


תאריך הגשה לציון: 15/9. אין הארכות, מגישים ציעים סופיים.

התרגיל אמור לקחת כשבוע.



כלייה כג -  $k$ -means: יש העוק צאטא שא מופרז בצורה כחורה:

אלא אמש:  - אין קו אינארי שופרז.

כג: אלתנוצז עם הבריה פתחי אלתית  $Spectral\ clustering$ . ממיריט אלתה שמייצג את הצאטא

ואז ניק אלתפוז בין ה-  $clusters$  בצורה אינאריט.  $loss$  אלזריתים כמו  $PCA$ .

טניק: מקבלים נקודות ממייצג מסומ  $d$ . יש מטרצה  $d \times d$  ובנימ בגל ממך הנקודות.

ניצג בגל עי מטרצה שכניות. אם  $z$  נקודות קרובות אחת לשניה זה יתבטא בגל.

$RBF-kernal$  - יש  $2^2$  זוגות של צמתים. כג: לחשב ממקרים אוקחים את המרחק האוקלידי.

האלכסון יהיה 0 כי אין מרחק של נקודה מנקודה.

• הגל לא מכון.

• המטרצה סימטרית כי אין תשיבות לכיוון.

השיטה זו מניחים שהמרחק בין כל  $z$  נקודות נורמלי, מחשבים מרחק אוקלידי:

$$\forall i \neq j: w_{ij} = \exp\left[-\frac{(x_i - x_j)^2}{2}\right] \quad \text{שלא יהיה שום}$$

מקבלים מטרצית שכניות סימטרית עם עדרים אי-שליליים, ואין אולאות עזניות.

• האופן יצ' מאפסים את האלכסון.

כג: אלתפוז את הנתיים אלתה שמייצג נתיים בצורה טובה נקח את  $w$  שבנינו לפי ה- $kernal$ .

נקבל בגל הפט ולא מכון.

אחר מכן נבנה  $A$ -מטרצה  $D$  אלכסנית. כל שורה סופרים מה הצורה של כל צומת (האלכסון)

ולת השאר 0.

נבנה מטרצה שנקראת בגל אפליסאן. מפתחים את  $w$  ממטרצית הצורות  $D$ . מסמנים  $K-L$ .

המטרצה הסופית נקראת  $normalized\ graph\ Laplacian$  ( $L_{norm}$ ):

$$L_{norm} = D^{-1/2} \cdot L \cdot D^{-1/2}$$

• מלים מטרצה במסקת  $1/2$  עם יצ' 1 חלקי שורש האלכסון.

$L_{norm}$  מטריצת בינים. נחשב זה לדרכים ווקטורים עצמיים. כל המטריות שלנו סימטריות וזה יבדוק

יתברר:

האלגוריתם שנבנה נקרא: *normalized spectral clustering*.

(1) מקבלים את הצאטא. כל  $x$  ממייצג מסוים.

(2) בינים את  $L_{norm}$  כמו שהסברנו.

(3) אם לא יוצאים את מספר ה-clusters, קובעים את  $k$  האופטימלי.

(4)  $n$ - $L_{norm}$  בינים את מטריצת הוקטורים העצמיים  $U$ . סימטרית וריבועית ולכן

הערכים העצמיים שלה ממשיים, והוקטורים העצמיים אורתונורמליים.

לוקחים את  $k$  העמודות הראשונות (כמו שלוש ב-PCA).

$$X_{n \times d} \rightarrow U_{n \times k}$$

זה הייצוג החזק של ה-data.

(5) מנרמלים כל עמודה משלה  $U$ .

(6) מפעילים את  $k$ -means מתוצרת 1.

נקבע  $k$  קואסטרים ממייצג  $k$ .

בלט: הקואסטרים הסופיים ממייצגים  $k$ .

\* איך קובעים את  $k$  האופטימלי?

• או שמקבעים את  $k$  כמו קודם.

• אם לא מצליחים  $k$  מוצאים  $k$  אופטימלי.

המטריצה סימטרית הוקטורים ממשיים ואי-שליליים. ממייצגים את הערכים העצמיים

ומחפשים את הקפיצה הכי גדולה בין הערכים העצמיים.

ממייצגים את הערכים העצמיים מהצדון לקדן כך שכל ערך עצמי יתאים וקטור עצמי.

$$\text{Sort: } \lambda_1 \geq \dots \geq \lambda_k \geq 0$$

$$\text{Delta: } \delta_i = |\lambda_i - \lambda_{i+1}|$$

$$\text{Gap: } k = \arg \max_i (\delta_i), i = 1, \dots, \frac{n}{2}$$

חישוב וקטורים וערכים עצמיים-אלגוריתם יעקובי למטריצות סימטריות:

מטריצות צימודיות  $B = M^{-1} A M$  (וקטורים עצמיים לא בהכרח).

