MIS Homework 2

Xinhao Du

January 2023

Problem 3.4

Solution:

(a) $(a+b)^T(a-b)$

$= a^T(a-b) + b^T(a-b)$

$= a^Ta - a^Tb + b^Ta - b^Tb$

$= \|a\|^2 - \|b\|^2$

(b) $\|a+b\|^2 + \|a-b\|^2$

$= (a+b)^T(a+b) + (a-b)^T(a-b)$

$= a^Ta + 2a^Tb + b^Tb + a^Ta - 2a^Tb + b^Tb$

$= 2a^Ta + 2b^Tb$

$= 2(\|a\|^2 + \|b\|^2)$

Problem 3.7

Solution:

According to Chebyshev inequality, number of X_i with $|X_i| \geq 3$ is no more than $\frac{\|X\|^2}{\alpha} = \frac{100}{9} \approx 11$. Also the example of 11 entries ≥ 3 : set $X = (31_{11}, 1, 0_{88})$

Problem 3.14

Solution:

$$\begin{aligned} \|x - e_i\|^2 &= \|x\|^2 - 2\|x\|\|e_i\| + \|e_i\|^2 \\ &= \|x\|^2 - 2x^T e_i + 1 \\ &= \|x\|^2 + 1 - 2x_i \end{aligned}$$

$\|x\|^2$ and 1 are constant, we could choose $x_i = \max_j x_j$

Problem 3.21

Solution:

(a) $\mathcal{D} = \|x_{1:T-1} - x_{2:T}\|^2$

(b) The smallest value of $\mathcal{D}(x)$ is 0, and signals are constant and all entries are equal.

(c). When $x_i = (-1)^i$ or $x_i = -(-1)^i$ makes the value of $\mathcal{D}$ largest, so the sum equal to 4 and $\mathcal{D} = 4(T-1)$.

Problem 3.26

Solution:

(a). $R(0)$ is correlation coefficient, so $R(0) = 1$.

For $R(1)$, both vectors have mean μ, so de-meaned of vectors are $(0, x-\mu 1_{n})$ and $(x-\mu 1_{n}, 0)$.

$$R(1) = \frac{(0_n x-\mu 1)^T (x-\mu 1, 0_n)}{\|x-\mu 1\|^2} = 0$$

(b) $$R(1) = \left(\frac{(0_n, x-\mu 1)}{\|x-\mu 1\|}\right)^T \left(\frac{(x-\mu 1, 0_n)}{\|x-\mu 1\|}\right)$$
$$= \left(\frac{(0_n, x-\mu 1)}{\sqrt{T}\ std(x)}\right)^T \left(\frac{(x-\mu 1, 0_n)}{\sqrt{T}\ std(x)}\right)$$
$$= \frac{1}{T}(0_n, z)^T (z, 0_n)$$
$$= \frac{1}{T}\sum_{t=1}^{T-n} z_t z_{t+n}$$

(c) $$R(n) = \frac{1}{T}\sum_{t=1}^{T-n} x_t \cdot x_{t+n}$$
$$= \frac{1}{T}\sum_{t=1}^{T-n} (-1)^{t+1} (-1)^{t+1+n}$$
$$= \frac{T-n}{T}(-1)^n$$
$$= \begin{cases} \frac{1-n}{T} & \text{when } n \text{ even} \\ \frac{n-1}{T} & \text{when } n \text{ odd} \end{cases}$$

(d) $R(7)$ is high means that x_t and x_{t+7} are higher or lower than mean value, x_{t+7} is the number of meals served in the week after x_t.

Problem A4.2

a

Solution:
using VMLS
using LinearAlgebra
articles, dictionary, titles=wikipedia_data()
Lth=length(articles)
k=5
assignment, reps=kmeans(articles, k)

b

Solution:
d = zeros(500)
for i in 1:500
d[i] = norm(articles[i] - reps[assignment[i]])
end

c and d

Solution:
for j = 1:k
group=[i for i=1:Lth if assignment[i]==j]
println()
println("Cluster: ", j)
tmp=sortperm(reps[j], rev=true)
println("5 most common words ", dictionary[tmp[1:5]])
println("Articles closest to representative 1: ")
tmp = sortperm(d[group]) for i= 1:5
println(" ", titles[group[tmp[i]]])
end
end

After we run the code, we can see 5 most common words and the 5 articles closest to representative 1 on the console:

```
Cluster: 1
5 most common words
 ["match", "win", "fight", "event", "title"]
Articles closest to representative:
 Wrestlemania_32
 Andre_the_Giant
 Floyd_Mayweather,_Jr.
 Ronda_Rousey
 Kimbo_Slice
```

e

Solution: Now we can repeat parts c and d for the other clusters. The 5 most common words in representative 1 are film, star, million, role, release, and then we can see the 5 articles closest for cluster 2, 3, 4 and 5 to each representative on the console

```
Cluster: 2
5 most common words
 ["film", "star", "million", "role", "release"]
Articles closest to representative:
 Leonardo_DiCaprio
 Kate_Beckinsale
 Maureen_O'Hara
 Star_Wars_Episode_I:_The_Phantom_Menace
 Star_Wars:_The_Force_Awakens

Cluster: 3
5 most common words
 ["album", "release", "song", "music", "single"]
Articles closest to representative:
 David_Bowie
 Kanye_West
 Celine_Dion
 Ariana_Grande
 Kesha

Cluster: 4
5 most common words
 ["season", "game", "team", "win", "play"]
Articles closest to representative:
 Kobe_Bryant
 Lamar_Odom
 Yogi_Berra
 Johan_Cruyff
 Halo_5:_Guardians

Cluster: 5
5 most common words
 ["series", "united", "film", "family", "american"]
Articles closest to representative:
 Ben_Affleck
 Mahatma_Gandhi
 Sigmund_Freud
 Carly_Fiorina
 Frederick_Douglass
```