Subject
Date

مبانی هوش مصنوعی

ابزارهای عامل — روش های جست و جو
- ۱- همه فضای حالت را بررسی می کند.
- ۲- نقطه جاهایی از فضای حالت را نمی گردد که بهینه است جواب آن نیست مثل A^*
- ۳- نقطه جاهایی را می گردد که امید دارد جواب آن جا باشد.

منطق (عامل های منطقی)

هوش مصنوعی — هوش کلاسیک ← روش های احتمالاتی، روش های جست و جو

هوش محاسباتی ← سیستم های فازی، الگوریتم های تکاملی، شبکه های عصبی

منطق
- رتبه اول
- گزاره ای
- فازی
- ...

عامل هوشمند یعنی معیار کارایی را بهتر کردن.

انواع تغییرها
- قطعی
- تصادفی (احتمالاتی)
- فازی (نادقیق)

تغییر نادقیق: نمی توانیم یک مقدار مشخص برای آن تعیین کنیم پس برای آن یک دامنه (محدوده) تعیین می کنیم. مثلاً برای قد کوتاه می گوییم قد بین ۱۱۰ تا ۱۵۰ سانتی متر.

می توانیم برای هر یک از اعداد بین ۱۱۰ تا ۱۵۰ یک مقدار تعلق تعریف کنیم و با عددی بین ۰ و ۱ مشخص کنیم. مثلاً بگوییم قد ۱۲۰ با مقدار تعلق ۰/۷ کوتاه است.

تغییر تصادفی: مثلاً یک تاس را قبل از اینکه بیندازیم نمی توانیم بگوییم چند می آید.

تغییر تصادفی با تعداد سروکار دارد، هرچه تعداد بیشتر باشد، بهتر است.

PAPCO

Scanned by CamScanner

Subject
Date

وقتی می‌گوییم با احتمال ۷۰٪ فردا باران می‌آید، منظورمان چیست؟

یعنی با توجه به آمارهای گذشته اگر ۱۰۰۰ بار شرایط آب و هوایی شبیه فردا اتفاق افتاده، ۷۰۰ بار باران آمده است.

* وقتی ۱۰۰ بار تاس را می‌اندازیم تعداد دفعاتی که ۱، ۲، ۳، ۴، ۵، ۶ می‌آید اگر با هم جمع کنیم، ۱۰۰ می‌شود. اما در یک متغیر نادقیق لزومی ندارد جمع تعلق‌ها ۱ شود. مثلاً تعلق قد ۱۲۰ به قد کوتاه اگر ۰٫۶ باشد، تعلق ۱۳۰ می‌تواند ۰٫۶ باشد و از لحاظ فیزیکی ارتباطی به هم ندارند.

نظریه مجموعه‌های فازی تعمیمی از نظریه مجموعه‌های کلاسیک است.

مجموعه ← تعدادی از اشیا که کنار هم قرار گرفته باشند. به هر شیء یک عضو گفته می‌شود

تعداد اعضا می‌تواند قابل شمارش (گسسته) یا غیرقابل شمارش (پیوسته) باشد.

مجموعه با حروف بزرگ لاتین مشخص می‌شود. با ، یا + اعضا از هم جدا می‌شوند

$$A = \{a, b, c\} \equiv A = \{a + b + c\}$$

مجموعه مرجع تمام اشیای ممکن برای حل مسئله مورد نظر ما را در بر می‌گیرد.

PAPCO

Scanned by CamScanner

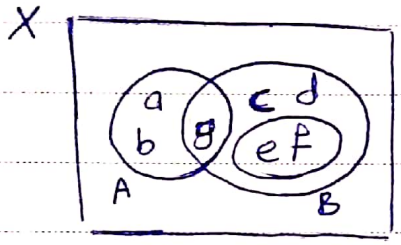
Subject
Date

مجموعه مرجع با حرف بزرگ X نشان داده می‌شود.

زیرمجموعه :

$A \subset B \quad . \quad \{x \mid x \in A \Rightarrow x \in B\}$

نمایش گرافیکی (ون) :



$C \subset B$
$A \subset X$
$B \subset X$

تعریف تابع تعلق :

$\mu \longrightarrow$ تابع تعلق

* نمایش با تابع تعلق :

$$\mu_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

* $A = \left\{ \frac{\mu_A(a)}{a}, \frac{\mu_A(b)}{b}, \frac{\mu_A(g)}{g} \right\} = \left\{ \frac{1}{a}, \frac{1}{b}, \frac{1}{g} \right\}$

$B = \left\{ \frac{1}{g}, \frac{1}{d}, \frac{1}{e}, \frac{1}{f} \right\} \qquad C = \left\{ \frac{1}{e}, \frac{1}{f} \right\}$

* نمایش برداری :

$X = \left\{ \frac{1}{a}, \frac{1}{b}, \frac{1}{g}, \frac{1}{e}, \frac{1}{f}, \frac{1}{d} \right\}$

مجموعه مرجع ترتیب را مشخص می‌کند.

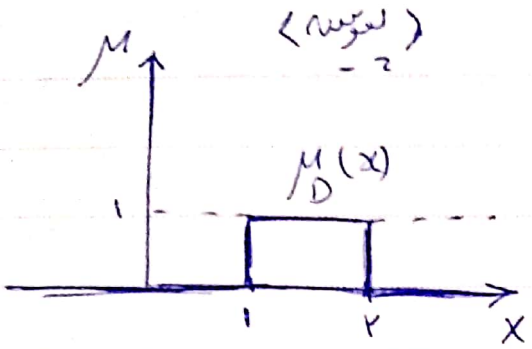
$A = \{1, 1, 1, 0, 0, 0\}$

$B = \{0, 0, 1, 1, 1, 1\} \qquad C = \{0, 0, 0, 1, 1, 0\}$

PAPCO

Scanned by CamScanner

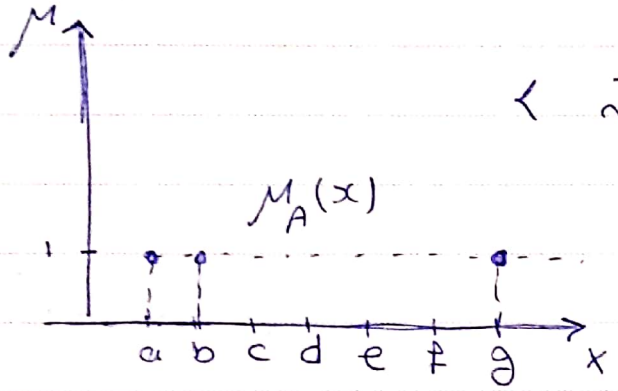
Subject
Date



$D=\{x \mid x\in R \ , \ 1\le x\le 2\}$

نشانه نمودار دن ، نمایش لازنگی

این گونه خواهد بود :



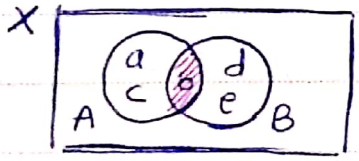
(گسسته)

عملیات در مجموعه ها :

$A\cap B$

چگونه نمایش داریم :

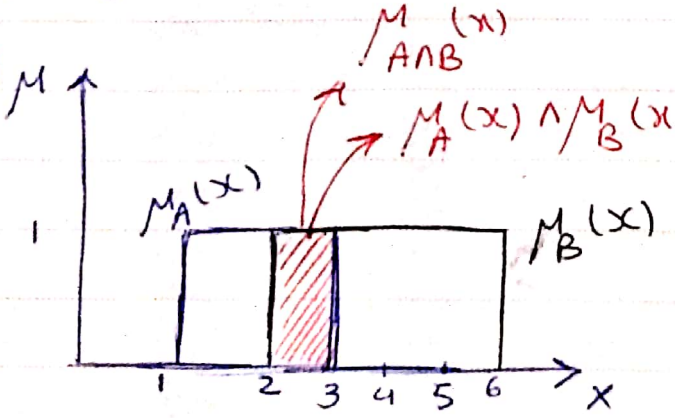
① $A\cap B=\{x \mid x\in A \ \wedge \ x\in B\}$

② 

$A=\{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\}$

$B=\{\frac{1}{b}+\frac{1}{d}+\frac{1}{e}\}$

③ $\mu_{A\cap B}(x)=\mu_A(x)\wedge\mu_B(x)=\text{Min}\{\mu_A(x),\mu_B(x)\}$

④ 

$A=\{x \mid x\in R \ , \ 1\le x\le 3\}$

$B=\{x \mid x\in R \ , \ 2\le x\le 7\}$

PAPCO

Scanned by CamScanner

Subject
Date

نکته ∞

$\wedge$ $\wedge$

* دلیل اینکه "و" و اشتراک علامت هایشان فرق می کند، از لحاظ زبانی است.

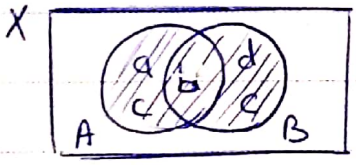
در مجموعه ها مفهوم اشتراک معنا دارد و هنگامی که با مقادیر تعلق کار داشته باشیم، "و" و "یا" کارکرد دارد. (ولی در عمل از لحاظ نمادگی $\wedge$ و $\cap$ یکی است).

$A \wedge B$ ✗ غلط است

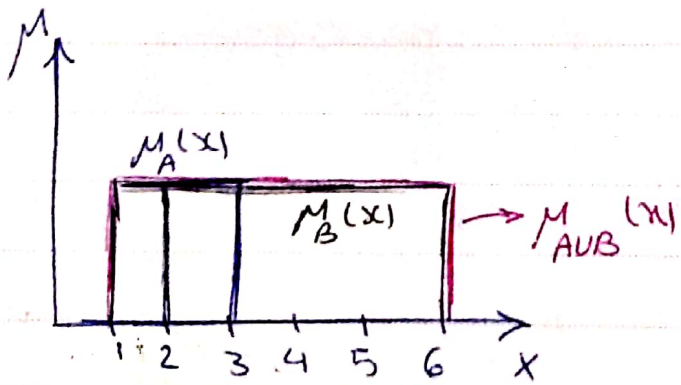
$\mu_A(x) \wedge \mu_B(x) \equiv A \cap B$

$A \cup B$

① $A \cup B = \{x \mid x \in A \vee x \in B\}$

② 

③ $\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \text{Max}\{\mu_A(x), \mu_B(x)\}$

④ 

PAPCO

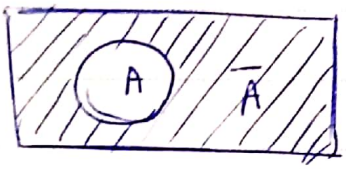
Scanned by CamScanner

Subject
Date

$\bar{A}$

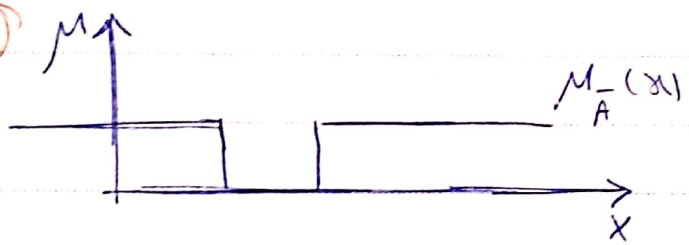
① $\bar{A} = \{x \mid x \in X \wedge x \notin A\}$

② X



③ $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$

④



به یک سری عملگرها کلاس نرمال گفته می‌شود به شرطی که خاصیت جابه‌جایی، شرکت‌پذیری، توزیع‌پذیری و دمورگان را داشته باشند. (عملگرهای اشتراک، اجتماع، تمم)

8✓

در کلاس نرمال این است که تعمیمی بر نظریه مجموعه‌ها است

$$\begin{cases} A \cap B \equiv \mu_B(x) \wedge \mu_A(x) \\ \mu_A(x) \vee \mu_B(x) \equiv B \cup A \end{cases}$$

