Alon Bar 201639663
Daniel Meller 038198149

Homework 1

Convexity

1.1.1

Lets prove that

$$R^d_{\geq 0}$$

is convex

let

$$0 \leq t \leq 1$$

and

$$x, y \in R^d_{\geq 0}$$

then for any

$$0 \leq i \leq d$$

if

$$z = x + y$$

$$z_i = x_i + y_i \geq x_i \geq 0$$

so

$$z \in R^d_{\geq 0}$$

so

$$R^d_{\geq 0}$$

is convex, as required

1.1.2

Lets prove by induction for any $N \geq 2$ that for $\{\alpha\}_{i=1}^{N}$ with $\sum_{i=1}^{N} \alpha_i = 1$ and $\{x_i \in C\}_{i=1}^{N}$ the following occurs: $\sum_{i=1}^{N} \alpha_i x_i \in N$

For N = 2, it occurs due to the the definition of a convex set.
Lets assume correctness for any N - 1 and prove for N

So $\sum_{i=1}^{N} \alpha_i x_i = (\alpha_1 + ... + \alpha_{n-1})(\frac{\alpha_1}{\sum_{i=1}^{N-1} \alpha_i} x_1 + ... + \frac{\alpha_{n-1}}{\sum_{i=1}^{N-1} \alpha_i} x_{n-1}) + \alpha_n x_n$

From the induction base we have that $(\frac{\alpha_1}{\sum_{i=1}^{N-1} \alpha_i} x_1 + ... + \frac{\alpha_{n-1}}{\sum_{i=1}^{N-1} \alpha_i} x_{n-1}) \in C$ since

$\frac{\alpha_1}{\sum_{i=1}^{N-1} \alpha_i} + ... + \frac{\alpha_{n-1}}{\sum_{i=1}^{N-1} \alpha_i} = 1$

Also $(\alpha_1 + ... + \alpha_{n-1}) = 1 - a_n$ thus we have 2 terms in C with coefficients that sum up to 1.

There for $\sum_{i=1}^{N} \alpha_i x_i \in C$ as required.

1.1.3

This is not true

Consider C as the sphere $x_1^2 + x_2^2 - 2 = 0$

Then C is a convex set. Now maybe $\{x_i\}$ are all points on the line $x_1 + x_2 = 0$which intersects with the sphere for instance $(0, 0)$, $(1, -1)$, $(-0.5, 0.5)$.....

The the point $(1, 1)$ for instance is in the sphere but may not be represented by a convex combination of the points

1.1.4

Lets define $z := \alpha x + (1 - \alpha)y$

By definition

$f(y) \geq f(z) + f'(z)(y - z) => (1 - \alpha)f(y) \geq (1 - \alpha)f(z) + (1 - \alpha)f'(z)(y - z),$

$f(x) \geq f(z) + f'(z)(x - z) => \alpha f(x) \geq \alpha f(z) + \alpha f'(z)(x - z)$

Now lets add the two terms

$(1 - a)f(y) + \alpha f(x) \geq \alpha f(z) + (1 - \alpha)f(z) + f'(z)((1 - \alpha)(y - z) + \alpha(x - z)) =$

$f(z) + f'(z)((1 - \alpha)y + \alpha x - z) = f(z) + f'(z) \cdot 0 = f(\alpha x + (1 - \alpha)y)$

So we got :

$(1 - a)f(y) + \alpha f(x) \geq f((1 - \alpha)y + \alpha x)$
Thus f is convex

2.3 Hierarchical Clustering

2.3.1

Lets prove that $d^2_{complete-link}(C_1, C_2)$ is a metric

We will show that the 3 properties of a metric hold, and deduce that it is indeed an metric.

