תרגיל בית 2

תאריך הגשה: 14.05.19

מוריאל בן משה 304832512, דניאל אנגלסמן 300546173

א. הראו כי מתקיים:

$$J=\sum_{k=1}^{n}x_k^Tx_k-\left(\frac{1}{n_1}\sum_{k\in G_1}\sum_{l\in G_1}x_k^Tx_l+\frac{1}{n_2}\sum_{k\in G_2}\sum_{l\in G_2}x_k^Tx_l\right)=\frac{1}{2n_1}\sum_{k\in G_1}\sum_{l\in G_1}d_{k,l}+\frac{1}{2n_2}\sum_{k\in G_2}\sum_{l\in G_2}d_{k,l}$$

1

$$J\triangleq J_1+J_2\triangleq\sum_{k\in G_1}\|x_k-\mu_1\|^2+\sum_{k\in G_2}\|x_k-\mu_2\|^2$$

$$J\triangleq\sum_{k\in G_1}\left\|x_k-\frac{1}{n_1}\sum_{l\in G_1}x_l\right\|^2+\sum_{k\in G_2}\left\|x_k-\frac{1}{n_2}\sum_{l\in G_2}x_l\right\|^2$$

$$J_i=\sum_{k\in G_i}\left[x_k^Tx_k-\frac{2}{n_i}\sum_{l\in G_i}x_lx_k+\frac{1}{n_i^2}\sum_{l\in G_i}x_l\sum_{l\in G_i}x_l\right]$$

Plug in each of the terms :

$$J_1=\sum_{k\in G_1}x_k^Tx_k-\frac{2}{n_1}\sum_{k\in G_1}x_k\sum_{l\in G_1}x_l+\frac{1}{n_1}\sum_{l\in G_1}x_l\sum_{l\in G_1}x_l=\sum_{k\in G_1}x_k^Tx_k-\frac{1}{n_1}\sum_{k\in G_1}\sum_{l\in G_1}x_k^Tx_l$$

$$J_2=\sum_{k\in G_2}x_k^Tx_k-\frac{2}{n_2}\sum_{k\in G_2}x_k\sum_{l\in G_2}x_l+\frac{1}{n_2}\sum_{l\in G_2}x_l\sum_{l\in G_2}x_l=\sum_{k\in G_2}x_k^Tx_k-\frac{1}{n_2}\sum_{k\in G_2}\sum_{l\in G_2}x_k^Tx_l$$

$$J=\sum_{k=1}^{n}x_k^Tx_k-\Big(\frac{1}{n_1}\sum_{k\in G_1}\sum_{l\in G_1}x_k^Tx_l+\frac{1}{n_2}\sum_{k\in G_2}\sum_{l\in G_2}x_k^Tx_l\Big)$$

We'll now develop the RHS of the equation :

$$\frac{1}{2n_1}\sum_{k\in G_1}\sum_{l\in G_1}d_{k,l}=\frac{1}{2n_1}(n_1\sum_{k\in G_1}x_k^Tx_k-2\sum_{k,l\in G_1}x_k^Tx_l+n_1\sum_{l\in G_1}x_l^Tx_l)=\sum_{k\in G_1}x_k^Tx_k-\frac{1}{n_1}\sum_{k,l\in G_1}x_k^Tx_l$$

$$\frac{1}{2n_2}\sum_{k\in G_2}\sum_{l\in G_2}d_{k,l}=\frac{1}{2n_2}(n_2\sum_{k\in G_2}x_k^Tx_k-2\sum_{k,l\in G_2}x_k^Tx_l+n_2\sum_{l\in G_2}x_l^Tx_l)=\sum_{k\in G_2}x_k^Tx_k-\frac{1}{n_2}\sum_{k,l\in G_2}x_k^Tx_l$$

Combine them together we get the proof :

$$\frac{1}{2n_1}\sum_{k,l\in G_1} + \frac{1}{2n_2}\sum_{k,l\in G_2} = \sum_{i=1}^{n} x_k^T x_k - \Big(\frac{1}{n_1}\sum_{k\in G_1}\sum_{l\in G_1} x_k^T x_l + \frac{1}{n_2}\sum_{k\in G_2}\sum_{l\in G_2} x_k^T x_l\Big) \tag{1.1}$$

