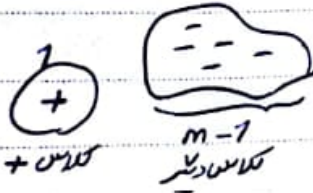


①

لایه‌های مختلف به این صورت است که یکی از کلاس‌ها را به عنوان کلاس + و بقیه کلاس‌ها را منفی (مثلاً SVM) و روی آن کلاس train می‌کنیم و این کار را به تکرار می‌کنیم تا تمام کلاس‌ها (مجموعاً $m-1$ بار) در نهایت باید به تعداد m (کلاس‌ها) کلاس‌های train کنیم.



در مرحله test، m عدد کلاس‌های بررسی و امتحان می‌شوند. برای هر داده‌ای که می‌گیریم، یک جواب توسط کلاس‌های دیگر تولید می‌شود. هر کلاس‌هایی که جواب + برای داده‌ای جدید تولید کند، آن داده متعلق به آن کلاس است (کلاس مثبت‌ترین کلاس‌ها).

این کار برای m کلاس‌های مختلف تکرار شود.

در واقع این کار برای هر کلاس، یک کلاس‌های داریم.

هم این روش، one vs all است.

این متد شبیه متد class validation است. (داده‌ها را به دو بخش تقسیم می‌کنیم و روی یکی از آن‌ها iteration صوت می‌گیریم و m بار برای تقسیم‌های مختلف، کلاس‌های آموزش داده می‌شود. پس m بار، m مسأله‌ی دو کلاس را حل می‌کنیم و لذا قابل انجام است.

②

به ازای $k=1$ ، واضح است که ترتیب قرار است.

به ازای $k > 1$ ، خواهیم داشت: یک frequent itemset به اندازه‌ی k داریم: $X = \{x_1, x_2, \dots, x_k\}$

این مجموعه را به دو زیرمجموعه‌ی $k-1$ و k می‌توان افزایش کرد:

$$\begin{cases} X_1 = \{x_1, x_2, \dots, x_{k-2}, x_{k-1}\} \\ X_2 = \{x_1, x_2, \dots, x_{k-2}, x_k\} \end{cases}$$

$X = X_1 \cup X_2 \rightarrow$ هر یک $k-1$ عضو دارند frequent هستند $\rightarrow X_1, X_2 \in F_{k-1}$

نکته: X در هر طریقی امتداد و تولید frequent itemset‌ها تولید می‌شود. و در مرحله بعدی حذف نمی‌شود.

پس اگر از یک مجموعه‌ی frequent آغاز کنیم، تمام زیرمجموعه‌های به دست آمده از آن، frequent خواهند بود.

در نتیجه، می‌تواند ترکیب دو زیرمجموعه‌ی $k-1$ تایی، که $k-2$ تایی اول آن‌ها یکسان است، کامل شود.

(هر یک frequent itemset‌های k به این ترتیب تولید می‌شوند)

