CS 229: Machine Learning
Problem Set 0

William Ma

July 20, 2017

1 Question 1

1a Part a

Given $f(x) = \frac{1}{2}x^T Ax + b^T x$ where A is a symmetric matrix and and $b \in \mathbb{R}^n$ is a vector, we can calculate $\nabla_x f(x)$ by taking the partial derivative

$$
\begin{aligned}
\frac{\partial f(x)}{\partial x_k} &= \frac{\partial}{\partial x_k}\Big[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \ x_i A_{ij} x_j + \sum_{i=1}^{n} \ b_i x_i\Big] \\
&= \frac{\partial}{\partial x_k}\frac{1}{2}\Big[\sum_{i\neq k}\sum_{j\neq k} A_{ij}x_i x_j + \sum_{i\neq k} A_{ik}x_i x_k + \sum_{j\neq k} A_{kj}x_k x_j + A_{kk}x_k\Big] \\
&\quad + \frac{\partial}{\partial x_k}\sum_{i=1}^{n} \ b_i x_i \\
&= \frac{1}{2}\sum_{i\neq k} A_{ik}x_i + \frac{1}{2}\sum_{j\neq k} A_{kj}x_j + A_{kk}x_k^2 + b_k \\
&= \frac{1}{2}\sum_{i=1}^{n} A_{ik}x_i + \frac{1}{2}\sum_{j=1}^{n} A_{kj}x_j + b_k \\
&= \sum_{i=1}^{n} A_{ik}x_i + b_k
\end{aligned}
$$

Now we can easily see that, if $\nabla_x f(x) = 2Ax + b$

1b Part b

Given that $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable, we can expand $f(x)$ to arrive at the solution

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} g(h(x))$$

By invoking Chain Rule,

$$\frac{\partial f(x)}{\partial x_k} = g'(h(x))\frac{\partial}{\partial x_k}h(x)$$

Combining these back into a vector,

$$\nabla f(x) = \begin{bmatrix} g'(h(x))\frac{\partial}{\partial x_1}h(x) \\ \vdots \\ g'(h(x))\frac{\partial}{\partial x_n}h(x) \end{bmatrix} = g'(h(x))\nabla h(x)$$

1c Part c

Given $f(x) = \frac{1}{2}x^T Ax + b^T x$ where A is a symmetric matrix and and $b \in \mathbb{R}^n$ is a vector, we can calculate the Hessian as follows

$$
\begin{aligned}
\frac{\partial^2 f(x)}{\partial x_k^2} &= \frac{\partial^2}{\partial x_k^2}\Big[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j + \sum_{i=1}^{n} b_i x_i\Big] \\
&= \frac{\partial^2}{\partial x_k^2}\frac{1}{2}\Big[\sum_{i\neq k}\sum_{j\neq k} A_{ij}x_i x_j + \sum_{i\neq k} A_{ik}x_i x_k + \sum_{j\neq k} A_{kj}x_k x_j + A_{kk}x_k^2\Big] \\
&\quad + \frac{\partial^2}{\partial x_k^2}\sum_{i=1}^{n} b_i x_i \\
&= \frac{1}{2}\sum_{i\neq k} A_{ik} + \frac{1}{2}\sum_{j\neq k} A_{kj} + 2A_{kk}x_k \\
&= \frac{1}{2}\sum_{i=1}^{n} A_{ik} + \frac{1}{2}\sum_{j=1}^{n} A_{kj} \\
&= \sum_{i=1}^{n} A_{ik}
\end{aligned}
$$

Thus, the $\nabla^2 f(x) = A$.

1d Part d

Given $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector, we can calculate $\nabla f(x)$ using the result we got from problem 1a and 1b

$$
\begin{aligned}
\nabla f(x) &= g'(a^T x)\ \nabla(a^T x) \\
&= g'(a^T x)a
\end{aligned}
$$

However, for the Hessian, we have to expand, apply Chain rule to each term, then recombine back into a vector.

$$
\begin{aligned}
\frac{\partial^2 f(x)}{\partial x_i \partial x_j} &= \frac{\partial^2}{\partial x_k \partial x_i} x_j g(a^T x) \\
&= g''(a^T x)\frac{\partial}{\partial x_i}\sum_{k=1}^{n} a_k x_k \frac{\partial}{\partial x_j}\sum_{l=1}^{n} a_l x_l \\
&= g''(a^T x)a_i a_j = \begin{bmatrix} g''(a^T x)a_1 a_1 & \dots & g''(a^T x)a_1 a_n \\ \vdots & \ddots & \vdots \\ g''(a^T x)a_n a_1 & \dots & g''(a^T x)a_n a_n \end{bmatrix} \\
&= g''(a^T x)aa^T
\end{aligned}
$$

Thus, $\nabla^2 f(x) = g''(a^T x)aa^T$.