ב. הראו כי מזעור משוואה (1), שקול לפתרון של בעיית **המקסימיזציה** של הפונקציה הבאה :

$$J_D \triangleq \frac{n_1 n_2}{n}\left[2\frac{d(G_1,G_2)}{n_1 n_2} - \frac{d(G_1,G_1)}{n_1^2} - \frac{d(G_2,G_2)}{n_2^2}\right]$$

Let us open the J_D expression and try to prove the maximization $J = C - 0.5J_D$:

$$\begin{aligned}
d(G_1,G_1) &= \sum_{k\in G_1}\sum_{l\in G_1} d_{k,l} = 2n_1 \sum_{k\in G_1} x_k^T x_k - 2\sum_{k,l\in G_1} x_k^T x_l\\
d(G_2,G_2) &= \sum_{k\in G_2}\sum_{l\in G_2} d_{k,l} = 2n_2 \sum_{k\in G_2} x_k^T x_k - 2\sum_{k,l\in G_2} x_k^T x_l\\
d(G_1,G_2) &= n_2 \sum_{k\in G_1} x_k^T x_k - 2\sum_{k\in G_1}\sum_{l\in G_2} x_k^T x_l + n_1 \sum_{l\in G_2} x_l^T x_l
\end{aligned}$$

After plugging, several terms are subtracting each other, and we get :

$$\begin{aligned}
J_D &= \frac{n_1 n_2}{n}\Big[-\frac{4}{n_1 n_2}\sum_{k\in G_1}\sum_{l\in G_2} x_k^T x_l - \frac{2}{n_1^2}\sum_{k,l\in G_1} x_k^T x_l + \frac{2}{n_2^2}\sum_{k,l\in G_2} x_k^T x_l\Big]\\
J_D &= 2\Big[-\frac{2}{n}\sum_{k\in G_1}\sum_{l\in G_2} x_k^T x_l - \frac{n_2}{n_1 n}\sum_{k,l\in G_1} x_k^T x_l - \frac{n_1}{n_2 n}\sum_{k,l\in G_2} x_k^T x_l\Big]\\
J_D &= 2\Big[-\frac{2}{n}\sum_{k\in G_1}\sum_{l\in G_2} x_k^T x_l - \frac{n-n_1}{n_1 n}\sum_{k,l\in G_1} x_k^T x_l - \frac{n-n_2}{n_2 n}\sum_{k,l\in G_2} x_k^T x_l\Big]\\
J_D &= 2\Big[-\frac{2}{n}\sum_{k\in G_1}\sum_{l\in G_2} x_k^T x_l + \Big(\frac{1}{n_1}-\frac{1}{n}\Big)\sum_{k,l\in G_1} x_k^T x_l + \Big(\frac{1}{n_2}-\frac{1}{n}\Big)\sum_{k,l\in G_2} x_k^T x_l\Big]\\
J_D &= 2\Big[-\frac{1}{n}\sum_{k,l=1}^{n} x_k^T x_l + \frac{1}{n_1}\sum_{k,l\in G_1} x_k^T x_l + \frac{1}{n_2}\sum_{k,l\in G_2} x_k^T x_l\Big] \qquad (1.2)
\end{aligned}$$

Plugging it back into the equation : $J = Const. - 0.5J_D$, would show equality.

ג. הסבירו את משמעות האלגוריתם K-Means על סמך התוצאה של סעיף ג'. בפרט,
הסבירו את התפקיד של כל אחד משלושת האיברים ב- J_D.