נרצה לחשב  $L$ -norm, וזה לא.

סביר מטריצת סיבוב  $P$  מתקיים:  $A' = P^T A P$

$P$  מטריצה אורתוגונלית.

השני האיברים מחוץ לאלכסון קטן מ- $\epsilon$  והאלגוריתם מתחיל.

מתחילים כל פעם  $P$  חדשה, כל פעם מאבדים איברים מחוץ לאלכסון ומקבלים מטריצה אלסגונית של ערכים עצמיים.

מטריצת סיבוב - שמים 1 על האלכסון ומוציאים איבר  $\text{pivot}$  שגרסה לאפס מחוץ לאלכסון.

נקראת  $\text{Jacobi rotation matrix}$ .

יש נוסחה ל- $\text{pivot}$  במוצאת.

לא כל האיברים משתנים ולכן נותן לעלות צורה מקוצרת:  $\text{matrix efficiency}$  במוצאת.

יש במוצאת בהקדמה צימוד של העצת האלגוריתם עם מטריצה  $3 \times 3$ .

$$A = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix} \quad \begin{matrix} i=1 \\ j=3 \\ t=1 \\ c=1/\sqrt{2} \\ s=1/\sqrt{2} \end{matrix}$$

$$P_1 = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$$

$$\begin{matrix} i=2 \\ j=3 \\ t=1/2 \sqrt{2} \\ c=4/3 \sqrt{2} \\ s=1/3 \end{matrix}$$

$$A' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & \sqrt{8} \\ 0 & \sqrt{8} & 1/\sqrt{2} \end{pmatrix} \quad \begin{matrix} \text{איבר מקטעני} \\ \text{שחצים לאפס} \end{matrix}$$

$$A' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

$$U = P_1 \cdot P_2 = \begin{pmatrix} 1/2 & -1/3 \sqrt{2} & 2/3 \\ 0 & 4/3 \sqrt{2} & 1/3 \\ -1/2 & -1/3 \sqrt{2} & 2/3 \end{pmatrix}$$

## מבנה מבין

מותק צף נוסמאת מוכת 1

המבחן שלת"ס

4.5.1

1) גאליה פתוחה ב-C

(2) 2 שאלות עם תשובה קצרה בפיגור

3) גאלה קצבה אל החומר של אתי

## חומר בפ"ח

כלים בתחום data science & בינה עמוקה

• לא שאנו חושבים, מה לעשות ביניהם.

• לביטויים - np, pd, ועד סקייט.

• אם צריך לנסוע 3 מ' 13' 13" שני אדם.

לקוח 13/2 מ

1 nke

חלק מפתחונים בקיץ וחלק בתשואה.

טבלת production של שירים ונתק 3 ו-10.

נרצה למשק תפוקת חברים  $\gamma$  : מסק  $x \cdot d$ .

ד- פתור של חברים ב' נמצא תפוקה של חברים 3,4 במחלת השנים הראשונה.

• אם לא ניתן מעט של חומרים של `numpy`.

שומרים את התוצאה ב- new emp.

צביר אבות 4 מוצה תוצה: production:  $\alpha \times d$ .

`d = production[:5, [2, 3]].mean()`

מחזיקים 0-4 ← 5 שורות הראשונות → תוצאה מסוג float

```
new_emps['production'] = new_emps['vetek'].production[:5, [2, 3]].mean()
```

## שאלה 2

טבלה של מדינות, יבשת שלידת אלו... צללן יבשת כמה מדינות יש ביבשת כך למספר האוכלוסייה  
צולל מ-10 מיליון.

יש טבלת world שמחזיקה את הנמנים. בנקודת זמנה חציה שבה האוכלוסייה הצלה מ-10 מיליון.

`big_countries = world[world['population'] > 10,000,000]`

צריך לראות יבשת מספר כמה מדינות זצולל מ-10 מיליון. לשים `group by` ומכריז לראות יבשת:

`big_per_cont = big_countries.groupby(['continent']).name.count()`  
← לזכור גם לכתוב `NaN`

## שאלה 3

נתן מוצא מאומן ע"י רגסיה לוגיסטית `model` (ע"י מתוצה דוגל). נכזה אלמנט נקודה  $(x, y)$   
 $x$  → `feature I`  
 $y$  → `feature II`  
• `features` - על כמה פיזרים אומנו את המוצל.

`model.predict(x)` →

מחשבה לקבל צימור מומנז מתמים  
צריך לתת מלכיה ב-מלכיות בגלל  
מל הפיזרים

• אם כותב מוצל מסומן לא צריך לכתוב `!`

`make_blobs` - יוצר צאטאן תצלה, מומנז מסומן.

• `features` - חיים לשים רשמה בתוך רשמה כי הוא מוצל לקבל מלכיה ב-מלכיות.