خاصیت جابه‌جایی

$$\begin{cases} (A \cap B) \cap C \equiv \mu_A(x) \wedge (\mu_B(x) \wedge \mu_C(x)) \\ \mu_A(x) \vee (\mu_B(x) \vee \mu_C(x)) = (A \cup B) \cup C \end{cases}$$

خاصیت شرکت‌پذیری

PAPCO

Scanned by CamScanner

Subject
Date

خاصیت توزیع پذیری

$$A \cap (B \cup C) \equiv (\mu_A(x) \wedge \mu_B(x)) \vee (\mu_A(x) \wedge \mu_C(x))$$

$$\mu_A(x) \vee (\mu_B(x) \wedge \mu_C(x)) = (A \cup B) \cap (A \cup C)$$

دمورگان

$$(A \cup B)' = \mu_{A'}(x) \wedge \mu_{B'}(x)$$

$$1 - (\mu_A(x) \wedge \mu_B(x)) \equiv (1 - \mu_A(x)) \vee (1 - \mu_B(x))$$

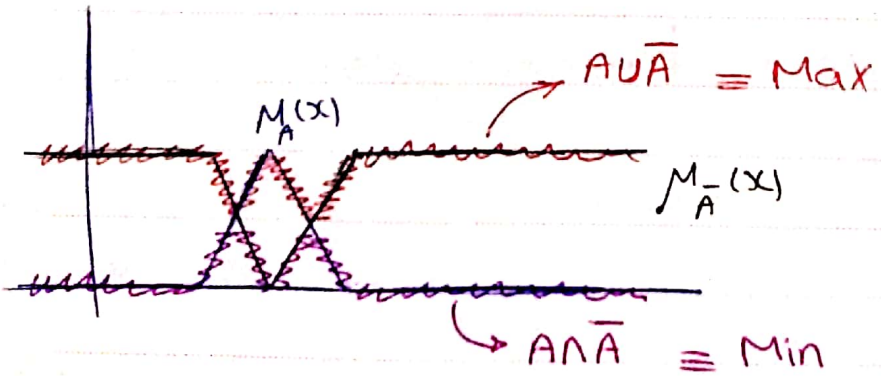
اگر اشتراک را ضرب بگیریم و اجتماع را Max بگیریم و متمم همان $1 - A$ ، طبق دمورگان است

$$\left.\begin{aligned} a \wedge b &= ab \\ a \vee b &= a + b - ab \\ \bar{a} &= 1 - a \end{aligned}\right\} \Longrightarrow$$ این کلاس دمورگان است.

دمورگان

* اگر مقادیر بین ۰ و ۱ باشند $A \cup \bar{A} = X$ برقرار نیست
* " " $A \cap \bar{A} = \phi$ می‌تواند برقرار نباشد.

* این خواص فرمی هستند.



PAPCO

Scanned by CamScanner

Subject
Date ۹۷/۱۱/۲۱

جلسه سوم

$$A \times B = \{(x,y) \mid x \in A \ , \ y \in B\}$$

اگر مجموعه‌ی مادر A و B به ترتیب X و Y باشند، آنگاه مجموعه مادر $A \times B$ ، $X \times Y$ خواهد بود.

تعداد اعضا

$|A| = m \quad , \quad |B| = n \Longrightarrow |A \times B| = m \times n$

$$\underset{m_1}{A_1} \times \underset{m_2}{A_2} \times \underset{m_3}{A_3} \times \cdots \times \underset{m_n}{A_n} = \{(x_1, x_2, \dots, x_n) \mid x_1 \in A_1 , x_2 \in A_2 , \dots , x_n \in A_n\}$$

** هر رابطه زیر مجموعه ای از مجموعه‌ی حاصل از ضرب کارتزین است.

$$\begin{array}{c|ccc} & 12 & 15 & 17 \\ \hline \text{حسن} & 0 & 1 & 0 \\ \text{علی} & 0 & 1 & 0 \\ \text{تقی} & 0 & 0 & 1 \end{array}$$

$$M_R(x,y) = \begin{cases} 1 & (x,y) \in R \equiv xRy \\ 0 & (x,y) \notin R \equiv x \not R y \end{cases}$$

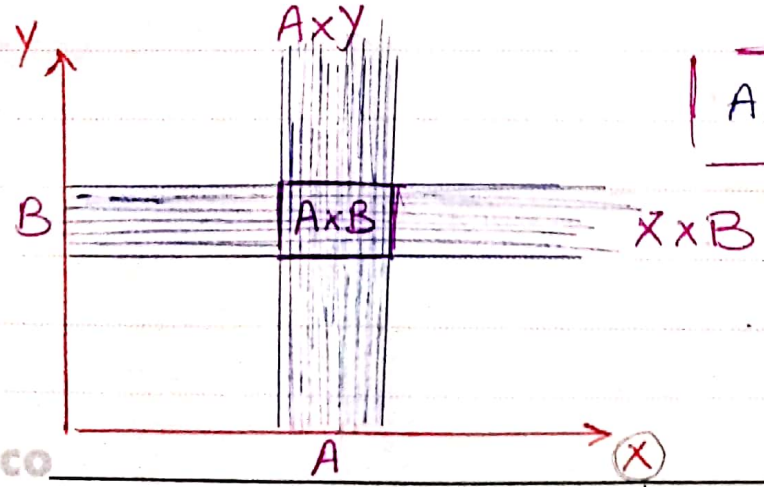
(برای n بعدی نیز تعمیم دارد)

* برای نمایش برداری اعضای مجموعه در این جا از ماتریس استفاده می‌شود.

* در صورتی می‌توان دو مجموعه را اشتراک یا اجتماع گرفت که مجموعه‌ی مادرشان یکی باشد.

راه حل (توسعه استوانه ای)!!

$A \times Y \ , \ X \times B \Longrightarrow$ توسعه‌ی استوانه ای



$$\boxed{A \times B = (A \times Y) \cap (X \times B)}$$

چون این ها همه مرجع شده اند (با توسعه استوانه ای) می‌توان اشتراک گرفت.

با توسعه استوانه ای A ، انتظار بعد ۱ به ۲ ان اضافه می‌شود و داریم $A \times Y$)

مجموعه مرجع A

Scanned by CamScanner

Subject
Date

$A \times B$ = اشتراک توزیع استوانه ای

* برای اشتراک ، راه میان بر همان ضرب کارتزین است . اما در $\cup$ راه میان بر نداریم

و باید حتماً اول مجموعه های مرجع را یکسان کنیم . $(A \times Y) \cup (X \times B)$

مجموعه مرجع

$$A \xrightarrow{R_1} B \xrightarrow{R_2} C$$

$X \quad Y \quad Z$

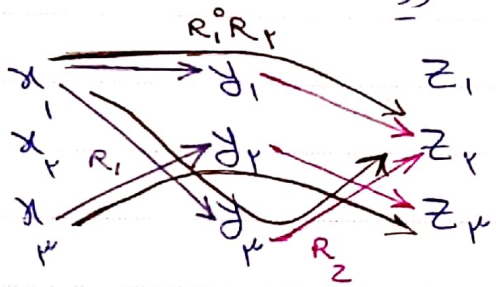
ترکیب

$$\underset{X \times Z}{R_3} = \underset{X \times Y}{R_1} \circ \underset{Y \times Z}{R_2} = \underset{\text{تصویر}}{\cup}\left(R_1 \underset{\uparrow}{\cap} R_2\right) = \underset{\text{تصویر}}{\cup}\underbrace{\left(\overset{X \times Y \times Z}{R_1 \times Z}\right) \cap \left(\overset{X \times Y \times Z}{X \times R_2}\right)}_{\text{یکی شدن مرجع}}$$

(برای کاهش یک بعد در این مثال)

ترکیب Max.Min

(به جای جمع Max و به جای ضرب Min)



تصویر نسبت به x ها

$$\begin{array}{c|ccc} R_1 & y_1 & y_2 & y_3 \\ \hline x_1 & 1 & 0 & 1 \\ x_2 & 0 & 0 & 0 \\ x_3 & 0 & 1 & 0 \\ \hline & 1 & 1 & 1 \end{array}$$

$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

تصویر نسبت به y ها

نکته : هنگامی که از یک رابطه نسبت به تمامی ابعاد تصویر بگیریم و سپس تصاویر را توسط استوانه ای دهیم و اشتراک بگیریم ، لزوماً به رابطه ی اول نمی رسیم .

نمایش با تابع تعلق :

$$\mu_{R_3}(x_1, z_2) = \left(\mu_{R_1}(x_1, y_1) \wedge \mu_{R_2}(y_1, z_2)\right) \vee \left(\mu_{R_1}(x_1, y_2) \wedge \mu_{R_2}(y_2, z_2)\right) \vee \left(\mu_{R_1}(x_1, y_3) \wedge \mu_{R_2}(y_3, z_2)\right)$$

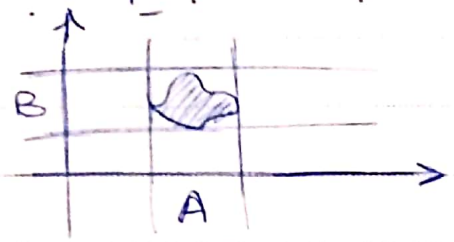
برای همه ی زوج ها باید تکرار شود تا اگر یک راه موجود بود ، بفهمیم رابطه بین A و C نیز وجود دارد .
که برای فهمیدن آن راه استفاده از ماتریس بهتر است .

Scanned by CamScanner

Subject
Date

کاربرد زبان بین Max, Min ؟

* می توانیم ماتریس های R_1 و R_2 را در هم ضرب کنیم و به جای ضرب اشتراک بگذاریم و به جای جمع اجتماع بگذاریم.



* ترکیب خاصیت جابه جایی ندارد.

$R_1 \times R_2$

$\hookrightarrow$ تعداد ستون R_1 = تعداد سطر R_2

و

مجموعه ی وسط بین دو رابطه و ترتیب نشان ملی باشد، در غیر این صورت عملیات غلط است.

$R_1 \times R_2 \Longrightarrow$ هر سطر R_1 در هر ستون R_2 ضرب می شود، بین آنها اجتماع بگیر

جمع و ضرب در استک

$S_{ds}(a,b) = \{a \text{ if } b=0,\ b \text{ if } a=0,\ 1 \text{ else}\}$

$t_{dp}(a,b) = \{a \text{ if } b=1,\ b \text{ if } a=1,\ 0 \text{ else}\}$

$\mu_{R \circ S} = Max(Min(\mu_R, \mu_S))$ ضرب ← Min ، جمع ← Max

$\mu_{R \circ S} = Max(\mu_R \circ \mu_S)$ ضرب ← $\bullet$ (dot) ، جمع ← Max

PAPCO

Scanned by CamScanner

Subject
Date

شرایط مرزی

اجتماع فازی - اس نرم

$S(0,a) = S(a,0) = a$

$S(1,1) = 1$

0	0	0
0	1	1
1	0	1
1	1	1

۰ عضو خنثی اجتماع است.

$t(1,a) = t(a,1) = a$

$t(0,0) = 0$

اشتراک فازی - تی نرم

0	0	0
0	1	0
1	0	0
1	1	1

۱ عضو خنثی اشتراک است.