$$H(s) = \frac{1}{2}$$

CAD	①			
CAB	②			
EABD	③			
EBD	④			
ECA	⑤			
ECAB	⑥			
E	⑦			
ECA	⑧			
ECB	⑨			
EC	⑩			

code

نمبر انگوری

A

رنگ

B

خوبی

C

عنکبی

D

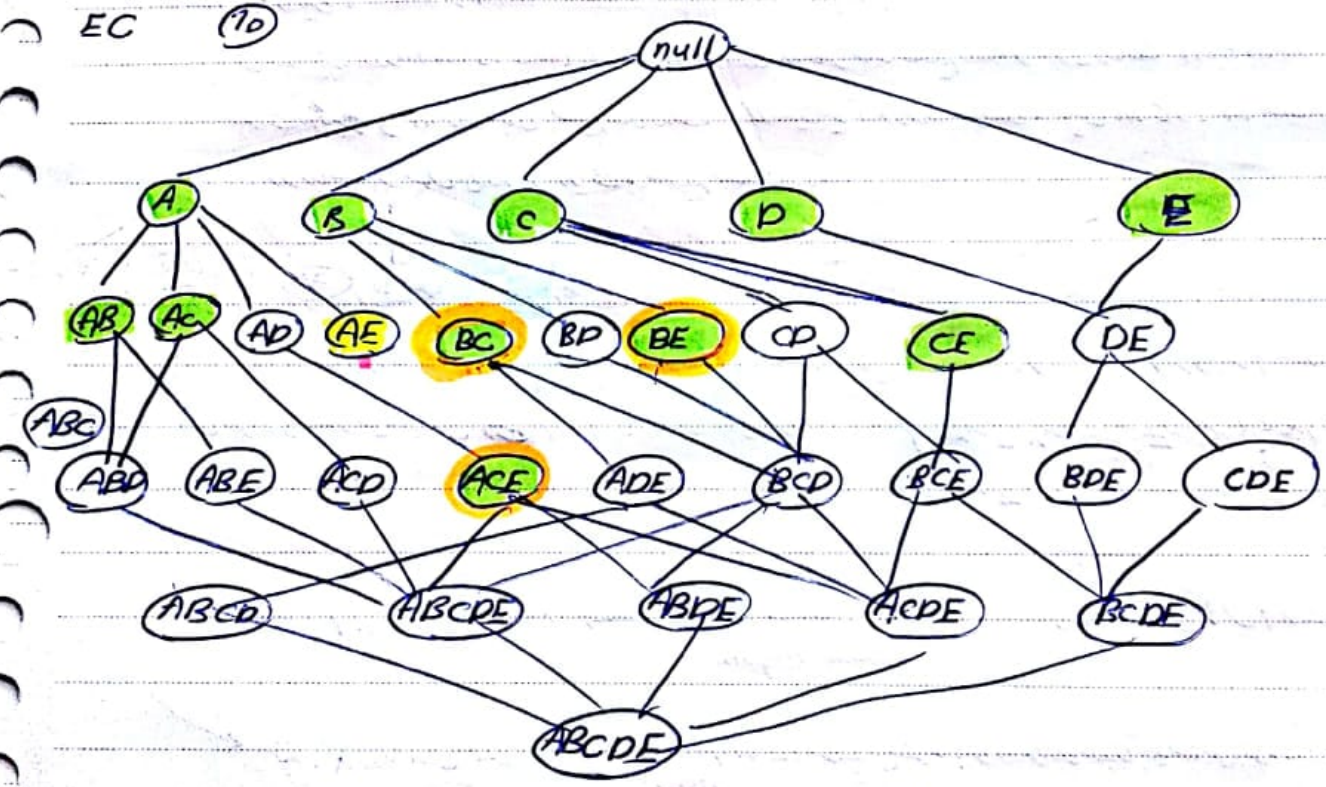
شعر

E

روانشناسی

③

الف



بجیب زنی:

maximal frequent itemsets → (M) مثبتي

close 1 " → (C) سبز

frequent But not maximal or close → (N) زي

PAPCO

فہم جامعہ مذہبی کم
کم زبان، روانشناسی کم

(ب) frequent items ہے۔ یہ رواں اختصار ہوتا ہے:

کم مذہبی، زبان کم
کم روانشناسی، زبان کم

کم مذہبی، روانشناسی، زبان کم
روانشناسی، مذہبی،

(ج) تمام قواعد انجینی استعمل شدہ از سمت ب

روانشناسی، مذہبی، زبان کم

روانشناسی، زبان کم، مذہبی

مذہبی، زبان کم، روانشناسی

زبان کم، روانشناسی

روانشناسی، زبان کم

زبان کم، مذہبی

مذہبی، زبان کم

④ اولین فرد نقطه فاصله اقلیدسی را مسدود کنیم

شماره دانشجویی	1	2	3	4
1	0	3.3	1.7	3.5
2	3.3	0	2.8	4.24
3	1.7	2.8	0	2.4
4	3.5	4.2	2.4	0

* برای خوشه بندی سه مرحله داریم: 1. فاصله بین دو نقطه را محاسبه می کنیم (هر داده یک خوشه است) 2. اگر فاصله بین دو خوشه کمتر از مقدار مشخص شده باشد، آن دو خوشه را merge می کنیم. 3. این کار را تا زمانی که فقط یک خوشه باقی بماند تکرار می کنیم.

$$1,2 \rightarrow \sqrt{(15-17)^2 + (20-17)^2 + (78-17)^2} = \sqrt{4+9+1} = \sqrt{14}$$

$$1,3 \rightarrow \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$$

$$1,4 \rightarrow \sqrt{9 + 4 + 0} = \sqrt{13}$$

در مرحله اول، فاصله بین دانشجوی 1 و 3 (کمترین مقدار است).