The K-means algorithm **optimizes** differently the given terms in the expression :

$$J_D = 2\Big[\mathbf{(1)} - \frac{1}{n}\sum_{k,l=1}^{n} x_k^T x_l + \mathbf{(2)}\frac{1}{n_1}\sum_{k,l\in G_1} x_k^T x_l + \mathbf{(3)}\frac{1}{n_2}\sum_{k,l\in G_2} x_k^T x_l\Big] \tag{1.3}$$

(1) : Is the average square distance between both clusters >> maximized.
(2) : Is the average square distance inside G_1 >> minimized.
(3) : Is the average square distance inside G_2 >> minimized.

א. הפעילו את אלגוריתם K-means הרגיל (עם מדד המרחק האוקלידי) עבור $K=2$ מחלקות, עבור סט הנקודות הבא:

$x_1=(1,1),\quad x_2=(3,3),\quad x_3=(20,20)$ ועם אתחול המרכזים:

$x_4=(1,-1),\quad x_5=(3,-3),\quad x_6=(20,-20),$ $\quad c_1=(1,1),\quad c_2=(20,20)$.

הראו את שלבי האלגוריתם, וחשבו את הערכים אליהם יתכנס.

2

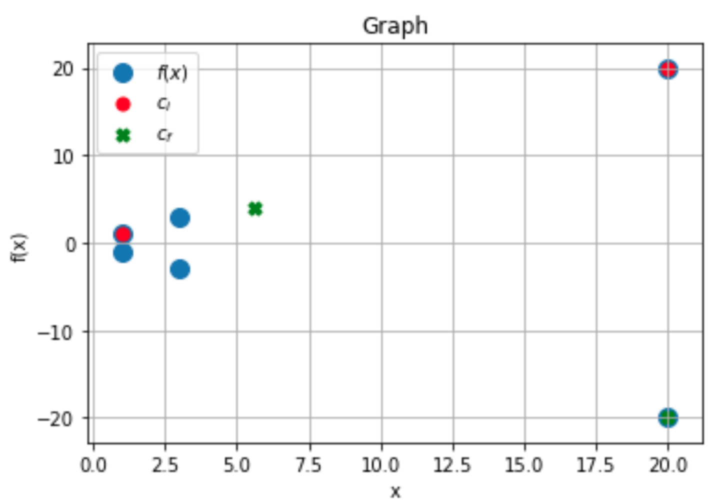
Willing to minimize the cost function from previous question, we shall calculate $x_i \to C(i)$:

$$\textbf{[1]} \qquad C(i) = argmin\ \ \|x_i - \mu_k\|^2$$

$$\textbf{[2]} \qquad \mu_k = \frac{1}{n_k} \sum_{i:C(i)=k} x_i, \quad k = 1 : K$$

$$\textbf{[3]} \qquad J(C) = \sum_{k=1}^{K} \sum_{i:C(i)=k} \|x_i - \mu_k\|^2$$

We iterate the process until reaching sufficient convergence criterion of J :



The final centroids ($\underline{c}_f$) are located at :

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} (5.6, 4.0) \\ (20, -20) \end{bmatrix} \tag{2.1}$$

כעת נדון בחלופה של האלגוריתם המתאימה לנקודות מתויגות. נניח כי נתון אוסף של
n נקודות $\{x_i, y_i\}$ כאשר $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^d$. נניח כי המידע מגיע מ- K מחלקות
שונות. עבור נקודה השייכת למחלקה , מתקיים: $y = w_k^T x + N, \quad N \sim \mathcal{N}(0,\sigma^2) \quad (iid)$
ו- $w_1,\dots,w_K$ סט המשקלים של המחלקות השונות.

יש לחלק את הנקודות ל- K מחלקות, וכן לחשב את המשקלים $w_1,\dots,w_K$ של כל מחלקה.