عملگرهای توسط گیری ← از Min تا وسط

از وسط تا Max

max → کوچکترین اجتماع ممکن

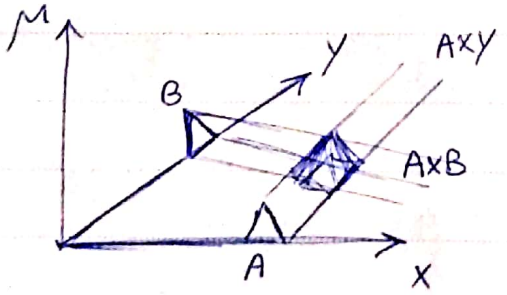
جمع دراستیک → " " بزرگترین " = 1

min → بزرگترین ضرب ممکن

ضرب دراستیک → " " کوچکترین " = 0

(نزدیک بودن، فازی است! (چه قدر نزدیک است؟))

توسعه استوانه ای ⇐ ضرب کارتزین مجموعه فازی در مجموعه ای قطعی که نداریم!

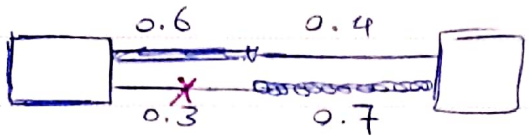


PAPCO

Scanned by CamScanner

Subject
Date

توصیه فیزیکی ترکیب Max . Min :



تا وقتی بین Min ها، Max وجود دارد،

ارتباط برقرار است.

اول طناب ۰/۳ پاره می‌شود، اگر باز طنابی قرار باشد پاره شود، ۰.۴ است پس تا وقتی طناب

۰.۴ هست، ارتباط برقرار است.

* هر وقت مشتق گیری بخواهیم می‌توانیم از Max. (max dot) استفاده کنیم.

(به جای اشتراک از ضرب استفاده می‌شود)

توسعه استوانه ای

$$\mu_{A\times B}(x,y) = \mu_A(x,y) \wedge \mu_B(x,y)$$

* حاصل ضرب کارتزین مجموعه های فازی برای با اشتراک توسعه استوانه ای آن ها است.

مثال/

$$A = \left\{ \frac{0.1}{a} + \frac{0.7}{b} \right\} \qquad Y = \left\{ \frac{1}{1} + \frac{1}{2} + \frac{1}{3} \right\}$$

$$A' = A\times Y = \left\{ \frac{0.1}{(a,1)}, \frac{0.1}{(a,2)}, \frac{0.1}{(a,3)}, \frac{0.7}{(b,1)}, \frac{0.7}{(b,2)}, \frac{0.7}{(b,3)} \right\}$$

(کاری به تعلق مجموعه فرضی نداره، همه اعضا فرضی را اضافه می‌کنیم)

$$B = \left\{ \frac{0.8}{2}, \frac{0.5}{3} \right\} \qquad X = \{a, b, c\}$$

$$B' = X\times B = \left\{ \frac{0.8}{(a,2)} + \frac{0.8}{(b,2)} + \frac{0.8}{(c,2)} + \frac{0.5}{(a,3)} + \frac{0.5}{(b,3)} + \frac{0.5}{(c,3)} \right\}$$

Scanned by CamScanner

Subject
Date

$$A' \cap B' = \left\{ \frac{0.1}{(a,2)} + \frac{0.1}{(a,3)} + \frac{0.7}{(b,2)} + \frac{0.5}{(b,3)} \right\}$$

$$R = A \times B = \left\{ \frac{0.1}{(a,2)} + \frac{0.1}{(a,3)} + \frac{0.7}{(b,2)} + \frac{0.5}{(b,3)} \right\}$$

⟸ نتیجه یکسان شد.

برعکس توسعه استوانه ای که بُعد اضافه می کند، تصویر روی بُعد کم می کند.

→ $A'' = \left\{ \frac{0.1}{a} , \overset{\cup}{\frac{0.7}{b}} \right\}$ $\quad B'' = \left\{ \frac{0.7}{2} + \frac{0.5}{3} \right\}$

تصویر روی A

توسعه استوانه ای

$$B''' = \left\{ \frac{0.7}{(a,2)} + \frac{0.7}{(b,2)} + \frac{0.7}{(c,2)} + \frac{0.5}{(a,3)} + \frac{0.5}{(b,3)} + \frac{0.5}{(c,3)} \right\}$$

$$A''' = \left\{ \frac{0.1}{(a,1)} + \frac{0.1}{(a,2)} + \frac{0.1}{(a,3)} + \frac{0.7}{(b,1)} + \frac{0.7}{(b,2)} + \frac{0.7}{(b,3)} \right\}$$

$$\Rightarrow A'' \cap B''' = \left\{ \frac{0.1}{(a,2)} , \frac{0.1}{(a,3)} , \frac{0.7}{(b,2)} , \frac{0.5}{(b,3)} \right\}$$

مثال //

$X = \{a, b, c\}$ $\quad Y = \{1, 2, 3\}$

$A = \left\{ \frac{0.8}{a} , \frac{0.1}{b} \right\}$ $\quad B = \left\{ \frac{0.7}{2} , \frac{0.4}{3} \right\}$

$$A' = \left\{ \frac{0.8}{a,1} , \frac{0.8}{a,2} , \frac{0.8}{a,3} , \frac{0.1}{b,1} , \frac{0.1}{b,2} , \frac{0.1}{b,3} \right\}$$

مرجع $X \times Y$

توسعه استوانه ای A

PAPCO

Scanned by CamScanner

Subject
Date

به صورت ماتریسی: (توسعه استوانه ای A)

	1	2	3
a	0.8	0.8	0.8
b	0.1	0.1	0.1
c	0	0	0

$$\mu_{A'}(x,y) = \mu_A(x,y)$$
$$y \in Y$$

توسعه استوانه ای B: $B' = \left\{ \frac{0.7}{a,2}, \frac{0.7}{b,2}, \frac{0.7}{c,2}, \frac{0.4}{a,3}, \frac{0.4}{b,3}, \frac{0.4}{c,3} \right\}$

$X \times Y$

	1	2	3
a	0	0.7	0.4
b	0	0.7	0.4
c	0	0.7	0.4

$$\Longrightarrow R = A' \cap B' = B' \cap A' = \left\{ \frac{0.7}{a,2}, \frac{0.1}{b,2}, \frac{0.4}{a,3}, \frac{0.1}{b,3} \right\} = A \times B$$

* اون صفرهایی که هستند نمی توانند در اشتراک وجود داشته باشند (اشتراک با صفر ⇐ صفر)

$$\xrightarrow{\text{راه دیگر}} A \times B = \left\{ \frac{0.7}{a,2}, \frac{0.4}{a,3}, \frac{0.1}{b,2}, \frac{0.1}{b,3} \right\}$$

$$\left.\begin{array}{l} A' = A \times Y \\ B' = X \times B \end{array}\right\} \Longrightarrow$$ تعریف توسعه استوانه ای با استفاده از ضرب کارتزین

* در حالتی که تنها مجموعه A دارای مرجع X باشد و B هم دارای مرجع X باشد باز هم ممکن است بخواهیم اشتراک A و B را در مرجع $X \times X$ محاسبه کنیم، در این حالت باز توسعه استوانه ای می دهیم.

Scanned by CamScanner

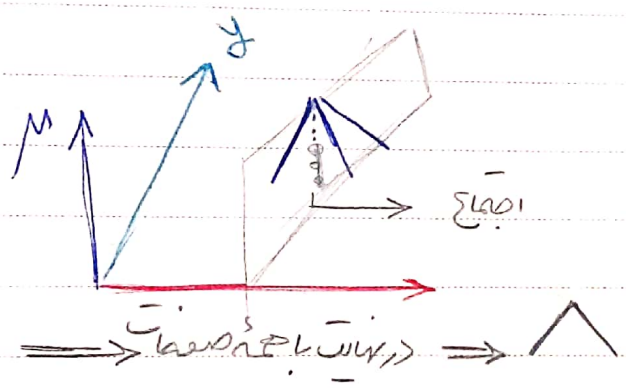
Subject
Date

خیلی اوقات چون تعلق در مجموعه مرجع 1 است، آن را نمی‌نویسیم.

تصویر کردن:

فرض می‌خواهیم تصویر را روی X بدست آوریم، صفحه‌هایی موازی MOY رسم می‌کنیم

این صفحه رابطه را در نقاطی قطع می‌کند و Max (اجتماع) این نقاط را بدست می‌آوریم.



تعریف تصویر اجتماع مقادیر تعلق است.

$$A = \bigcup_{y \in Y} R$$

$$\mu_A(x) = \bigvee_{y \in Y} \mu_R(x,y)$$

ادامه مثال قبل))

$$A'' = \bigcup R = \left\{ \frac{0.7}{a}, \frac{0.1}{b} \right\} = \bigvee_{y \in Y} \mu_R(x,y) = \mu_A(x)$$

تصویر R روی A

$$B'' = \bigcup R = \left\{ \frac{0.7}{2}, \frac{0.4}{3} \right\} = \bigvee_{x \in X} \mu_R(x,y) = \mu_B(y)$$

تصویر R روی B

به صورت ماتریسی:

	1	2	3
a	0.7	0.7	0.4
b	0.1	0.1	0.1

PAPCO

Scanned by CamScanner

Subject
Date

توسعه استوانه A'' و B'' :

$$\begin{array}{cccc} & 0.7 & 0.4 & \\ 0.7 & \begin{matrix}0.7\\0.7\end{matrix}\Big)^{0.7} & \begin{matrix}0.7\\0.4\end{matrix}\Big)^{0.4} & \\ 0.1 & \begin{matrix}0.1\\0.7\end{matrix}\Big)^{0.1} & \begin{matrix}0.1\\0.4\end{matrix}\Big)^{0.1} & \end{array}$$

$$A''' = \left\{ \frac{0.7}{a,1}, \frac{0.7}{a,2}, \frac{0.7}{a,3}, \frac{0.1}{b,1}, \frac{0.1}{b,2}, \frac{0.1}{b,3} \right\}$$

$$B''' = \left\{ \frac{0.7}{a,2}, \frac{0.7}{b,2}, \frac{0.7}{c,2}, \frac{0.4}{a,3}, \frac{0.4}{b,3}, \frac{0.4}{c,3} \right\}$$

$$A'' \times B'' = A''' \cap B''' = B''' \cap A''' = \left\{ \frac{0.7}{a,2}, \frac{0.4}{a,3}, \frac{0.1}{b,2}, \frac{0.1}{b,3} \right\}$$

← نتیجه با R یکسان شد چون جداپذیر است.

* رابطه جداپذیر از روی تصاویرش قابل بازسازی است.

رابطه ای جداپذیر است که بین متغیرهایش وابستگی وجود نداشته باشد. (همه متغیرها مستقل باشند)

مثلاً در دایره $x^2 + y^2 = 1$ ← هرکدام از x و y را داشته باشیم دیگری بدست می آید

در نتیجه مستقل نیستند (وقتی استوانه را تصویر می کنیم در واقع این وابستگی از بین می رود به خاطر همین بازسازی تصاویر شکل اولیه را نمی دهد)

PAPCO

Scanned by CamScanner

Subject
Date

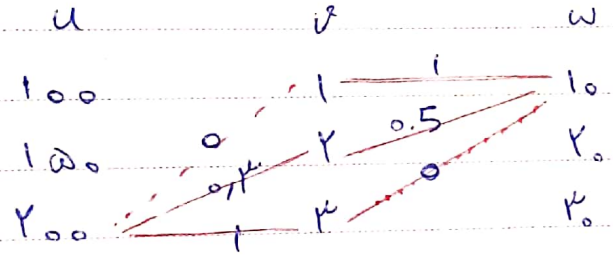
$R: U \to V$, $S: V \to W$; X, Y, Z

$$R \circ S = \underset{\text{تصویر}}{\cup} (R' \cap S')$$

$$R' = R \times Z$$
$$S' = X \times S$$

$$\Longrightarrow \boxed{R \circ S = \underset{\text{تصویر}}{\bigcup_{y \in Y}} (R \times Z \cap X \times S)}$$

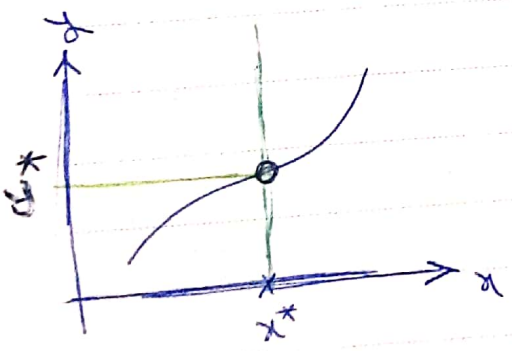
روی صفحه $X \times Z$ تصویر می‌شود



می‌خواهیم ببینیم بین ۱۰۰ و ۲۰۰ رابطه وجود دارد؟

$$\mu_{R \circ S} = \left(\mu_R(u_3, v_1) \wedge \mu_R(v_1, w_1)\right) \vee \left(\mu_R(u_3, v_2) \wedge \mu_R(v_2, w_1)\right) \vee$$
$$\left(\mu_R(u_3, v_3) \wedge \mu_R(v_3, w_1)\right)$$

توابع و روابط



۱- توسعه استوانه‌ای ورودی را در اختیار خروجی بدست می‌آوریم.