1. Identity:
   a. By definition, if $C_1 = C_2$ then the function yields 0 so $x = y \Rightarrow d(x, y) = 0$
   b.
2. Symmetry:
   a. Since $\forall x_i, x_j ||x_i - x_j||_2^2 = ||x_j - x_i||_2^2$ then
      $d^2_{complete-link}(C_1, C_2) = d^2_{complete-link}(C_2, C_1)$
3. Triangle inequality:
   a. Need to show that
      $d^2_{complete-link}(C_1, C_2) + d^2_{complete-link}(C_2, C_3) \geq d^2_{complete-link}(C_1, C_3)$
      For some $x_1 \in C_1,\ x_3 \in C_3$
      $d^2_{complete-link}(C_1, C_3) = ||x_1 - x_3||_2^2$
      Let's choose some random $x_2 \in C_2$ then from triangle inequality
      $||x_1 - x_3||_2^2 \leq ||x_1 - x_2||_2^2 + ||x_2 - x_3||_2^2$
      Also lets mark with $x^*,\ x^{**} \in C_2$ the appripriates members in C_2 that produce the values of the metrics $d^2_{complete-link}(C_1, C_2),\ d^2_{complete-link}(C_2, C_3)$

      So we get that
      $d^2_{complete-link}(C_1, C_3) =$
      $||x_1 - x_3||_2^2 \leq ||x_1 - x_2||_2^2 + ||x_2 - x_3||_2^2 \leq ||x_1 - x^*||_2^2 + ||x^{**} - x_3||_2^2 =$
      $d^2_{complete-link}(C_1, C_2),\ d^2_{complete-link}(C_2, C_3)$
      As needed

      Thus this is a metric

May 9, 2021

1 1.2.1

we need to prove that $\forall h \in \mathrm{R}^d$ $\bigtriangledown \mathrm{f}(\mathrm{x}_0)[h] =< g_0, h > \Longrightarrow \mathrm{g}_0 = \bigtriangledown f(\mathrm{x}_0)$
by defenition, if $f(x_0)[h] =< g_0, h >$, it is also true for h of the standard base form such as:
$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \vdots \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$
$\bigtriangledown \mathrm{f}(\mathrm{x}_0)[e_i] = \frac{\partial f}{\partial x_i}(x_0)$, we know that $g = (g_1, g_2 \vdots, g_d)$, so the inner product $< g, e_i >= g_i$ and we can conclude that $g_i = \frac{\partial f}{\partial x_i}$ which derives that $g(x_0) = \bigtriangledown f(x_0)$

2 1.2.2

$\bigtriangledown f(x)[h] := \lim_{x \to\ 0} \frac{f(x+ht)-f(x)}{t} = \lim_{x \to\ 0} \frac{f(x)+tf(x)-f(x)}{t} = \frac{tf(h)}{t} = f(h)$

3 1.2.3.1

$f(x) = x^T Ax$
$\bigtriangledown f(x)[h] = \lim_{t \to\ 0} \frac{f(x+th)-f(x)}{t} = \lim_{t \to\ 0} \frac{(x+th)^T A(x+th)-x^T Ax}{t} =$
$= \lim_{t \to\ 0} \frac{x^T Ax+tx^T Ah+th^T Ax+t^2 x^T Ax-x^T Ax}{t}$
$= \lim_{t \to\ 0} \frac{t(x^T Ah+h^T Ax+tx^T Ax)}{t}$
$= x^T Ah + h^T Ax* = x^T Ah + x^T A^T h = x^T (A + A^T)h =< (A + A^T)x, h >$
$\bigtriangledown f(x)[h] = (A + A^T)x$

4 1.2.3.2

$\bigtriangledown f(x)[h] = \bigtriangledown < g, h > (X)[H] = ** < \bigtriangledown g(X)[H], l(X) > + < g(X), \bigtriangledown h(X)[H] >=< H, AX > + < X, AH >=< (A + A^T)X, H >$
**using the product rule to derive this formula

5 1.2.3.3

$f(x) = ||y - Ax|| =< y - Ax, y - Ax > \bigtriangledown f(x)[h] =< -Ah, y - Ax > + < y - Ax, -Ah >= 2 < y - Ax, -Ah >= -2 < A^T(y - Ax), h > \bigtriangledown f(x) = -2A^T(y - Ax)$
using the product rule to derive this formula

