MATH 7502 - Semsester 2, 2018

Mathematics for Data Science 2

Created by Zhihao Qiao, Maria K.... and Yoni Nazarathy

Assignment 1

Question (1)

(a) If $u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, v = \begin{bmatrix} -6 \\ 1 \\ -2 \end{bmatrix}, w = \begin{bmatrix} 4 \\ -6 \\ -1 \end{bmatrix}$,

prove and disprove that u, v and w lie in the same plane.

(b) Use the dot product to determine the angle between u and v.

(c) Find the length(L2 norm) of u and$v$, show that Cauchy-Schwarz inequality and the Triangle Inequality holds.

(d) Find the unit vectors associated with u,v and w.

solution

(a) Three vectors are in the same plane iff they are dependent. but u is independent from v and w, therefore they are not in the same plane. See also the Julia code below indicating they are independent.

(b) $u \cdot v = -6 + 2 - 6 = -10$.

$|u||v| = \sqrt{14}\sqrt{41}$

$\cos(\theta) = \frac{-10}{\sqrt{14}\sqrt{41}}$,

$\theta = -0.417$.

(c) as (b),$|u \cdot v| = 10$,

$\|u\|_2 = \sqrt{14}, \|v\|_2 = \sqrt{41}$,

we have that $10 < \sqrt{14}\sqrt{41}$. Therefore CS holds in L2 norm.

(d) $\hat{u} = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \hat{v} = \frac{1}{\sqrt{41}} \begin{bmatrix} -6 \\ 1 \\ -2 \end{bmatrix}, \hat{w} = \frac{1}{\sqrt{53}} \begin{bmatrix} 4 \\ 6 \\ -1 \end{bmatrix}$

In [2]:

```
A = [1 -6 4;
     2 1 -6;
     3 -2 -1];
nullspace(A),det(A)
#since the nullspace is {0} and they are indpendent
#Alt, since the det is not equal to 0 they are independent
```

Out[2]:

```
(Array{Float64}(3,0), 55.000000000000036)
```

Question 2

(a)Find unit vector u_1 and u_2 in the direction of $v = (1, 4)$ and $w = (-2, 1, 2)$ respectively.

(b)Find unit vectors v_1 and v_2 that are parallel to u_1 and u_2 (Are there such vectors not u_1 and u_2?).

(c)Find unit vectors w_1 and w_2 that are perpendicular to u_1 and u_2.

(d)Find the dot product of any unit vector $u + w$ and $u - w$.

solution

(a) $u_1 = \frac{1}{\sqrt{17}}(1, 4), u_2 = \frac{1}{3}(-2, 1, 2)$

(b) Suppose v_1 exists, then $v_1 = au_1$ and a is a real scalar. We know that v_1 is also a unit vector, therefore

$$\|v_1\|_2 = \sqrt{a}\|u_1\|_2 = 1$$

Since $\|u_1\|_2 = 1$, therefore we must have $\sqrt{a} = 1$, then $u_1 = v_1$...

(c) no unique answer, as long as the dot products are 0.

(d)

$$(u + w) \cdot (u - w) = u \cdot u - w \cdot w = \|u\|_2^2 - \|w\|_2^2 = 0$$

Question 3

Show the implementation of k-means algorithm (using k=2) for the following table by hand:

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Initialize with individual (1) and (4), in this case the centroid are $m_1 = (1.0, 1.0), m_2 = (5.0, 7.0)$.

(a) Using the initilizations given, verify two clusters of next step is $\{1, 2, 3\}$ and $\{4, 5, 6, 7\}$.

(b) Find the new centroids using (a).

(c) Verify the most optimized cluster with k =2.

Solution

(a) We can find the distance of each indivdiual from the centroid m_1 and m_2. e.g For individual (1), the distance to m_1 is 0, and distance to m_2 is $\sqrt{(1-5)^2+(1-7)^2} = \sqrt{52} = 7.21$

Individual	distance to m1	distance to m2
1	0.0	7.21
2	1.12	6.10
3	3.61	3.61
4	7.21	0
5	4.72	2.5
6	5.31	2.06
7	4.30	2.92

Therefore , the clusters are $\{1, 2, 3\} and \{4, 5, 6, 7\}$. or $\{1, 2\} and \{3, 4, 5, 6, 7\}$. There is not a unique partition due to the entry 3.61 for individual 3

(b) the new centroid, $m_1 = \left(\frac{1}{3}(1+1.5+3), \frac{1}{3}(1+2+4)\right) = (1.83, 2.33)$.