۲- تصویر اشتراک بدست آمده را با رابطه بین ورودی و خروجی بدست می‌آوریم.

۳- تصویر اشتراک بدست آمده را روی خروجی بدست می‌آوریم.

مرحله ۱: $(A \times Y)$ ، تابع تعلق: $\mu_A(x, y)$ ، $y \in Y$

PAPCO

Scanned by CamScanner

Subject
Date

مرحله ۲ : $(A\times Y)\cap R$ ، تابع تعلق : $\underset{y\in Y}{\mu_A(x,y)\wedge\mu_R(x,y)}$

مرحله ۳ : $\bigcup_{x\in X}\left[\overset{X\times Y}{(A\times Y)}\cap \overset{X\times Y}{R}\right]=B \Longrightarrow A\circ R=B$

تابع تعلق : $\underset{y\in Y}{\bigvee_{x\in X}\mu_A(x,y)\wedge\mu_R(x,y)}=\mu_B(x,y)$

نتیجه :

اگر رابطه بین ورودی و خروجی یک سیستم را داشته باشیم، این سیستم شناخته شده است.

از این ۳ تا هر ۲ تایی را داشته باشیم، سومی بدست می‌آید :

$R \quad A \quad B$

$A, R \Longrightarrow B = A\circ R$

$B, R \Longrightarrow A = R\circ B$ $\quad \underset{X\times Y}{R}\cap\underset{X\times Y}{(X\times B)}$

$A, B \Longrightarrow$ R بدست می‌آید (در آینده بحث می‌شود)

$A\circ R \begin{cases} = B & (\text{R جدا پذیر}) \\ \neq B & (\text{R جدا ناپذیر}) \end{cases}$

حتی در حالتی که $A\circ R\neq B$ است، اثبات می‌شود که این B، بهترین تقریب B است

اثبات: $A_1\circ R = B_1$ اثبات اینکه فرمول $A\circ R=B$ خوب است.

$A_1 \to A$ (A به A نزدیک شود)

$B_1 \to B$ $\quad B_1$ ای که فرمول می‌دهد، بهترین تقریب برای B است.

$B = A\circ R$

$\hookrightarrow A = A_1\times A_2\times\cdots\times A_n$

$B = B_1\times B_2\times\cdots\times B_m$

PAPCO

Scanned by CamScanner

Subject
Date

مثال //

$y = x^2$

$A = \left\{ \frac{0.2}{1}, \frac{1}{2}, \frac{0.3}{3} \right\}$

$x = A$

$y = B$

R = B

A

0.2	1	0.3
1	2	3

	1	4	9
1	1 (0.2)	0 (0.2)	0 (0.2)
2	0 (1)	1 (1)	0 (1)
3	0 (0.3)	0 (0.3)	1 (0.3)
	0.2	1	0.3

$\Rightarrow B = \left\{ \frac{0.2}{1}, \frac{1}{4}, \frac{0.3}{9} \right\}$

مثال //

$y = 2x_1^2 - x_2 + 5$

$x_1 = A_1 \qquad A_1 = \left\{ \frac{0.2}{1} + \frac{1}{2} + \frac{0.3}{3} \right\}$

$x_2 = A_2 \qquad A_2 = \left\{ \frac{0.5}{4} + \frac{1}{5} + \frac{0.4}{6} \right\}$

$y = B$

$\Rightarrow A = A_1 \times A_2 = \left\{ \frac{0.2}{1,4}, \frac{0.2}{1,5}, \frac{0.2}{1,6}, \frac{0.5}{2,4}, \frac{1}{2,5}, \frac{0.4}{2,6}, \frac{0.3}{3,4}, \frac{0.3}{3,5}, \frac{0.3}{3,6} \right\}$

PAPCO

Scanned by CamScanner

Subject
Date

اجماع

	3	2	1	9	8	7	19	18	17
1,4	1 0.2	0	0	0	0	0	0	0	0
1,5	0	1 0.2	0	0	0	0	0	0	0
1,6	0	0	1 0.2	0	0	0	0	0	0
2,4	0	0	0	1 0.5	0	0	0	0	0
2,5	0	0	0	0	1 1	0	0	0	0
2,6	0	0	0	0	0	1 0.4	0	0	0
3,4	0	0	0	0	0	0	1 0.3	0	0
3,5	0	0	0	0	0	0	0	1 0.3	0
3,6	0	0	0	0	0	0	0	0	1 0.3
	0.2	0.2	0.2	0.5	1	0.4	0.3	0.3	0.3

$$B = \left\{ \frac{0.2}{1} + \frac{0.2}{2} + \frac{0.2}{3} + \frac{0.4}{7} + \frac{1}{8} + \frac{0.5}{9} + \frac{0.3}{17} + \frac{0.3}{18} + \frac{0.3}{19} \right\}$$

اعداد فازی

A_0 ← مجموعه مرجع

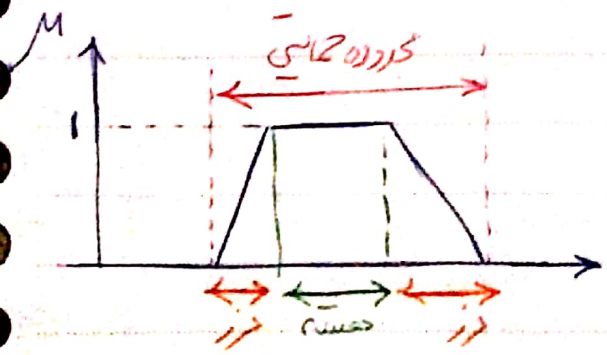
A_+ ← مجموعه Support

مرکز های لا دارای یک مجموعه مساوی یک تا زه است. اگر محدب نباشد ممکن است چند تا زه باشد

تابع تعلق: مرزها ← مقادیر کمتر از یک

هسته ← مقادیر یک

محدوده حمایتی ← مقادیر غیر صفر



مقدار Max تابع تعلق ⟸ ارتفاع / تابع تعلق نرمال، دارای ارتفاع یک است

PAPCO

Scanned by CamScanner

Subject
Date

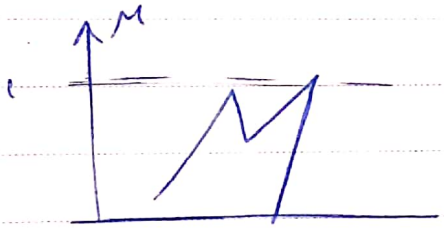
* نقاط تقاطع، نقطه هایی هستند که مقدار تابع تعلق آن ها برابر ۰٫۵ است.

تابع تعلق محدب

If $x < y < z \Rightarrow \mu(y) \geq \text{Min}[\mu(x), \mu(z)]$

← حداکثر یک قسمت صعودی و یک قسمت نزولی دارد.

اشتراک دو تابع تعلق محدب ⇐ محدب اجتماع دو تابع تعلق محدب ⇐ معلوم نیست



X غیر محدب

عدد فازی ⇐ تابع تعلق محدب با یک نقطه نرمال

برش های لاندا مجموعه فازی ⇐ برش های لاندای یک مجموعه فازی، مجموعه های کلاسیکی هستند که اعضای آن ها تابع تعلق بزرگتر یا مساوی لاندا دارند.

$$A_\lambda = \{x \mid \mu_A(x) \geq \lambda\}$$

$$A^\lambda = \left\{\frac{\lambda}{x} \mid x \in A_\lambda\right\} \Rightarrow \bigcup_{\lambda \in [0,1]} A^\lambda = A$$

$$\alpha \geq \lambda \Rightarrow A_\alpha \subseteq A_\lambda$$

** برش لاندای یک عدد فازی، بازه ای را تعریف می کند که در آن بازه مقدار تابع تعلق بزرگتر از λ است

$$A_\lambda = [a, b]$$

PAPCO

Scanned by CamScanner

Subject
Date

مثال))

$$A = \left\{ \frac{0.5}{a} + \frac{0.2}{b} + \frac{0.7}{c} \right\}$$

$$A_{0.7} = \{c\} \qquad A_{0.5} = \{a, c\} \qquad A_{0.2} = \{a, b, c\}$$

$$\Rightarrow \overset{0.7}{A} = \left\{ \frac{0.7}{c} \right\} \qquad \overset{0.5}{A} = \left\{ \frac{0.5}{a} + \frac{0.5}{c} \right\}$$

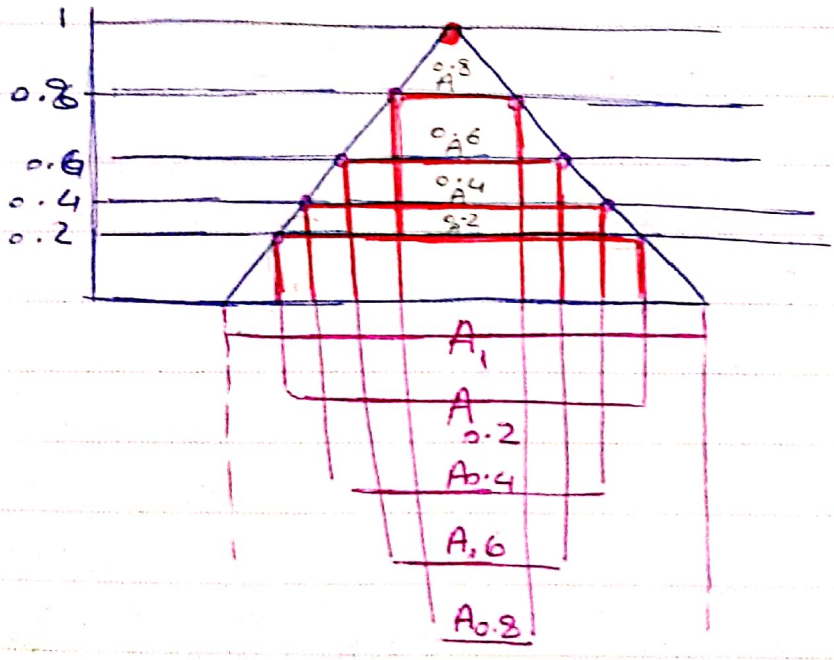
$$\overset{0.2}{A} = \left\{ \frac{0.2}{a} + \frac{0.2}{b} + \frac{0.2}{c} \right\}$$

$$A = \overset{0.7}{A} \cup \overset{0.5}{A} \cup \overset{0.2}{A} = \left\{ \frac{0.5}{a} + \frac{0.2}{b} + \frac{0.7}{c} \right\}$$ ⇐ اجتماع می‌گیریم

در سطح مجموعه اولیه بازسازی شود

با همهٔ برش‌های لایه‌ای متفاوت، مجموعه اولیه قابل بازسازی است.