6 1.2.3.4

same like previous section
$f(X) = ||Y - AX||_F^2 =< Y - AX, Y - AX > \bigtriangledown f(X)[H] =< -AH, Y - AX > + < Y - AX, -AH >= 2 < Y - AX, -AH >= -2 < A^T(Y - AX), H > \bigtriangledown f(X) = -2A^T(Y - AX)$
using the product rule to derive this formula

7 1.2.3.5

$f(X) =< X^T A, Y^T >$
$\bigtriangledown f(X)[H] = \bigtriangledown < X^T A, Y^T > [H] =<^T A[H], Y^T > + < X^T A, ^T [H] >=< AY, H > \longrightarrow \bigtriangledown f(X) = AY$

8 1.2.3.7

$f(X) =< A, log[X] >, g(X) = log[X]$

$\bigtriangledown g(X)[H] = \bigtriangledown log(X)[H] = \lim_{t \to 0} \frac{\log[X+tH] - \log[X]}{t} = \lim_{t \to 0} \frac{\log[\frac{X+tH}{X}]}{t} = H \circ X^{\circ -1} = **(X)[H] =< 0, \log[X] > + < A, H \circ X^{\circ -1} >=< A \circ X^{\circ -1}, H > \longrightarrow (X) = A \circ X^{\circ -1}$
**product rule

9 1.2.3.8

$f(X) =< a, diag(X) >$
$\bigtriangledown f(X)[H] = \lim_{t \to 0} Tr[\frac{adiag(X+tH) - adiag(X)}{t} = \lim_{t \to 0} Tr[a\frac{diag(X) + diag(tH) - diag(X)}{t} = Tr[adiag(H)) =< a, diag(H) > \longrightarrow \bigtriangledown f(X) = diag(a)$

10 1.4

$G(x, y) = sin(x + y)$

11 1.4.1.1

$min_x max_y G(x, y)$
exists x_0 such that $max_y sin(x_0 + y) = 1$ (sin function properties)
meaning, $min_x 1 = 1$

12 1.4.1.2

$max_y min_x G(x, y)$
exists y_0 such that $min_x sin(x + y_0) = -1$ (sin function properties)
meaning, $max_y - 1 = -1$

13 1.4.2.1

according to Rayleigh quotient $f(x) = \frac{x^T Ax}{x^T x}$, by the defenition in class, we know that $f(x) = \frac{x^T Ax}{x^T x} = \frac{<x,Ax>}{<x,x>}$.
for some scalar k, we know that $f(kx) = \frac{x^T Ax}{x^T x} = \frac{<kx,kAx>}{<kx,kx>} = \frac{k^2<x,Ax>}{k^2<x,x>} = \frac{<x,Ax>}{<x,x>} = f(x)$.
we can conclude that for every x that minimize f(x), kx also minize f(x), so we can write $min(x^T Ax)$ such that $||X||^2 = 1$ and we can conclude that $minf(x) = min < x, Ax >= min(x^T Ax)$, what we were asked to prove

14 1.4.2.2

$Ł(x, \lambda) = x^T Ax - \lambda(x^T x - 1)$

15 1.4.2.3

To prove: $\bigtriangledown Ł(x, \lambda) = 0 \leftrightarrow Ax = \lambda x$
$\bigtriangledown Ł(x, \lambda) = 2Ax - \lambda 2x$
$2Ax - \lambda 2x = 0 \leftrightarrow Ax = \lambda x$

