Ass1

August 5, 2022

1 Assignment1

HengCheng Zhang

u7096187

1.1 Sets

(a) A $\cap$ B = [0,1)

(b) A $\cup$ B = [-1,3]

(c) A $\cap$ C = [-1,0]

(d) A $\cup$ C = [-1,1)

(e) B $\cap$ C = {0}

(f) B $\cup$ C = [-1,3]

(g) A $\times$ B = { (x,y) | a $\in$ [-1,1) , y $\in$ [0,3] }

(h) sup(A B) = 0

(i) inf(A $\cap$ $\mathbb{R}$) = -1

(j) sup($\mathbb{R}$ B) = ∞

1.2 Diagonal matrices

(a)

for $y \in \mathbb{R}^n$ $y_{1} = *{i=1}^n* *D*{1,i} x_i\ y_{2} = *{i=1}^n* *D*{2,i} x_i\ \ y_{n} = *{i=1}^n* *D*{n,i} x_i\ $

Since D is a diagonal matrix, for all $i \neq j, D_{i,j} = 0$

$\therefore y_1 = D_{1,1} \cdot x_1$
$y_2 = D_{2,2} \cdot x_2$
$\vdots$
$y_n = D_{n,n} \cdot x_n$

$\therefore$ By changing n for $D \in \mathbb{R}^{n \times n}$, the cost of computing each element in y won't be affected.

Since $y \in \mathbb{R}^n$, the change of n will affect the size of y which will leads to a linear change in time complexity.

$\therefore$ The cost of cumputing y = Dx is O(n)

(b)

$(D \cdot A)_{ij} = \sum_{k=1}^{n} D_{i,k} \cdot A_{k,j} = D_{i,i} \cdot A_{i,j}$
$(A \cdot D)_{ij} = \sum_{k=1}^{n} A_{i,k} \cdot D_{k,j} = A_{i,j} \cdot D_{j,j}$

$\therefore$ As long as $D_{i,i} \neq D_{j,j}$, $DA \neq AD$

An counter example is for

$D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$

$D \cdot A = \begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \neq A \cdot D = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

(c)

$det(D) = \sum_{\sigma} sgn(\sigma) \prod_{i=1}^{n} D_{i,\sigma(i)}$
$\prod_{i=1}^{n} D_{i,\sigma(i)} = \begin{cases} \prod_{i=1}^{n} D_{i,\sigma(i)} & \text{for } \sigma = (1,2,3,\cdots,n) \\ 0 & \text{for the rest permutation of } \sigma \end{cases}$
Because for all the subdiagonals of D, all the elements are 0. Therefore the product of them will also be 0

(d)

When the diagonal elements of D are all positive.

1.3 Triangular matrices

From 2(c) we know that for a diagonal matrix D $\in \mathbb{R}^{n\times n}$, det(D) = \$ *{i=1}^n D*{i,i} \$

Similarly, for a triangular matrix A, there always has zero element in each product of subdiagonals.

Therefore $det(A) = \prod_{i=1}^{n} D_{i,i}$

For eigenvalues λ, $det(A - \lambda I) = 0 = \prod_{i=1}^{n}(A_{i,i} - \lambda)$

$\therefore \lambda = A_{i,i}$

$\therefore$ the eigenvalues of a triangular matrix equal to its diagonal elements

Since A needs to be non-singular if it is invertible.

Therefore $det(A) \neq 0$.

Thus all eigenvalues needs to be positive for non-negative determinant.

$\therefore$ A needs to be a positive difinite trianglular matrix when its invertible.

1.4 Determinant of a transpose

Let $A \in \mathbb{R}^{n*n}$

For n = 1

$det(A) = A_{1,1} = det(A^T)$

Assume $det(A^T) = det(A)$ when n = k

For n = k+1, let A_{11} denotes the matrix A without the row and column that contains $A_{1,1}$ $det(A) = A_{1,1} \cdot det(A_{11}) - A_{1,2} \cdot det(A_{12}) + \cdots + (-1)^{k+2} \cdot A_{1,k+1} \cdot det(A_{1(k+1)})$

$$\begin{aligned} det(A^T) &= A^T_{1,1} \cdot det(A^T_{11}) - A^T_{2,1} \cdot det(A^T_{21}) + \cdots + (-1)^{k+2} \cdot A^T_{k+1,1} \cdot det(A^T_{(k+1)1}) \\ &= A_{1,1} \cdot det(A^T_{11}) - A_{1,2} \cdot det(A^T_{21}) + \cdots + (-1)^{k+2} \cdot A_{1,k+1} \cdot det(A^T_{(k+1)1}) \end{aligned}$$