$m_2 = \left(\frac{1}{4}(5+3.5+4.5+3.5), \frac{1}{4}(7+5+5+4.5)\right) = (4.12, 5.38)$.

(c) repeat (a) with the new centroid, should get the same cluster.

Note: The fact that k-means stoped does NOT imply an optimal solution - but rather a "stopping of the algorithm" - this (heuristic) is in practice taken as "optimal". However, for pedagogical purposes, if we would like to actually find the optimum, it can be done by considering all possible 2^7=128 partions of these individuals. We have not done this for the solution.

Question 4

Let f be a mapping from $R^2 \rightarrow R^2$,

$$f(x,y) = \left(e^x \sin\left(y\right), e^x \cos\left(y\right)\right)$$

(a)Compute the Jacobian Matrix.

(b)Find its determinant.

(c) State conditions for invertability of x and y.

Solution

(a) we have $f_1(x,y) = e^x \sin(y), f_2(x,y) = e^x \cos(y)$.

$d_x(f_1) = e^x \sin(y), d_y(f_1) = e^x \cos(y)$,

$d_x(f_2) = e^x \cos(y), d_y(f_2) = -e^x \sin(y)$.

Therefore,

$$J = \begin{bmatrix} e^x \sin(y) & e^x \cos(y) \\ e^x \cos(y) & -e^x \sin(y) \end{bmatrix} = e^x \begin{bmatrix} \sin(y) & \cos(y) \\ \cos(y) & -\sin(y) \end{bmatrix}$$

.

(b)

$$det(J) = (e^x)^2 \left(- \sin^2(y) - \cos^2\left(y\right)\right) = -e^{2x}$$

(c) It is always invertable.

Question 5

True or false (give a reason or prove if true and a counter example if false).

(a) If $u = (1,1,1)$ is perpendicular to v and w for any v and w, then v is parallel to w.

(b) If u is perpendicular to v and w, then u is perpendicular to $v+w$.

(c) If u and v are perpendicular unit vectors, then $\|u-v\| = \sqrt{2}$.

(a) Not true, e.g $v = (1,-1,0)$ and $w = (0,-1,1)$.

(b) True, $u \cdot (v+w) = u \cdot v + u \cdot w = 0$.

(c) $\|u-v\|_2 = \sqrt{(u-v)\cdot(u-v)} = \sqrt{u \cdot u - u \cdot v - v \cdot u + v \cdot v} = \sqrt{u \cdot u + v \cdot v} = \sqrt{2}$

Question 6

Suppose each of the vectors $u_1, \ldots, u_k$ is a linear combinaton of the vectors $v_1, \ldots, v_m$. Assume that w is a linear combination of $u_1, \ldots, u_k$.

(a)Show that w is also a linear combianton of $v_1, \ldots, v_m$ for the case $m = k = 2$.

(b)Show the above for the genernal m and k. (Note: m does not necessilary equal to k)

Solution

(a) $u_1 = a_1 v_1 + b_1 v_2, u_2 = a_2 v_1 + b_2 v_2$, since w is a linear combination of u_1 and u_2,
$$w = pu_1 + qu_2 = p(a_1 v_1 + b_1 v_2) + q(a_2 v_1 + b_2 v_2) = (pa_1 + qa_2)v_1 + (pb_1 + qb_2)v_2.$$

(b)Let
$$u_j = a_{1,j} v_1 + \ldots + a_{m,j} v_m$$
and
$$w = b_1 u_1 + \ldots b_k u_k.$$

subsitute u_j to w, we have
$$w = b_1(a_{1,1} v_1 + \ldots + a_{m,1} v_m) + \ldots + b_k(a_{1,k} v_1 + \ldots + a_{m,k} v_m)$$

By rearrange the coefficents, we have
$$w = (b_1 a_{1,1} + b_2 a_{1,2} + \ldots + b_k a_{1,k})v_1 + (b_1 a_{2,1} + b_2 a_{2,2} + \ldots + b_k a_{2,k})v_2 + \ldots + (b_1 a_{m,1} + b_2 a_{,,}$$

Question 7

Determine whether each of the following scalar-valued functions of n-vectors is linear. If it is a linear function, find its inner product representation, i.e Find an n-vector a such that $f(x) = a^T x$. If it is not linear, give specific x, y, α, β such that
$$f(\alpha x + \beta y) \neq \alpha f(x) + \beta f(y)$$

(a) $f(x) = \max_k x_k - \min_k x_k$ (The spread of values of the vector).