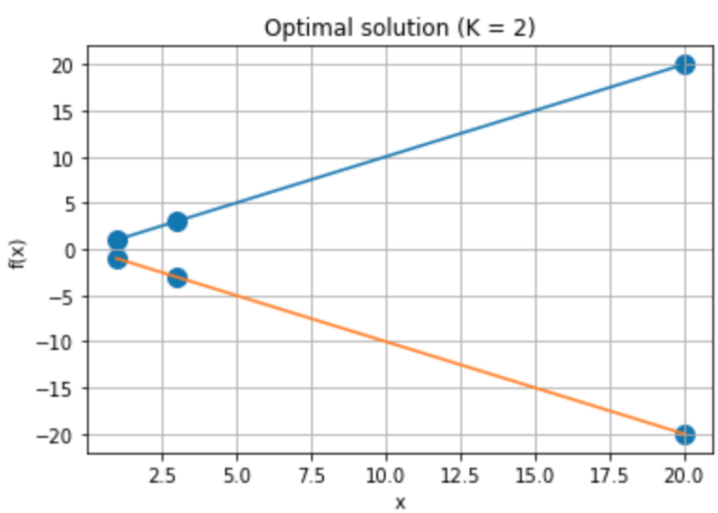
We would like to minimize the total average distances of $\{x_i, y_i\} \in G_k$ (inner loop):

$$J(C) = argmin \sum_{k=1}^{K} \Big(\sum_{i \in G_k} \Big\| y_i - w_k^T x_i \Big\|^2 \Big) \tag{2.2}$$

From here on we'll proceed similarly to the mentioned at section (2.1).

The optimal solution will be obtained via derivation (k = const.) :

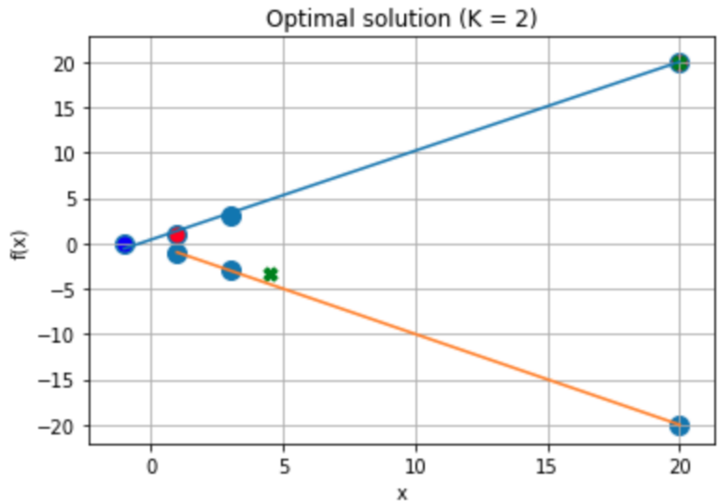
$$\begin{aligned} w^* &= \frac{d}{dw}\Big(\sum_{i \in G} y_i^2 - 2 \sum_{i \in G} y_i w^T x + \sum_{i \in G_k} (w^T x_i x_i^T w) \Big) \\ w^* &= \frac{d}{dw}\Big(c - 2w^T b + w^T Q w \Big) = 0 \quad \rightarrow \quad w^* = Q^{-1} b \end{aligned} \tag{2.3}$$



Using the set of x_i from (2.1), we get 2 linear separators:

$$w_1 = 1, \;\; w_2 = -1 \tag{2.4}$$

Adding 7th point $x_7 = (-1, 0)$ would slightly influence the w values :



There is a small shift of the centroids and the weights :

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} (4.5, -3.333) \\ (20, -20) \end{bmatrix}; \qquad \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 410/411 \\ -1 \end{bmatrix}; \tag{2.5}$$

The log-likelihood function is:

$$\begin{aligned} log\ P((x,y)|w) &= log \prod_{G_k} \prod_{(x_i,y_i)\in G_k} \frac{1}{\sqrt{2\pi\sigma_k}} exp\Big(-\frac{1}{2\sigma_k}\Big\|y_i - w^T x_i\Big\|^2\Big) \\ &= -\frac{N}{2} log(2\pi) + \frac{1}{2}\sum_{G_k}\sum_{(x_i,y_i)\in G_k} \frac{\big\|y_i - w^T x_i\big\|^2}{\sigma_k} - log(\sigma_k) \end{aligned} \tag{2.6}$$

Either here we will try to minimize the cost such that :

$$w^* = \frac{d}{dw}\Big(c - 2w^T b + w^T Q w\Big) = 0 \quad \rightarrow \quad w^* = Q^{-1} b$$

בשאלה זאת נפתח וריאציות שונות של אלגוריתם PCA ללמידה לא מודרכת (Unsupervised).