Since for $A \in \mathbb{R}^{k*k} det(A) = det(A^T)$

$\therefore$ $det(A^T_{ij}) = det(A_{ji})$ for $i, j = 0, 1, 2, \cdots, k$

$\therefore$ for $n = k + 1, det(A) = det(A^T)$

$\therefore$ for any square matrix A, $det(A) = det(A^T)$

1.5 Cauchy-Schwarz inequality

Since x $\in \mathbb{R}^n$

let y $\in \mathbb{R}^n$ = (1,1,1,...,1), where $\|y\|_2 = \sqrt{n}$

from the Cauchy-Schwarz inequality we can get $|x^T y| \leq \|x\|_2 \|y\|_2$

since $y_i = 1$ for $i = 1,2,3,...,n $

$x^T y = \sum_{i=1}^n x_i \cdot y_i = \sum_{i=1}^n x_i = 1$

$\therefore 1 \leq \|x\|_2 \cdot \sqrt{n}$

$\therefore \|x\|_2 \geq \frac{1}{\sqrt{n}}$

The lower bound doesn't change as the only assumption made is $\sum_{i=1}^n x_i = 1$ and there's no restriction on whether x should be non-negative or not.

1.6 Least-squares

[30]:
```
import numpy as np
```

[31]:
```
A = np.array([[8, 4], [6, 6], [5, 8], [0, 0]])
B = np.array([2, 0, 2, 2]).T
sol = np.linalg.inv(A.T@A)@A.T@B
print(sol)
```

```
[0.14950635 0.06770099]
```

1.7 Gradients

(a) $f(x) = \alpha^T x + b$

$$\begin{aligned}
\nabla_x f(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{\alpha^T(x+h) + b - (\alpha^T x + b)}{h} \\
&= \lim_{h\to 0} \frac{\alpha^T h}{h} \\
&= \alpha^T \\
&= \alpha
\end{aligned}$$

(b) $f(x) = \frac{1}{2}x^T P x + q^T x + r$

$$\begin{aligned}
\nabla_x f(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{2}(x+h)^T P(x+h) + q^T(x+h) + r - (\frac{1}{2}x^T P x + q^T x + r)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{2}(x^T P x + x^T P h + h^T P x + h^T P h) + q^T h - (\frac{1}{2}x^T P x)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{2}x^T P h + \frac{1}{2}h^T P x + \frac{1}{2}h^T P h + q^T h}{h} \\
&= \lim_{h\to 0} (\frac{1}{2}x^T P + \frac{1}{2}h^T P + q^T + \frac{\frac{1}{2}(Px)^T h}{h}) \\
&= \frac{1}{2}P^T x + q^T + \frac{1}{2}Px \\
&= \frac{1}{2}(P + P^T)x + q
\end{aligned}$$

(c) $f(x) = \frac{1}{2}x^T P x$ with $P = P^T$

$$\begin{aligned}
\nabla_x f(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{2}(x+h)^T P(x+h) - (\frac{1}{2}x^T P x)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{2}(x^T P x + x^T P h + h^T P x + h^T P h - x^T P x)}{h} \\
&= \lim_{h\to 0} (\frac{1}{2}x^T P + \frac{1}{2}h^T P + \frac{(Px)^T h}{2h}) \\
&= \frac{1}{2}P^T x + \frac{1}{2}Px \\
&= Px
\end{aligned}$$

(d) $f(x) = (a^T x - b)(c^T x - d)$

$$
\begin{aligned}
\nabla_x f(x) &= \lim_{h \to 0} \frac{(a^T(x+h)-b)(c^T(x+h)-d)-(a^Tx-b)(c^Tx-d)}{h} \\
&= \lim_{h \to 0} \frac{(a^Tx-b+a^Th)(c^Tx-d+c^Th)-(a^Tx-b)(c^Tx-d)}{h} \\
&= \lim_{h \to 0} \frac{(a^Tx-b)(c^Tx-d)+(a^Tx-b)c^Th+a^Th(c^Tx-d+c^Th)-(a^Tx-b)(c^Tx-d)}{h} \\
&= \lim_{h \to 0} \left((a^Tx-b)c^T + \frac{a^Thc^Tx-a^Thd+a^Thc^Th}{h}\right)
\end{aligned}
$$