16 2.1.1.1

we will show equivalence for the minimization target.
$\Sigma_{x_i \in D_k} ||x_i - \mu_k||_2^2 = \Sigma_{x_i \in D_k} ||x_i||_2^2 + ||\mu_k||_2^2 - 2 < x_i, \mu_k >$
$= \Sigma_{x_i \in D_k} ||x_i||_2^2 + \Sigma_{x_i \in D_k} ||\mu_k||_2^2 - 2\Sigma_{x_i \in D_k} < x_i, \mu_k >=$
$= \Sigma_{x_i \in D_k} ||x_i||_2^2 + N||\mu_k||_2^2 - 2N < \frac{1}{N} \Sigma_{x_i \in D_k} x_i, \mu_k >=**$
$= \Sigma_{x_i \in D_k} ||x_i||_2^2 + N||\mu_k||_2^2 - 2N||\mu_k||_2^2$
$= \Sigma_{x_i \in D_k} ||x_i||_2^2 - N||\mu_k||_2^2$

$\Sigma_{x_i,x_j\in D_k}||x_i - x_j||_2^2 = \Sigma_{x_i\in D_k}||x_i||_2^2 + \Sigma_{x_j\in D_k}||x_j||_2^2 - \Sigma_{x_i,x_j\in D_k} 2 < x_i, x_j >$
$= N\Sigma_{x_i\in D_k}||x_i||_2^2 + N\Sigma_{x_j\in D_k}||x_j||_2^2 - \Sigma_{x_i,x_j\in D_k} 2 < x_i, x_j >$
$= 2N\Sigma_{x_i\in D_k}||x_i||_2^2 - 2\Sigma_{x_i\in D_k}\Sigma_{x_i\in D_k} < x_i, x_j >$
$= 2N\Sigma_{x_i\in D_k}||x_i||_2^2 - 2\Sigma_{x_i\in D_k} < x_i, \Sigma_{x_i\in D_k} x_j >$
$= 2N(\Sigma_{x_i\in D_k}||x_i||_2^2 - \Sigma_{x_i\in D_k} < x_i, \mu_k >$
$= 2N(\Sigma_{x_i\in D_k}||x_i||_2^2 - N||\mu_k||_2^2)$

this equivalence is correct since we have only constant difference between them.

17 2.1.1.2

$arg\min_{\mu_k} \Sigma_{i=1}^N \min_k ||x_i - \mu_k||_2^2$
by definition of the K-Means algorithm, every $x_i \in D_k$ it holds that $k = arg\min d(x_i, \mu_i)$
we will use that in the objective of kmeans $arg\min \Sigma_{k=1}^K \Sigma_{x_i\in D_k}||x_i - \mu_k||_2^2 = arg\min \Sigma_{k=1} =^K \min ||x_i - \mu_k||_2^2$

18 2.1.2

the statement is False
a counter example will be: 4 points that create a rectangle, where the centroids are in the longer edge of the rectangle. lets use the four points (1,2),(1,4),(7,2),(7,4).
if the centroids starts at (4,2),(4,4) - it will set two clusters and the centroids will be the center of the clusters, and it will not change.
. while the global solution will be (1,3), (7,3)

19 2.2.1

$\mu_y = E[Y] = E[a^T X + b] = E[a^T X] + b = a^T \mu_x + b$
$\Sigma_y = Var(a^T X + b) = Var(Ax) + Var(b) = a^T \Sigma_x a$
$Y \sim N(a^T \mu_x + b,\ a^T \Sigma_x a)$
$f_Y(y) = \frac{1}{(2\pi)^d det\Sigma_y} exp(-\frac{1}{2}(y - \mu_y)^T \Sigma_y^{-1}(y - \mu_y) =$
$= \frac{1}{(2\pi)^d det(a^T\Sigma_x a)} exp(-\frac{1}{2}(y - a^T\mu_x + b)^T (a^T\Sigma_x a)^- 1(y - a^T\mu_x + b)$

20 2.2.2

Symmetry:
$\Sigma_x^T = E[(X - \mu_x)(X - \mu_x)^T]^T = E[((X - \mu_x)^T)^T (X - \mu_x)^T] = E[(X - \mu_x)(X - \mu_x)^T] = \Sigma_x$
Positive:

let $Y = X - \mu_x$ and some vector v.
$v^T \Sigma_x v = v^T E[YY^T]v = E[v^T YY^T v] = E[(v^T Y)(v^T Y)^T] = E[||vY||^2] \geq 0$