נתון סט הדוגמאות הבא: $\{x_i\}_{i=1}^{n} \in \mathbb{R}^d$, כאשר ממוצע הדגימות הינו אפס,

$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0$. בכתיב וקטורי, מטריצת הדוגמאות הינה: $X = [x_1, \dots, x_n]^T$.

א. 1) כתבו והסבירו בקצרה את אלגוריתם PCA.

הסבר: ראשית, נסחו את P_n ו - z_i כתלות ב v_m ו- λ_m, כאשר P_n היא מטריצת הקווריאנס, z_i הינו ייצוג הדוגמא ה– i בממד $\mathrm{M} \le d$. השתמשו בהסבר בסימונים v_m ו λ_m עבור ווקטורים וערכיים עצמיים בהתאמה.

2) הוכיחו שמטריצת P_n הינה מטריצת positive semi-definite – P.S.D.

תזכורת: מטריצה היינה PSD אם מתקיים: $\forall u \in \mathbb{R}^d$ מתקיים: $u^T P_n u \ge 0$.

3

(3.a) PCA is a dimensionality reduction algorithm that helps presenting compactly, large vectors. Given a dataset $\{x_i, y_i\} \in \mathbb{R}^d$ we'll do the following steps :

(1) Center the the dataset such that $x_i \leftarrow x_i - \bar{x}$ where $\bar{x} = \frac{1}{n}\sum_i^n x_i$.

(2) Define the covariance matrix :

$$P_n \triangleq \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \quad \Rightarrow_{centering} \quad P_n = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$$

Using a vector script : $X = [x_1, x_2, ..., x_n]^T$ we get the following matrix : $P_n = \frac{1}{n}X^T X$. We'll extract its eigen-values $\{\lambda_m\}_{m=1}^d$ and eigen-vectors $\{v_m\}_{m=1}^d$, such as $P_n v_m = \lambda_m v_m$. The eigen-vectors are also called principal directions of P_n and can span the $\mathbb{R}^d$ space as :

$$x_i = \sum_{m=1}^{d} z_{im} v_m, \qquad z_{im} \triangleq \langle v_m, x_i \rangle = v_m^T x_i, \qquad V = [v_1, v_2, ..., v_d]^T$$

(3) Calculate the PC by let $1 \le M \le d$ be the desired new reduced dimension :

$$x_i = V z_i \quad \Leftrightarrow \quad z_i = V^T x_i \quad \rightarrow \quad z_i^{(M)} = \begin{bmatrix} z_{i1} \\ \vdots \\ z_{iM} \end{bmatrix} = \begin{bmatrix} v_1^T \\ \vdots \\ v_M^T \end{bmatrix} x_i = V_M^T x_i, \quad i = [1, n]$$

The new M centers are chosen by the $\{\lambda\}_{i=1}^M$ biggest eigen-values.

(3.b) Prove P_n is P.S.D, such that $u^T P_n u \geq 0 \quad \forall u \in \mathbb{R}^d :$ (2 approaches ...)

$$Recall: \quad V^T = V^{-1}, \quad \Lambda = diag(\lambda_1, ..., \lambda_d) \quad and \quad P_n = V\Lambda V^T = \sum_{k=1}^{d} \lambda_k v_k v_k^T$$

(i)
$$u^T P_n u = u^T V \Lambda V^T u = (u^T V)_{1\times d} \Lambda (V^T u)_{d\times 1} \qquad subs. \quad w = V^T u \in \mathbb{R}^{d\times 1}$$

$$\Rightarrow \quad w^T \Lambda w = \bar{\lambda} \cdot (w^T I_d w)_{1\times 1} = \bar{\lambda}\big(u^T (VV^T) u\big) = \bar{\lambda}\big(u^T u\big) = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_d \end{bmatrix} \|u\|^2 \geq 0$$