پس آن دو را ادغام می کنیم.

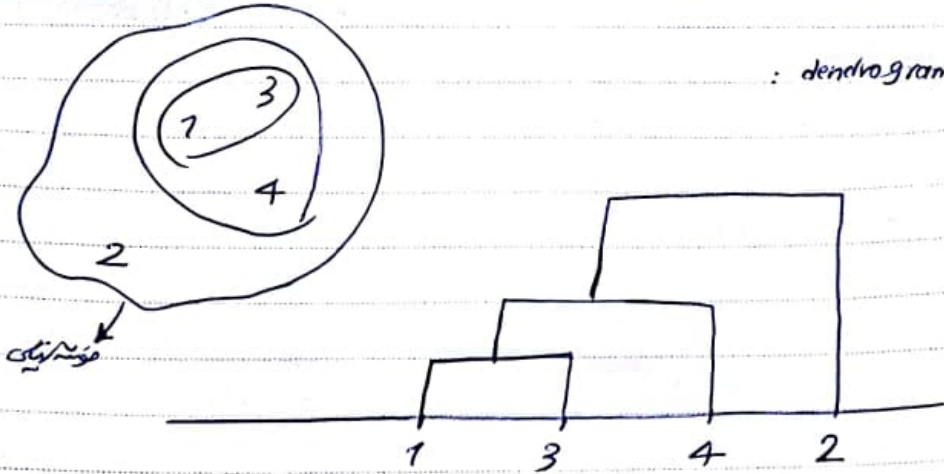
	1,3	2	4
1,3	0	2.82	2.44
2	2.82	0	
4	2.44	4.24	0

فاصله بین خوشه 1 و 3 و 4 (کمترین است) و 2. پس این دو ترکیب می شوند.

	1,3 و 4	2
1,3 و 4	0	2.82
2	2.82	0

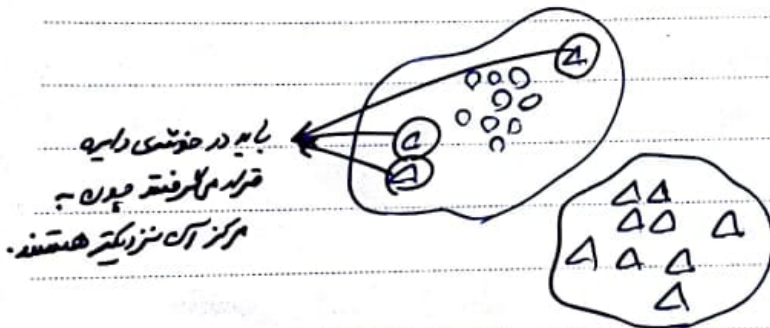
در انتها خوشه‌های 4 و 3 و 2 با 2 ترکیب می‌شوند و یک خوشه نهایی می‌شوند.

شماره dendrogram :



5) الگوریتم به کار رفته برای خوشه‌بندی، **می‌تواند Kmeans باشد.**

چون این الگوریتم بر اساس نزدیک‌ترین مراکز خوشه‌ها کار می‌کند و در این صورت باید خوشه‌بندی ۳ شکل زیر انجام می‌شد:



و Δ ها اضافه‌اند.

احتمالاً این روش خوشه‌بندی بر مبنای خوشه‌بندی فازی صورت گرفته است و از توزیع داده‌ها استفاده کرده است و با تحقیق این کارها، به خوشه‌بندی بالا رسیه است.

پس **خوشه‌بندی می‌تواند EM باشد.** (خوشه‌بندی بر اساس پارامترهای توزیع)
 ساده