(حالت پیوسته)



* هر چه تعداد برش‌های لایه‌ای بیشتر شود، اجتماع آن‌ها ($\overset{\lambda}{A}$ ها) به مجموعه اولیه بیشتر شبیه می‌شود.

PAPCO

Scanned by CamScanner

Subject
Date

منطق کلاسیک

$P: x \text{ is in } A$

متغیر (شی) — مجموعه

درستی بودن

$T(P) = \mu_A(x)$

* برای هر گزاره می‌توان یک مجموعه و یک عضو متعلق به آن تعریف کرد

$A =$ مجموعه دارای

$P = a \in A$

$\hookrightarrow T(P) = \mu_A(a)$

a عضو برای درست است.

* یک گزاره تعلق یک عضو به یک مجموعه است.

* درستی گزاره با تابع تعلق بیان می‌شود.

$P: A$

$Q: B \Rightarrow \boxed{P \wedge Q \equiv A \cap B}$

$P \to Q \equiv A \to B \equiv A' \cup B$ ← مجموعه‌های مرجع یکسان نیست ⇐ توسعه استوانه‌ای

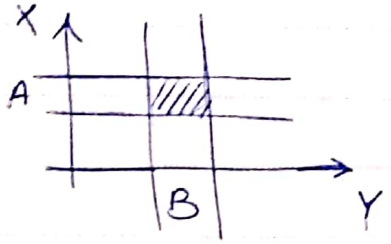
$\equiv P' \vee Q$

برای گزاره‌ها از $\wedge$ و $\vee$ استفاده می‌کنیم زیرا هر گزاره را به صورت یک قرار تعلق نشان می‌دهند.

$\Longrightarrow P \to Q \equiv A \to B \equiv (A' \times Y) \cup (X \times B)$

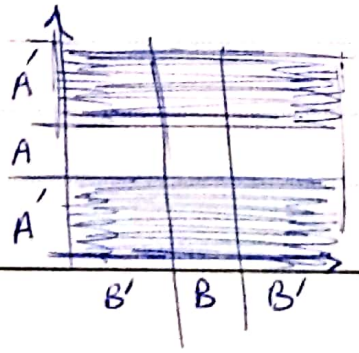
تعریف اول:

اگر A برقرار باشد:



$\xRightarrow{\cup} A \to B =$

اگر A برقرار نباشد:



$(A \times B) \cup (A' \times Y)$

قانون ارسطو: اگر شرط برقرار بود آنگاه B

اما اگر نبود هر چیزی خواستی بگو

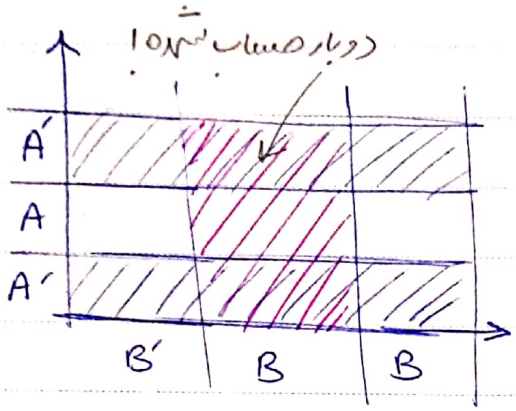
PAPCO

Scanned by CamScanner

Subject
Date

تعریف دوم :

$$A \to B = (A' \times Y) \cup (X \times B)$$



* در حالت کلاسیک دو تعریف باهم تفاوتی ندارند اما بسته به تعریف S نرم، avb ، ممکن است a شود یا نشود . اگر a نشود این ۲ تعریف باهم فرق دارند.

* در مجموعه های کلاسیک برای $A \circ R = B$ لزومی ندارد R جدا بزند باشد تا نتیجه B شود حتی اگر مجموعه $A_1 \subseteq A$ باشد باز هم $A_1 \circ R = B$ می شود.

منطق فازی

$$R = (A \times B) \cup (A' \times Y)$$

$$\mu_R(x,y) \overset{max}{=} \left[\min\left[\mu_A(x), \mu_B(y)\right], 1-\mu_A(x)\right]$$

$$R = (A' \times Y) \cup (X \times B)$$

$$\mu_R(x,y) = \max\left[1-\mu_A(x), \mu_B(y)\right]$$

این ۲ تا کاملاً باهم برابر می شوند. ($\max a, a = a$)

مثال ۱۱) اگر حجم کم باشد فشار باید زیاد باشد.

$$A \Rightarrow B$$

$$\underbrace{\text{حجم کم}}_{A} \qquad \underbrace{\text{فشار زیاد}}_{B}$$

حجم کم $A = \left\{\frac{0.2}{10}, \frac{1}{20}, \frac{0.5}{30}\right\}$ فشار زیاد $B = \left\{\frac{0.3}{40} + \frac{1}{50} + \frac{0.6}{60}\right\}$

PAPCO

Scanned by CamScanner

Subject
Date

* در این گونه رابطه‌ها، چون نمی‌توانیم وقتی شرط برقرار نیست، هر کاری بکنیم، پس هیچ کاری نمی‌کنیم

یعنی رابطه فقط حالت خالی می‌شود که شرط برقرار باشد یعنی $\boxed{R = A \times B}$

(نمی‌دانیم اگر حجم کم شود چی می‌شود)

* در سیستم‌های فیزیکی نمی‌توان از قانون ارسطو استفاده کرد (در فلسفه می‌توان)

مثال:

$$A = \left\{ \frac{0.2}{10}, \frac{1}{20}, \frac{0.5}{30} \right\} \Rightarrow A' = \left\{ \frac{0.8}{10} + \frac{0}{20} + \frac{0.5}{30} + \frac{1}{40} \right\}$$

$$X = \{10, 20, 30, 40\}$$

حل مثال:

$R = A \times B$

	40	50	60
10	0.2	0.2	0.2
20	0.3	1	0.6
30	0.3	0.5	0.5

نسبتاً کم $A' = \left\{ \frac{0.3}{10} + \frac{0.8}{20} + \frac{0.4}{30} \right\}$

$$\frac{0.3}{10} + \frac{0.8}{20} + \frac{0.4}{30}$$

(این جا همان ترکیب Max Min داریم)

	40	50	60	
10	0.2	0.2	0.2	0.3
20	0.3	1	0.6	0.8
30	0.3	0.5	0.5	0.4
	0.3	0.8	0.6	

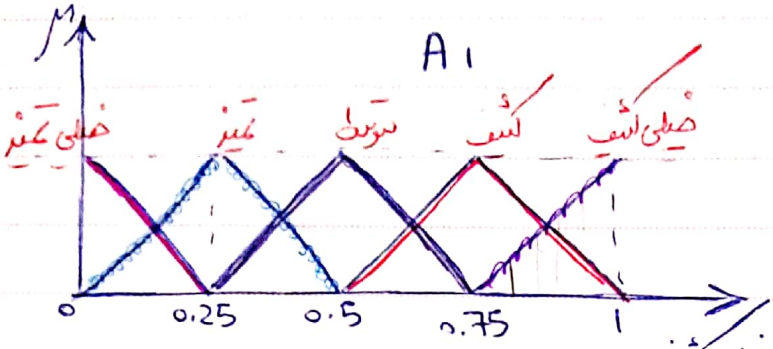
$$\Rightarrow B = \left\{ \frac{0.3}{40} + \frac{0.8}{50} + \frac{0.6}{60} \right\}$$

PAPCO

Scanned by CamScanner

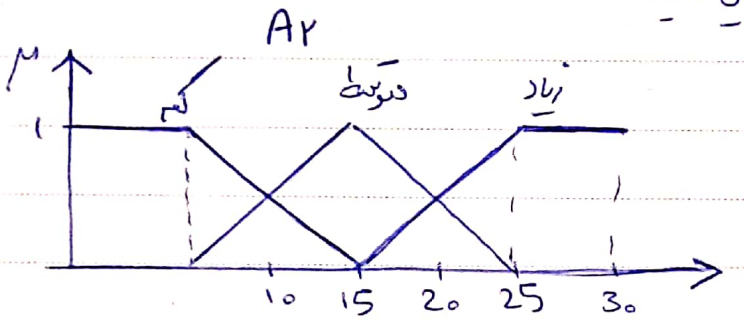
Subject
Date

مثال: اگر لباس‌ها خیلی کثیف باشند و زیاد باشند، آنگاه آب داغ است.



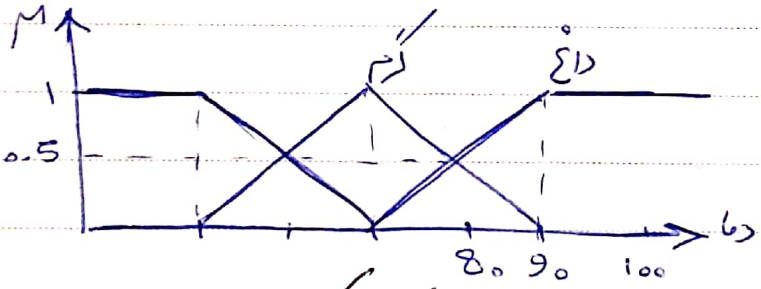
برای 0.8 ⟹ 0.2
0.9 ⟹ 0.6
1 ⟹ 1

A'_1 نسبتاً خیلی کثیف $= \left\{ \frac{0.3}{0.8} + \frac{0.7}{0.9} + \frac{0.9}{1} \right\}$



A'_2 نسبتاً زیاد $= \left\{ \frac{0.4}{18} + \frac{0.7}{20} + \frac{1}{22} \right\}$

برای 18 ⟹ 0.3
20 ⟹ 0.5
22 ⟹ 0.7



برای 80 ⟹ 0.5
85 ⟹ 0.75
90 ⟹ 1

پس داریم: اگر لباس‌ها نسبتاً خیلی کثیف باشند و نسبتاً زیاد باشند، آب حلونه است.

$$A'_1 \times A'_2 = \left\{ \frac{0.3}{0.8,18} + \frac{0.3}{0.8,20} + \frac{0.3}{0.8,22} + \frac{0.4}{0.9,18} + \frac{0.7}{0.9,20} + \frac{0.7}{0.9,22} + \frac{0.4}{1,18} + \frac{0.7}{1,20} + \frac{0.9}{1,22} \right\}$$

⟸ اول همیشه ورودی را بدست می‌آوریم بعد رابطه. زیرا ستون‌های ورودی سطرهای رابطه‌اند.