(ii)
$$u^T (\sum_{k=1}^{d} \lambda_k v_k v_k^T) u = \sum_{k=1}^{d} \lambda_k u^T v_k v_k^T u = \sum_{k=1}^{d} \lambda_k (u^T v_k)(v_k^T u) = ...$$

$$... = \sum_{k=1}^{d} \lambda_k (u^T v_k)(u^T v_k)^T = \sum_{k=1}^{d} \lambda_k (u^T v_k)^2 \geq 0 \qquad (3.1)$$

The mentioned above is fulfilled since - $\lambda_1 \geq ... \geq \lambda_d \geq 0$.

נרצה להוריד את ממד הדוגמאות ל-M כאשר M ≤ d, כך ששגיאת השחזור הממוצעת במובן
של שגיאה ריבועית תהיה לכל היותר ϵ. נסחו את הבעיה למציאת מספר מינימלי של
וקטורים, M, שבעזרתם נוכל לייצג את הדוגמאות ברמת הדיוק הנדרשת.

For $m < d$ let $A \in \mathbb{R}^{M\times d}, B \in \mathbb{R}^{d\times M}$ be the dimensionality reduction matrices ($\mathbb{R}^d \to \mathbb{R}^M$).

The reconstruction error is : $e_i = x_i - \hat{x}_i = (I - BA)x_i$

As seen in lecture, the minimal reconstruction error is obtained for :

$$A = V_M^T, \quad and \quad B = V_m \equiv [v_1, ..., v_M]$$

And the minimal square error is therefore :

$$minimize \ (E_n = \sum_{i=1}^{n} \|e_i\|^2) \quad \to \quad E_{n,min} = \sum_{i=M+1}^{d} \lambda_i \leq \epsilon \qquad (3.2)$$

נסחו את PCA כבעיית אופטימיזציה עם אילוצים על מנת למצוא את הווקטור העיקרי המשמעותי ביותר בעזרת P_n ו u כאשר מתקיים:

$$u^T u = 1$$

Express as an optimization problem, and using **Eqn. 3.1** :

$$arg\max_u \; u^T P_n u \;=\; arg\max_u \; u^T(\sum_{k=1}^{d} \lambda_k v_k v_k^T)u = arg\max_u \sum_{k=1}^{d} \lambda_k (u^T v_k)(v_k^T u)$$

Since all eigen-vectors are orthonormal, every $j \neq k$ will result in 0, hence we get $u = v_k$.

$$arg\max_v \; \sum_{k=1}^{d} \lambda_k (v_k^T (v_k v_k^T) v_k)\Big|_{u^T u=1} \quad \rightarrow \quad arg\max_u \; \sum_{k=1}^{d} \lambda_k$$

Finally, the maximal value is :

$$arg\max_{\lambda}\{\lambda_1 \geq ... \geq \lambda_d \geq 0\} \quad \Rightarrow \quad \underline{\underline{\lambda_1}} \tag{3.3}$$

4

Answer the following question on paper and add them to the theoretical part (החלק היבש) of the assignment.

For the given set of points: $(1, 2, 3, 4, 5)$ write the sum of $L1$ distances between this set of points and any given x. Show that the median is, in fact, the point which has the minimal sum of $L1$ distances to this set of points.

Let us check if the minimal value of L_1 distance intersects with the median function :

$$L_1 = \|x - y\|_1 = \sum_{i=1}^{n} |x - y_i| = nx - \sum_{i=1}^{n} y_i$$

$$arg\min_x \;\; L_1 \quad \rightarrow \quad x^* = \frac{1}{n}\sum_{i=1}^{n} y_i \;\Big|_{n=5} \quad \rightarrow \quad \frac{1}{5}\sum_{i=1}^{n} y_i = \frac{15}{5} = 3$$

$$median(y) = \frac{y\lfloor\frac{n+1}{2}\rfloor + y\lceil\frac{n+1}{2}\rceil}{2} = \frac{y(3) + y(3)}{2} = y(3) = 3 \tag{4.1}$$