Since both d and $c^Tx \in \mathbb{R}$, we can move them to the front

$$
\begin{aligned}
\nabla_x f(x) &= \lim_{h \to 0} \left((a^Tx-b)c^T + \frac{(c^Tx)\cdot a^Th - d\cdot a^Th + a^Thc^Th}{h}\right) \\
&= \lim_{h \to 0} \left((a^Tx-b)c^T + (c^Tx)\cdot a^T - d\cdot a^T + a^Thc^T\right) \\
&= 2a^Txc^T - bc^T - da^T
\end{aligned}
$$

(e) $f(x) = \frac{1}{1+exp(-g(x))}$

$$
\begin{aligned}
\nabla_x f(x) &= -\frac{\frac{\partial(1+e^{-g(x)})}{\partial x}}{(1+e^{-g(x)})^2} \\
&= -\frac{e^{-g(x)}\cdot\frac{\partial -g(x)}{\partial x}}{(1+e^{-g(x)})^2} \\
&= \frac{e^{-g(x)}\cdot\nabla_x g(x)}{(1+e^{-g(x)})^2} \\
&= f(x)\cdot(1-f(x))\cdot\nabla_x g(x)
\end{aligned}
$$

1.8 Trace

$(X^TY)_{i,j} = \sum_{k=1}^{m} X_{ki}Y_{kj}$

$$
\begin{aligned}
tr(X^TY) &= \sum_{i=1}^{n}(X^TY)_{i,i} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ji}Y_{ji}
\end{aligned}
$$

$\therefore$ The trace of X^TY is the sum of the dot products of each column of X and Y, which is equivelent to considering X and Y as a large matrix $\in \mathbb{R}^{(m*n)*1}$ that connects all of their columns together respectively.

Like $X' = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{m1} & X_{12} & X_{22} & \cdots & X_{m2} & \cdots & X_{mn} \end{bmatrix}$

Since $tr(X^TY)$ can be considered as a large dot prodcut, it is also an inner product itself.

1.9 Anti-symmetric matrix

$$
\begin{aligned}
x^TAx &= (x^TAx)^T \\
&= x^TA^Tx \\
&= -x^TAx
\end{aligned}
$$

$\therefore 2x^T Ax = 0$
$\therefore x^T Ax = 0$

$$\frac{1}{2}x^T(B+B^T)x = xTBx$$
$$xTBx - \frac{1}{2}x^T(B+B^T)x = 0$$
$$x^T(B-B^T)x = 0$$

since for any matrix B $\in \mathbb{R}^{n\times n} B - B^T$ is anti-symmetric
and for an anti-symmetric matrix A we have $x^T Ax = 0$
$\therefore x^T(B+B^T)x = 0$
$\therefore \frac{1}{2}x^T(B+B^T)x = x^T Bx$

1.10 Maximum entropy

(a)

Since $1^T x = 1$
$\therefore \sum_{i=1}^{n} x_i = 1 \therefore x_n = 1 - \sum_{i=1}^{n-1} x_i$

(b)

$$\sum_{i=1}^{n} x_i \log_{x_i} = \sum_{i=1}^{n-1} x_i \log x_i + (1 - \sum_{i=1}^{n-1} x_i) \log (1 - \sum_{i=1}^{n-1} x_i)$$

(c)

$$\begin{aligned}\nabla_{x_i}\left(\sum_{i=1}^{n} x_i \log_{x_i}\right) &= 1 + \log x_i + \frac{1}{1-\sum_{i=1}^{n-1} x_i} - \left(\log (1 - \sum_{i=1}^{n-1} x_i) + \sum_{i=1}^{n-1} x_i \times \frac{-1}{1-\sum_{i=1}^{n-1} x_i}\right) \\ &= 1 + \log x_i - \log (1 - \sum_{i=1}^{n-1} x_i) - 1 \\ &= \log \frac{x_i}{1-\sum_{i=1}^{n-1} x_i} = 0\end{aligned}$$

$\therefore \frac{x_i}{1-\sum_{i=1}^{n-1} x_i} = 1$

$\therefore$for any i = 0,1, $\cdots$,n,
$x_i = 1 - \sum_{i=1}^{n-1} x_i = x_n$

$\therefore \sum_{i=1}^{n} x_i = n \times x_n = 1$

$\therefore x_i = \frac{1}{n}$