* باید کل خروجی‌ها (از 0 تا 100) را بگیریم اما برای مثال چند را حساب می‌کنیم.

$$R = \overset{A_1 \times A_2}{A} \times B = A_1 \times A_2 \times B \Rightarrow \mu_R(x_1, x_2, y) = \mathrm{Min}\left[\mu_{A_1}(x_1), \mu_{A_2}(x_2), \mu_B(y)\right]$$

PAPCO

Scanned by CamScanner

Subject
Date

	80	85	90
0.8, 18	0.2 0.3	0.2 0.3	0.2 0.3
0.8, 20	0.2 0.3	0.2 0.3	0.2 0.3
0.8, 22	0.2 0.3	0.2 0.3	0.2 0.3
0.9, 18	0.3 0.4	0.3 0.4	0.3 0.4
0.9, 20	0.5 0.7	0.5 0.7	0.5 0.7
0.9, 22	0.5 0.7	0.6 0.7	0.6 0.7
1, 18	0.3 0.4	0.3 0.4	0.3 0.4
1, 20	0.3 0.7	0.5 0.7	0.5 0.7
1, 22	0.5 0.9	0.7 0.9	0.7 0.9
	0.5	0.7	0.7

(ترکیب Max Min)

$$\Longrightarrow B = \left\{ \frac{0.5}{80}, \frac{0.7}{85}, \frac{0.7}{90} \right\}$$

(فضای پوشش)

$$B = A \circ R$$

$$B_\lambda = A_\lambda \circ R_\lambda$$

$$R = A_1 \times A_2 \times B \Longrightarrow R_\lambda = A_{1\lambda} \times A_{2\lambda} \times B_\lambda$$

در این روابط تا زمانی که متمم نداریم، مشکلی نیست. اگر متمم بود به عنوان یک مجموعه مستقل به آن نگاه کنیم

$$(A')_\lambda \neq (A_\lambda)'$$

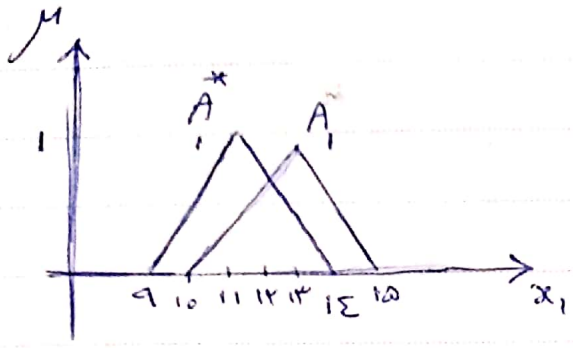
مثلاً $R = \underbrace{A'}_{C} \cup B \Rightarrow R = C \cup B \Rightarrow R_\lambda = C_\lambda \cup B_\lambda$

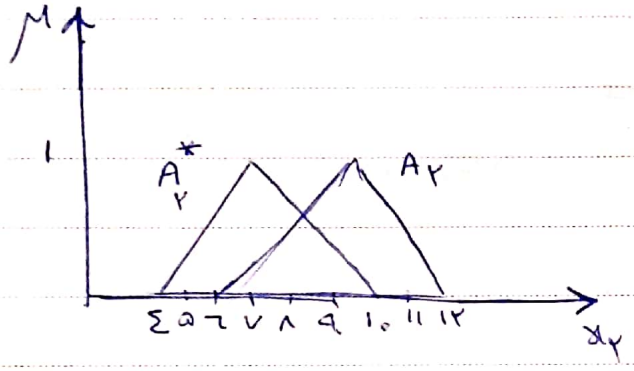
PAPCO

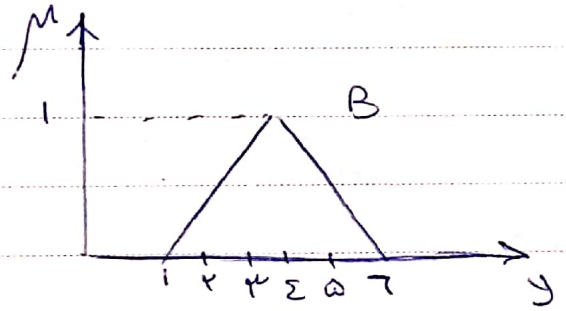
Scanned by CamScanner

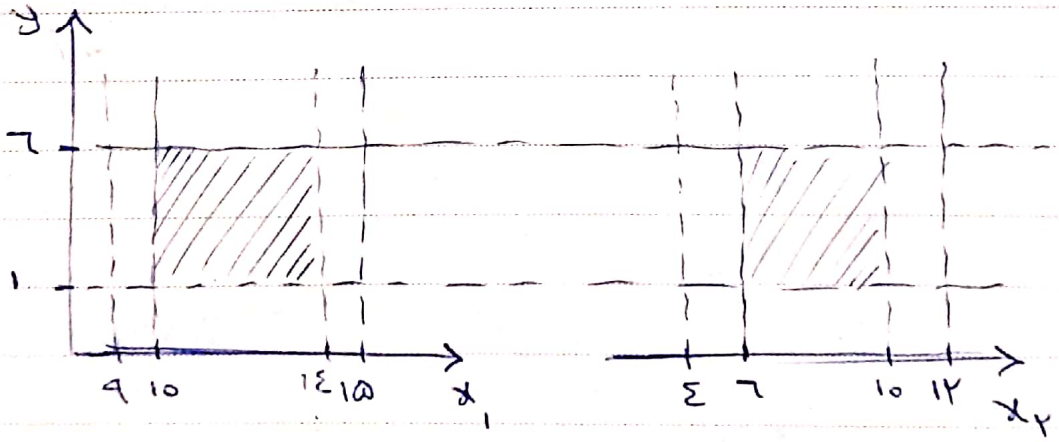
Subject
Date

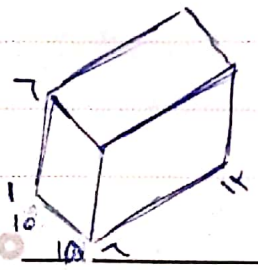
مثال از فضای پیوسته :











$R = A_1 \times A_2 \times B$

$A_{1\,0^+}^* = [9 \quad 14]$ $\quad A_{2\,0^+}^* = [4 \quad 10]$

$A_{1\,0^+} = [10 \quad 15]$ $\quad A_{2\,0^+} = [6 \quad 12]$

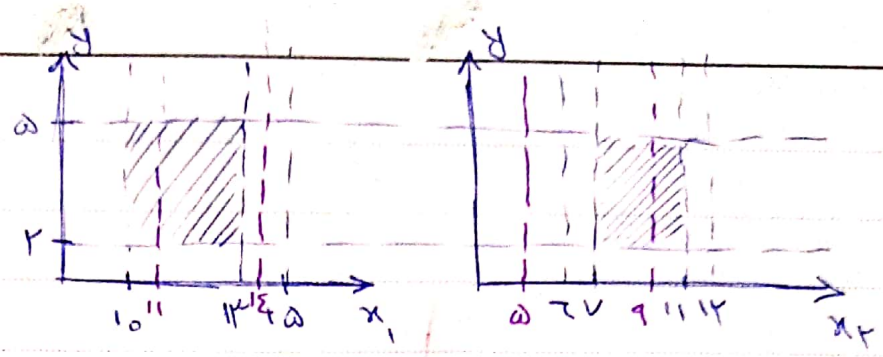
PAPCO

Scanned by CamScanner

Subject
Date

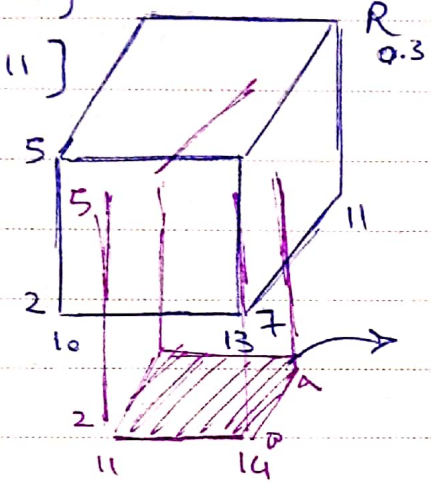
$B_{0^+} = [1 \quad 6]$

$B_{0.3} = [2 \quad 5]$



$A_{1\,0.3} = [10 \quad 13]$

$A_{2\,0.3} = [7 \quad 11]$



$A^*_{1\,0.3} = [11, 14]$

$A^*_{2\,0.3} = [5, 9]$

توسعه استوانه ای می دهیم تا ملحق شود ،
بعد اشتراک می گیریم .

$A_{11} = 12 \qquad A^*_{11} = 13$

$A_{21} = 8 \qquad A^*_{21} = 7 \quad \Longrightarrow B_1 = ?$

$x_1 = 12 \qquad x_2 = 8 \qquad y = 3$

A^*_{11} و A^*_{21} با x_1 و x_2 اشتراک ندارد در نتیجه خروجی تهی است . ($B_1 = \phi$)

← حتماً هر دو باید اشتراک داشته باشند تا خروجی تهی نشود .

$A_{1\,0.8} = [11, 12.5] \qquad A^*_{1\,0.8} = [12.5, 13.5]$

$A_{2\,0.8} = [7, 9.5] \qquad A^*_{2\,0.8} = [6, 8]$

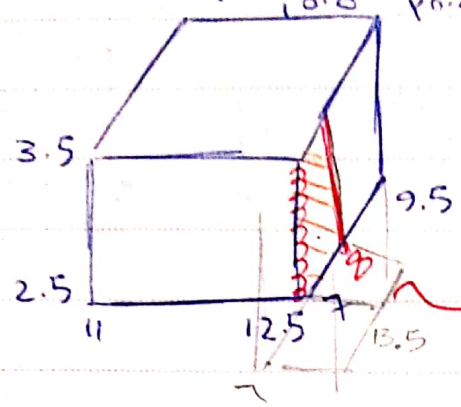
$\Longrightarrow B_{0.8} = [2.5 \quad 3.5]$

PAPCO

Scanned by CamScanner

Subject
Date

$R = A_{1\,0.8} \times A_{2\,0.2} \times B_{0.8}$



اشتراک ۲ دلتا این مستطیل می‌شود.

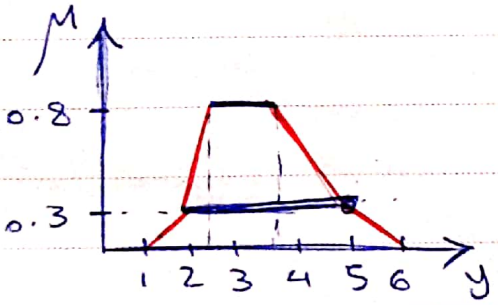
تصویر می‌کنیم روی لا و $B_{0.8}$ برای با [2.5 3.5] می‌شود

همیشه برش لا دارای 0^+ را بدست می‌آوریم که مجموعه Support را داشته باشیم بعد بین آن ها

برش "0.3 و 0.8 و 1 را هم بدست می‌آوریم.

$$B^{0^+} = \left\{ \frac{0^+}{[1\ \ 6]} \right\}$$

$$B^{0.3} = \left\{ \frac{0.3}{[2\ \ 5]} \right\} \qquad B^{0.8} = \left\{ \frac{0.8}{[2.5\ \ 3.5]} \right\} \qquad B_1 = \phi$$



* هر برش لا دارای مجموعه مرجع برای با خودش است.

* ورودی ها حتماً باید با A_1 و A_2 اشتراک داشته باشند تا از تصویر اشتراک آن ها بر روی لا خروجی حاصل شود.

$(R = A_1 \times \cdots \times A_n \times B)$

$A^*_{i\lambda} \cap A_{i\lambda}$

$i = 1, 2, \cdots, n$

* اگر حداقل یکی از اشتراک ها تهی باشد، خروجی ($B^*_\lambda = B_\lambda$) تهی است

در غیر این صورت (همه اشتراک ها غیر تهی باشند) خروجی غیر تهی است.

PAPCO

Scanned by CamScanner

Subject
Date

$$(\overset{R_1}{S_1} \cap \overset{R_2}{S_2}) \cup \overset{R_3}{S_3}$$

$S_1: \; A_1 \xrightarrow{R_1} B_1 \qquad B_1^* = A^* \circ R_1$

$S_2: \; A_2 \xrightarrow{R_2} B_2 \qquad B_2^* = A^* \circ R_2$

$S_3: \; A_3 \xrightarrow{R_3} B_3 \qquad B_3^* = A^* \circ R_3$

* در جاهایی که بین توابع اجتماع است می‌توان آن‌ها را ترکیب کرد و یک قانون کلی بدست آورد

S_i : $B_i^* = A^* \circ R_i$ (با اشتراک نمی‌توان این کار را کرد)

$$B^* = \cup B_i^* = \cup (A^* \circ R_i) = A^* \circ (\cup R_i)$$

$$\boxed{R = \cup R_i}$$

$$B = \overbrace{(B_1^* \cup B_2^*)}^{B_5^*} \cap \overbrace{(B_3^* \cup B_4^*)}^{B_6^*} \Longrightarrow B = B_5^* \cap B_6^*$$

$B_5^* = A^* \circ (R_1 \cup R_2) \qquad B_6^* = A^* \circ (R_3 \cup R_4)$

* به ازای هر قانون خروجی را بدست می‌آوریم و با توجه به اینکه بین توابع اجتماع یا اشتراک است، خروجی کلی را بدست می‌آوریم. ← قاعده کلی

اگر بین همه قانون‌ها اجتماع باشد، یک رابطه کلی که برابر با اجتماع آن‌ها است، حساب می‌کنیم و فقط یک ترکیب بین ورودی و رابطه حاصل شده خروجی را می‌دهد. وگرنه باید به ازای تمام اجتماع‌ها و اشتراک‌ها، یک بار ترکیب با ورودی حساب کنیم. (n بار خروجی جزئی و n بار هم اجتماع اشتراک) (ترکیب)

PAPCO

Scanned by CamScanner

Subject
Date

بین دانش $A \Rightarrow B$ else C اشتراک است.