2 Problem 2

2a Part a

Proof. Given $z \in \mathbb{R}^n$ and that $A = zz^T$, $A \in \mathbb{S}_+^{n\times n}$ if $A = A^T$ and $x^T Ax \geq 0$.

$$\begin{aligned} A &= A^T \\ zz^T &= (zz^T)^T \\ zz^T &= (z^T)^T z^T = zz^t \end{aligned}$$

Thus, $A = A^T$.

$$\begin{aligned} x^T Ax &\geq 0 \\ x^T zz^T x &\geq 0 \\ (x^T z)(x^T z)^T &\geq 0 \\ (x^T z)^2 &\geq 0 \end{aligned}$$

Thus, since $A = A^T$ and $x^T Ax \geq 0$, $A \in \mathbb{S}_+^{n\times n}$. $\square$

2b Part b

Given $z \in \mathbb{R}^n$ is a *non-zero* vector and $A = zz^t$, the null-space of A is 1 since, $Ax = 0$ only when x is orthogonal to z, which implies that $z^T x = 0$ as shown.

$$\begin{aligned} Ax &= 0 \\ zz^T x &= 0 \\ z(0) &= 0 \end{aligned}$$

Thus, the null-space is 1. Using the rank-nullity theorem, the rank of A is $n - 1$.

2c Part c

Proof. Given $A \in \mathbb{S}_+^{n\times n}$ and $B \in \mathbb{R}^{m\times n}$ is arbitary,

$$\begin{aligned} BAB^T &= (BAB^T)^T \\ BAB^T &= (B^T)^T A^T B^T \\ BAB^T &= BAB^T \end{aligned}$$

Thus, $BAB^T = (BAB^T)^T$.

$$\begin{aligned} x^T BAB^T x &\geq 0 \\ (x^T B)A(x^T B)^T &\geq 0 \end{aligned}$$

Since $A \in \mathbb{S}_+^{n\times n}$, then $yAy^T \geq 0$. We can simply let $y = x^T B$ for $(x^T B)A(x^T B)^T \geq 0$ to be true. Thus, since $BAB^T = (BAB^T)^T$ and $x^T BAB^T x \geq 0$, $BAB^T \in \mathbb{S}_+^{m\times m}$. $\square$

3 Problem 3

3a part a

Proof. Given that A is diagonalizable, such that $A = T\Lambda T^{-1}$, and $t^{(i)} \in \mathbb{R}^n$ is the i-th column of T,

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

The inverse of a matrix, $M \in \mathbb{R}^{n\times n}$ multiplied by $x^{(i)}$, the i-th column of M, returns always returns a $n \times n$ matrix, N, where

$$N_{jk} = \begin{cases} 1, & \text{if } j = i \text{ and } k = i \\ 0, & \text{otherwise} \end{cases}$$

Thus,

$$\begin{aligned} At^{(i)} &= T\Lambda T^{-1}t^{(i)} = T\lambda_{(i)} \\ &= t^{(i)}\lambda_i = \lambda_i t^{(i)} \end{aligned}$$

Thus, $At^{(i)} = \lambda_i t^{(i)}$ where $(t^{(i)}, \lambda_i)$ are the eigenvector/eigenvalue pair of A. $\square$

3b Part b

Proof. Given that A is symmetric, $A = U\Lambda U^{-1}$, U is orthogonal, and $u^{(i)} \in \mathbb{R}^n$ is the i-th column of T,

$$\begin{aligned} Au^{(i)} &= U\Lambda U^T u^{(i)} \\ &= U\Lambda U^{(-1)}u^{(i)} \end{aligned}$$

We can use the result we got from problem 3a and get that $Au^{(i)} = \lambda_i u^{(i)}$, where $(u^{(i)}, \lambda_i)$ are the eigenvector/eigenvalue pair of A. $\square$

3c Part c

Proof. Given $A \in \mathbb{S}_+^{n\times n}$ and λ_i is an eigenvalue of A,

$$\begin{aligned} x^T Ax &\geq 0 \\ x^T U\Lambda U^T x &\geq 0 \\ (x^T U)\Lambda(x^T U)^T &\geq 0 \end{aligned}$$

Since Λ is a diagonal matrix, $\Lambda \in \mathbb{S}_+^{n\times n}$, which implies that $\lambda_i \geq 0$. $\square$