(b) $f(x) = x_n - x_1$ (The difference of last element and the first).

(c) $f(x)$ = The median of an n-vector, suppose $n = 2k + 1$ is odd. then the median is the (k+1)th largest number among all entries of x.

(d) Define $x_{n+1} = f(x_n) = x_n + (x_n - x_{n-1})$, for $n \geq 2$ (This is a simple prediction of what x_{n+1} should be based on a straight line drawn through x_n and x_{n-1}.

Solution

(a) not true, i.e $x = (1, 0), y = (-1, 0), \alpha = -1, \beta = 1$

$$f(\alpha x + \beta y) = f((-2, 0)) = 2.$$
$$\alpha f(x) + \beta f(y) = -1(1 - 0) + 1(0 - (-1)) = 0$$

.

(b)

$f(\alpha x + \beta y) = (\alpha x_n + \beta y_n) - (\alpha x_1 - \beta y_1) = \alpha(x_n - x_1) + \beta(y_n - y_1)$

$\alpha f(x) + \beta f(y) = \alpha(x_n - x_1) + \beta(y_n - y_1)$

So this is linear.

i.e $f(x) = x_n - x_1$, then $f(x) = a^T x$, where $a^T = (-1, 0, 0, 0, \ldots, 0, 1)$

(c) not true, i.e $x = (-1, 0, 1), y = (1, 0, 1), \alpha = 0.5, \beta = 1$

$$f(\alpha x + \beta y) = f((0.5, 0, 1.5)) = 0.5$$
$$\alpha f(x) + \beta f(y) = 0.5 * 0 + 1 * 0 = 0$$

.

(d) This question was ill-posed - sorry.

Question 8

Clustering a collection of vectors into $k = 2$ groups is called 2-way partitioning, since we are partitioning the the vector into 2 groups, with index sets G_1 and G_2. Suppose we run k-means, with $k = 2$, on the n-vectors $x_1, \ldots, x_n$ with $x_i \in R^n$.

Show that there is a nonzero vector w and a scalar v that statisfy

$$w^T x_i + v \geq 0 \;\; \text{for} \;\; i \in G_1$$
$$w^T x_i + v \leq 0 \;\; \text{for} \;\; i \in G_2$$

In other words, the affine function $f(x) = w^T x + v$ is greater than or equal to zero in the first group, and less than or equal to zero in the second group. This is called the linear separation of the two groups. Hint: Consider the function $\|x - z_1\|^2 - \|x - z_2\|^2$, where z_1 and z_2 are the group representatives.

Solution,

Let $x_i \in G_1$ WLOG , then by definitions of clustering,

$$\|x_i - z_1\|^2 \leq \|x_i - z_2\|^2$$

Therefore, we must have

$$\|x_i\|^2 - 2\langle x_i, z_1\rangle + \|z_1\|^2 \leq \|x_i\|^2 - 2\langle x_i, z_2\rangle + \|z_2\|^2$$

Using inner product algebra, we must have

$$2\langle x_i, z_2 - z_1\rangle + \|z_1\|^2 - \|z_2\|^2 \leq 0$$

Similarily, if $x_i \in G_2$, then by definition of clusering,

$$\|x_i - z_1\|^2 \geq \|x_i - z_2\|^2$$

Therefore, we must have

$$\|x_i\|^2 - 2\langle x_i, z_1\rangle + \|z_1\|^2 \geq \|x_i\|^2 - 2\langle x_i, z_2\rangle + \|z_2\|^2$$

Using inner product algebra, we must have

$$2\langle x_i, z_2 - z_1\rangle + \|z_1\|^2 - \|z_2\|^2 \geq 0$$

Hence, if we let $w = z_1 - z_2$ and $v = \|z_2\|^2 - \|z_1\|^2$, where z_1 and z_2 are any group representatives from G_1 and G_2, this must exist and statisfies the condition.