$A \Rightarrow B$
$A' \Rightarrow C$

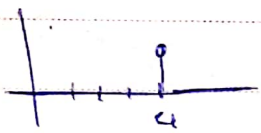
$$(A \Rightarrow B) \cap (A' \Rightarrow C)$$

فازی سازی

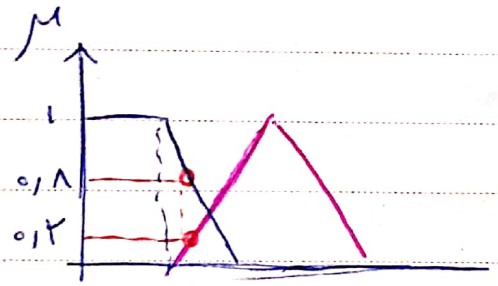
یک راه برای اینکه بتوانیم از قوانین فازی هنگامی که ورودی ما دقیق بود، استفاده کنیم، این است که تابع تعلق بنویسیم مانند مثال زیر.

$$\mu_A(x) = \begin{cases} 1 & x = 4 \\ 0 & x \neq 4 \end{cases}$$

شمارا برای 4



* جمع تعلق ها در هر نقطه برابر یک است ⇐ مجموعه های متعامد



آنتروپی: بی نظمی، عدم قطعیت

وقتی احتمال 0,5 است عدم قطعیت و ابهام کامل داریم. (وقتی احتمال ها برابر باشند شمارا برای ۳ تا کلاس 0,33 عدم قطعیت است)

P_i: احتمال تعلق داده i به کلاس یک

$1 - P_i$: " " " "

PAPCO

Scanned by CamScanner

Subject
Date

تابعی می‌خواهیم که وقتی احتمالات برابر باشند، Max باشد و وقتی احتمالات ۱ و ۰ هستند Min باشد

$$-k\sum_{i=1}^{L}\Big(\overbrace{p_i \ln p_i}^{①} + \overbrace{(1-p_i)\ln(1-p_i)}^{②}\Big)$$

مشتق می‌گیریم

$$-k\left(\ln p_i + 1 - (\ln(1-p_i)+1)\right) = 0$$

$$\ln p_i = \ln(1-p_i)$$

$$p_i = 1-p_i \Rightarrow \boxed{p_i = 0.5}$$

اگر $p_i = 0$ ① مبهم است $\Longrightarrow \dfrac{\ln p_i}{\frac{1}{p_i}} \xrightarrow{Hop} \dfrac{\frac{1}{p_i}}{\frac{-1}{p_i^2}} = -p_i = 0$

اگر $p_i = 1$ ② مبهم است $\Longrightarrow$ .....

در حالت کلی در صورتی که بیش از ۲ کلاس داشته باشیم، می‌توان آن را به صورت زیر نوشت:

$$E = -\sum_{i=1}^{L}\sum_{j=1}^{C} p_{ij}\ln p_{ij} - \sum_{i=1}^{L}\lambda_i\left(\sum_{j=1}^{C} p_{ij} - 1\right)$$

احتمال اینکه داده i به کلاس j تعلق داشته باشد p_{ij}:

$$\frac{\partial E}{\partial p_{ij}} = -\ln p_{ij} - 1 - \lambda_i = 0$$

$$\ln p_{ij} = -1-\lambda_i \Rightarrow p_{ij} = e^{-1-\lambda_i}$$

$$\frac{\partial E}{\partial \lambda_i} = \sum_{j=1}^{C} p_{ij} - 1 = 0 \Rightarrow \sum_{k=1}^{C} p_{ik} = 1$$

$$\Rightarrow \sum_{k=1}^{C} e^{-1-\lambda_i} = 1 \Rightarrow e^{-1-\lambda_i}\underbrace{\sum_{k=1}^{C}}_{C} = 1 \Rightarrow e^{-1-\lambda_i} = \frac{1}{C}$$

$$\Longrightarrow \boxed{p_{ij} = e^{-1-\lambda_i} = \frac{1}{C}}$$

PAPCO

Scanned by CamScanner

Subject
Date

** L داده های مختلف و C تعداد کلاس ها است.

قانون لاگرانژ: اگر بخواهیم Max و Min تابعی را که مشروط به یک شرط است حساب کنیم باید ضریب لاگرانژ را در آن شرط ضرب کرده و از خود تابع کم می کنیم.

از عبارت حاصل یکبار از همان متغیر قبلی و یکبار از ضریب لاگرانژ مشتق می گیریم (λ)

* در صفحه قبل آنتروپی $E = -\sum_{i=1}^{L}\sum_{j=1}^{C} P_{ij} \ln P_{ij}$ است ولی چون Max, Min را می خواستیم از قانون لاگرانژ استفاده کردیم.

تعریف آنتروپی برای داده های یک کلاس

* هر احتمال برای یک داده است.

احتمال اینکه در قسمت اول باشیم

$$S(x) = P_a(x) S_a(x) + P_b(x) S_b(x)$$

← از این جا برای جداسازی کلاس ها

* هر احتمال برای جمعی از داده ها است.

$$P_a = \frac{n_a}{n_a + n_b} \qquad P_b = 1 - P_a = \frac{n_b}{n_a + n_b}$$

$$S_a(x) = -P_{a1} \ln P_{a1}(x) - P_{a2} \ln P_{a2}(x) \qquad S_b(x) = -P_{b1} \ln P_{b1}(x) - P_{b2} \ln P_{b2}(x)$$

P_{a1}: داده های کلاس یک در قسمت a

P_{b2}: داده های کلاس دو در قسمت b

$$P_{b1} = \frac{n_{b1}}{n_b} \qquad P_{a1} = \frac{n_{a1}}{n_a}$$

$$P_{b2} = \frac{n_{b2}}{n_b} \qquad P_{a2} = \frac{n_{a2}}{n_a} \qquad P_{b2} = 1 - P_{b1}$$

$$n_{a1} + n_{a2} = n_a \qquad n_{b1} + n_{b2} = n_b$$

$S(x)$ را Min می کنیم می خواهیم کمترین آنتروپی را داشته باشد (داده های غیرمشترک کمتره)

تا بهترین مرز جداسازی را انتخاب کنیم. (بهترین جاهایی که تعداد داده ها عوض نشه آنتروپی را حساب کنیم) و جایی که Min باشد ← بهترین مرز برای جداسازی است

Scanned by CamScanner

Subject
Date

نمونه مثال Slide :

$$P_a = \frac{22}{22+18} \qquad P_b = \frac{18}{22+18}$$

$$P_{a1} = \frac{20}{22} \qquad P_{a2} = \frac{2}{22}$$

$$S_a = \frac{-20}{22} \ln \frac{20}{22} - \frac{2}{22} \ln \frac{2}{22}$$

$$P_{b1} = \frac{3}{18} \qquad P_{b2} = \frac{15}{18}$$

$$S_b = \frac{-3}{18} \ln \frac{3}{18} - \frac{15}{18} \ln \frac{15}{18}$$

$$\Longrightarrow S(x) = \frac{22}{40} S_a + \frac{18}{40} S_b$$

(در Slide ، 1+ در مخرج وجود دارد به خاطر اینکه مخرج صفر نشود اما این طور که ما محاسبه کردیم می دانیم صفر نمی شود و لزومی ندارد 1+ کنیم)

« Clustring »

در خوشه بندی label ها را نمی دانیم و بر اساس یک معیاری و کی تشخیص می دهیم کدام ها در یک خوشه اند.

در نتیجه یک معیار شباهت می خواهیم . (و کی های شبیه به هم دارند)

الگوریتم C-min یا k-min برای خوشه بندی است که تعداد خوشه ها C یا k است

و باید تشخیص کنیم چه داده هایی در این خوشه ها است .

HCM
(k-min or C-min)

PAPCO

Scanned by CamScanner

Subject
Date

اگر بخواهیم برای مسائل خوشه بندی مجموعه فازی تعریف کنیم، دو راه داریم:

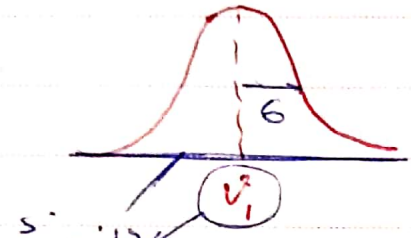
(۱)

$$u_{ik} = \frac{\frac{1}{D_{ik}}}{\sum_{j=1}^{c} \frac{1}{D_{jk}}}$$

تعلق داده X_k به مجموعه فازی i

(۲) راه حل ساده تر:

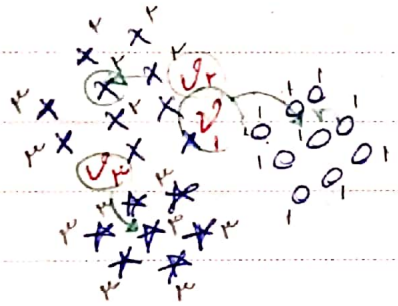
$$\mu_{A_i}(x) = e^{-\left(\frac{x-v_i}{\sigma}\right)^2}$$



فرض کنیم مجموعه فازی ای که داریم گوسی است.

انحراف معیار $\sigma \Rightarrow \sigma^2 = \frac{1}{N}\sum_{k=1}^{N}(x_k - v_i)^2$

C-min & K-min



ابتدا تعداد خوشه ها را مشخص می کنیم.

سپس به صورت تصادفی به تعداد خوشه ها،

مرکز ها را در نظر می گیریم.

یک معیار فاصله در نظر می گیریم. داده های خوشه باید به مرکز خوشه نزدیک باشند.

مختصات مرکز

$$v_i = \sum_{k \in v_i} \frac{x_k}{N_k}$$

میانگین داده های خوشه

$$u_{ik} = \begin{cases} 1 & x_k \in C_i \\ 0 & x_k \notin C_i \end{cases}$$

ماتریس $i \times k$

$$\boxed{v_i = \frac{\sum_{k=1}^{N} u_{ik} x_k}{\sum_{k=1}^{N} u_{ik}}}$$

تغییر تصادفی است که اگر داده k در خوشه i ام بود یک می شود و اگر نه صفر می شود.

$N \xrightarrow{\text{داده}} \text{ط}$

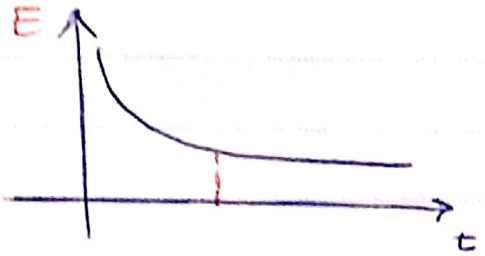
Scanned by CamScanner

Subject
Date

* همین طور مختصات مرکز را به دست می‌آوریم تا مکان واقعی آن بدست آید. مکان مدلر دیگر تغییر نمی‌کند.

