrices

A matrix $A \in \mathbb{R}^{n \times n}$ is *positive semi-definite* (PSD), denoted $A \succeq 0$, if $A = A^T$ and $x^T A x \geq 0, \forall x \in \mathbb{R}^n$. A matrix A is *positive definite* (PD), denoted $A \succ 0$, if $A = A^T$ and $x^T A X > 0, \forall x \neq 0$, i.e., all non-zero vectors x. The simplest example of a positive definite matrix is the identity matrix I, which satisfies $x^T I x = \|x\|_2^2 = \sum_{i=1}^n x_i^2$.

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

$$\begin{aligned} x^T A x &= x^T z z^T x \\ &= (x^T z)^2 \geq 0 && \text{by commutativity of dot product} \end{aligned}$$

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

$$\begin{aligned} Ax &= 0 \\ zz^T x &= 0 && \text{since } z \text{ is non-zero,} \\ z^T x &= 0 \end{aligned}$$

I.e., x is in the nullspace of A if it is orthogonal to z, so the nullity (dimension of the nullspace) is $n - 1$ (where the one missing dimension is the one occupied by z).

The rank-nullity theorem states (for $A \in \mathbb{R}^{m \times n}$):

$$\begin{aligned} \text{rank}(A) + \text{nul}(A) &= n \\ \text{rank}(A) + (n - 1) &= n \\ \text{rank}(A) &= 1 \end{aligned}$$

c. Let $A \in \mathbb{R}^n$ be positive semidefinite 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 note, give a counterexample with explicit A, B.

BAB^T is PSD if $x^T BAB^T x \geq 0 \; \forall \; x \in \mathbb{R}^m$:

$$\begin{aligned} & x^T B A B^T x \\ &= \left(x^T B\right) A \left(B^T x\right) \\ &= \left(B^T x\right)^T A \left(B^T x\right) \\ &= z^T A z \geq 0 && \text{where } z = B^T x \in \mathbb{R}^n, \text{ and since A is PSD} \end{aligned}$$

3. Eigenvectors, eigenvalues, and the spectral theorem

The eigenvalues of an $n \times n$ matrix $A \in \mathbb{R}^{n \times n}$ are the roots of the characteristic polynomial $p_A(\lambda) = \det(\lambda I - A)$, which may (in general) be complex. They are also defined as the values $\lambda \in \mathbb{C}$ for which there exists a vector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$. We call such a pair (x, λ) an *eigenvector, eigenvalue* pair. In this question, we use the notation diag$(\lambda_1, \dots, \lambda_n)$ to denote the diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$, i.e.,

$$\text{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

a. Suppose that the matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable, i.e. $A = T \Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n \times n}$, where $\Lambda =$ diag$(\lambda_1, \dots, \lambda_n)$. Use the notation $t^{(i)}$ for the columns of T, so that $T = \begin{bmatrix} t^{(1)} & \dots & t^{(n)} \end{bmatrix}$, where $t^{(i)} \in \mathbb{R}^n$. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalue/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$.

$$\begin{gathered} A = T \Lambda T^{-1} \\ AT = T\Lambda \\ A \begin{bmatrix} | & | & & | \\ t^{(1)} & t^{(2)} & \cdots & t^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ t^{(1)} & t^{(2)} & \cdots & t^{(n)} \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\ \begin{bmatrix} | & | & & | \\ At^{(1)} & At^{(2)} & \cdots & At^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 t^{(1)} & \lambda_2 t^{(2)} & \cdots & \lambda_n t^{(n)} \\ | & | & & | \end{bmatrix} \\ \therefore At^{(i)} = \lambda_i t^{(i)}, \quad i \in \{1, 2, \dots, n\} \end{gathered}$$

A matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if $U^T U = I$. The spectral theorem, perhaps one of the most important theorems in linear algebra, states that if $A \in \mathbb{R}^{n \times n}$ is symmetric, i.e. $A = A^T$, then A is *diagonalizable by a real orthogonal matrix*. I.e., there are a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $U^T A U = \Lambda$, or, equivalently,

$$A = U \Lambda U^T$$

Let $\lambda_i = \lambda_i(A)$ denote the ith eigenvalue of A.

b. Let A be symmetric. Show that if $U = \begin{bmatrix} u^{(1)} & \cdots & u^{(n)} \end{bmatrix}$ is orthogonal, where $u^{(i)} \in \mathbb{R}^n$ and $A = U \Lambda U^T$, then $u^{(i)}$ is an eigenvector of A and $Au^{(i)} = \lambda_i u^{(i)}$, where $\Lambda =$ diag$(\lambda_1, \dots, \lambda_n)$.

$$\begin{gathered} A = U \Lambda U^T \\ AU = U\Lambda \\ A \begin{bmatrix} | & | & & | \\ u^{(1)} & u^{(2)} & \cdots & u^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ u^{(1)} & u^{(2)} & \cdots & u^{(n)} \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\ \begin{bmatrix} | & | & & | \\ Au^{(1)} & Au^{(2)} & \cdots & Au^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 u^{(1)} & \lambda_2 u^{(2)} & \cdots & \lambda_n u^{(n)} \\ | & | & & | \end{bmatrix} \\ \therefore Au^{(i)} = \lambda_i u^{(i)} \quad i \in \{1, 2, \dots, n\} \end{gathered}$$

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

$$
\begin{aligned}
At^{(i)} &= \lambda_i t^{(i)} \\
{t^{(i)}}^T A t^{(i)} &= {t^{(i)}}^T \lambda_i t^{(i)} \\
&= \lambda_i {t^{(i)}}^T t^{(i)} \geq 0 \quad A \text{ is PSD} \\
\lambda_i &\geq \frac{0}{{t^{(i)}}^T t^{(i)}} \\
\lambda_i &\geq 0
\end{aligned}
$$