معیار خطا: $E=\sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik} D_{ik}$

$D_{ik} \rightarrow \|x_k - v_i\|^2$ اقلیدسی

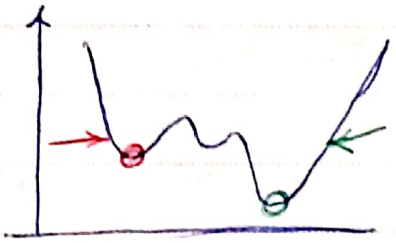


* هرچه انسترویی کمتر باشد، خوشه‌بندی بهتر است.

یک معیار برای خوشه‌بندی، انسترویی است.

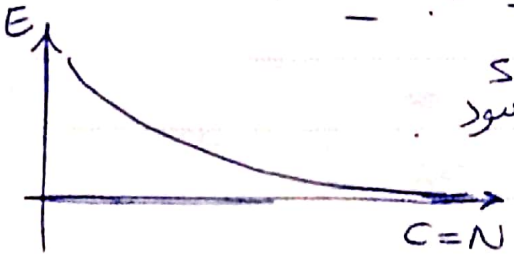
* یک عیب k-min این است که داده‌های نویز را در خوشه‌ها قرار می‌دهد.

* k-min به نقطه شروع حساس است و به اولین min که می‌رسد همان جا می‌ایستد. (min محلی)



* هرچه تعداد خوشه‌ها کمتر باشد، بهتر است زیرا تعمیم بیشتری دارد. (هرچه مدل ساده‌تر باشد (genera تر است) اگر تعداد خوشه‌ها زیاد باشد انگار به داده‌ها overfit شده است.

یعنی به داده‌های آموزش نتیجه می‌دهد ولی برای داده‌های جدید خوب نیست.



تعداد خوشه برای داده‌ها باشد، خطا صفر می‌شود.

PAPCO

Scanned by CamScanner

Subject
Date

FCM

در HCM برای اینکه Error ، مینیمم باشد مشتق نمی‌گیریم چون u_{ik} گسسته است و نمی‌شود از آن مشتق گرفت. در FCM ، u_{ik} را به صورت پیوسته مانند تابع تعلق که مقادیر بین ۰ و ۱ دارد در نظر می‌گیریم تا بتوان مشتق گرفت.

$$E = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}^{m} \| x_k - v_i \|^2$$

$$\frac{\partial E}{\partial u_{ik}}$$

* توان m اضافه شد چون در غیر آن صورت وقتی برحسب u_{ik} مشتق می‌گرفتیم، u_{ik} باقی نمی‌ماند تا مقدارش را بیابیم.

* چون محدودیت داریم (تعلق یک داده نسبت به تمام خوشه‌ها، جمعش باید یک باشد)

← ضریب لاگرانژ

$$(x_k - v_i)^T (x_k - v_i)$$

$$E = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}^{m} \| x_k - v_i \|^2 - \sum_{k=1}^{N} \lambda_k \left(\sum_{i=1}^{C} u_{ik} - 1 \right)$$

$$\frac{\partial E}{\partial v_i} = - \sum_{k=1}^{N} 2 u_{ik}^{m} (X_k - v_i) = 0$$

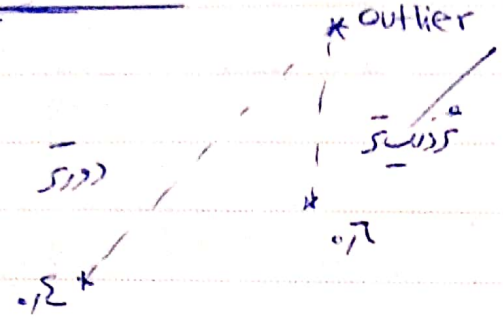
$$\sum_{k=1}^{N} u_{ik}^{m} X_k = v_i \sum_{k=1}^{N} u_{ik}^{m} \Longrightarrow \boxed{v_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} X_k}{\sum_{k=1}^{N} u_{ik}^{m}}}$$

PAPCO

Scanned by CamScanner

Subject
Date

* در FCM هم برای داده‌های outlier را به خوشه‌ها می‌دهد در حالی که به هیچ کدام تعلق ندارد.



ادامه اثبات قبل:

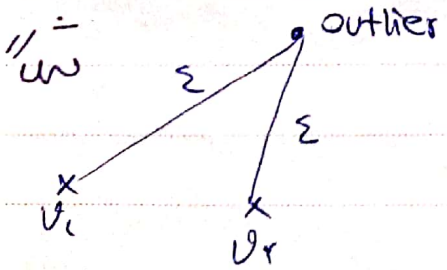
$$\frac{\partial E}{\partial u_{ik}} = m u_{ik}^{m-1} \underbrace{D_{ik}}_{\|x_k - v_i\|^2} - \lambda_k = 0 \Rightarrow u_{ik} = \left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} \frac{1}{D_{ik}^{\frac{1}{m-1}}} \quad (1)$$

$$\frac{\partial E}{\partial \lambda_k} = -\left(\sum_{i=1}^{c} u_{ik} - 1\right) = 0 \Rightarrow \sum_{j=1}^{c} u_{jk} = 1 \quad (2)$$

$$(1) \rightarrow (2) \quad \sum_{j=1}^{c} \left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} \frac{1}{D_{jk}^{\frac{1}{m-1}}} = 1 \quad (3)$$

$$\frac{(1)}{(3)} \quad \boxed{u_{ik} = \frac{\frac{1}{D_{ik}^{\frac{1}{m-1}}}}{\sum_{j=1}^{c} \frac{1}{D_{jk}^{\frac{1}{m-1}}}}}$$

چون دارد فاصله را نسبی حساب می‌کند نمی‌تواند outlier ها را تشخیص دهد و این مشکل FCM است.



$$\frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{4}} = 0.5$$

FCM هم مشابه k-min مشکل min محلی را دارد.

PAPCO

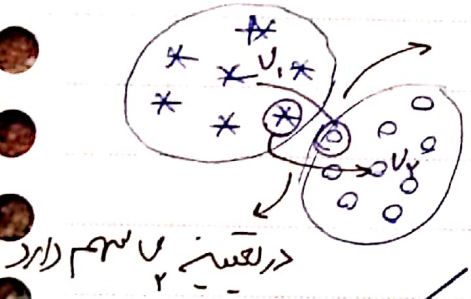
Scanned by CamScanner

Subject
Date

در FCM هم مراکز اولیه را به صورت تصادفی انتخاب می‌کنیم و با توجه به فرمول اینقدر اینکار را تکرار می‌کنیم تا error خیلی کم شود و ثبت شود.

تنها فرق FCM نسبت به k-min این است که تابع تعلق پیدا کرده و مرکزی که FCM درست می‌کند مراکز دقیق‌تری نسبت به k-min است.

* در k-min یک سری اطلاعات دور ریخته می‌شوند. در FCM هر داده برای مراکز هر خوشه ای سهم دارد.



* نماینده خوشه، مرکز خوشه است.

* m بزرگ خوب نیست هرچه m بزرگتر باشد خروجی به ۰ و ۱ بیشتر نزدیک می‌شود و فازی بودن کم‌تر می‌شود.

در FCM دو تا پارامتر داریم: ① تعداد خوشه‌ها ② m که میزان فازی بودن را نشان می‌دهد. (اگر m بزرگ باشد از حالت ۰ و ۱ بدتر است چون فقط ۱، یک می‌ماند و بقیه صفر می‌شوند) m معمولاً بین ۲ تا ۵ است. حالت default آن ۲ است.

m باید بزرگتر از یک باشد و هرچه به یک نزدیک‌تر باشد، میزان فازی بودن آن بیشتر است.

PAPCO

Scanned by CamScanner

Subject
Date

وقتی m بزرگ می‌شود داده‌هایی که نزدیک مراکز هستند تاثیر بیشتری دارند ولی بقیه مقدار صفر می‌گیرند.

u_{ik} : تعلق داده x_k به خوشه i ام

trace ← جمع عناصر روی قطر اصلی

$$\frac{1}{c} \leq \frac{\text{trace}\,|U U^T|}{N} \leq 1$$

(تعداد داده‌ها با تعداد خوشه‌ها برابر باشد)

حالت مساوی یک وقتی است که مقدار تعلق هر داده به یکی از خوشه‌ها 1 و به بقیه خوشه‌ها صفر باشد

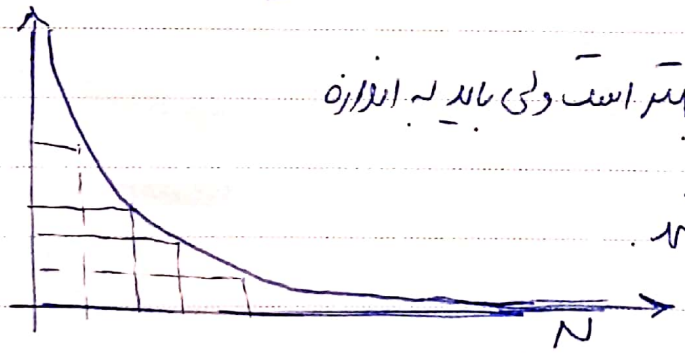
$$U = \begin{bmatrix} & & \\ & 1 & \\ - & - & - \\ & u_{ik} & \end{bmatrix}_{c \times N}$$

حالت $\frac{1}{c}$ وقتی اتفاق می‌افتد که $u_{ik} = \frac{1}{c}$ باشد یعنی ابهام کامل

که هر دو این حالات حدی هستند و اتفاق نمی‌افتند.

هرچه به $\frac{1}{c}$ نزدیک شود فازی بودن بیشتر است و هرچه به یک نزدیک‌تر باشد فازی بودن

کمتر است در نتیجه ابهام کمتر است و بهتر می‌توان نتیجه‌گیری (خوشه‌بندی) کرد.



* هرچه تعداد خوشه‌ها کمتر باشد بهتر است ولی باید به اندازه باشد تا جایی که خطاهای آن تقریباً کم باشد.

یک راه برای مشخص کردن m و c استفاده از انتروپی است.

$$-\sum u_{ik} \log u_{ik}$$

→ اگر u_{ik} : صفر و یک باشند ⇒ Min

اگر u_{ik} ها مساوی باشند ⇒ Max

PAPCO

Scanned by CamScanner

Subject
Date

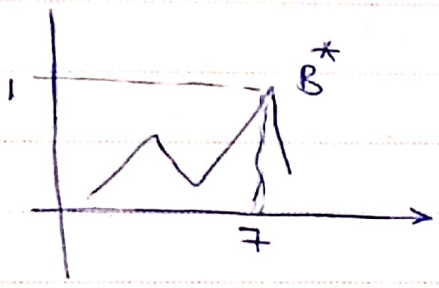
$$\underbrace{\sum_{i,k} u_{ik} \| x_k - v_i \|^2}_{\text{چه قدر داده های هر خوشه به مرکز خوشه نزدیک اند}} - \underbrace{\sum_i \| v_i - \bar{x} \|^2}_{\text{چه قدر مراکز از هم دورند}}$$

* هر چه حاصل کمتر باشد، بهتر است.

Defuzzification

روش های غیرفازی سازی:

① Max: جایی که تابع تعلق بیشترین مقدار را دارد.



مقدار غیرفازی B^* برابر ۷ است.

② مراکز ثقل:

$$z^* = \frac{\int z \cdot \mu_c(z)\, dz}{\int \mu_c(z)\, dz}$$

$$A = \left\{ \frac{0.3}{2} + \frac{0.7}{3} + \frac{0.1}{4} \right\}$$

$$z^* = \frac{0.3 \times 2 + 0.7 \times 3 + 0.1 \times 4}{0.3 + 0.7 + 0.1}$$

$$\boxed{\bar{x} = \frac{\sum_{x_i} x_i \mu_A(x_i)}{\sum_{x_i} \mu_A(x_i)}}$$

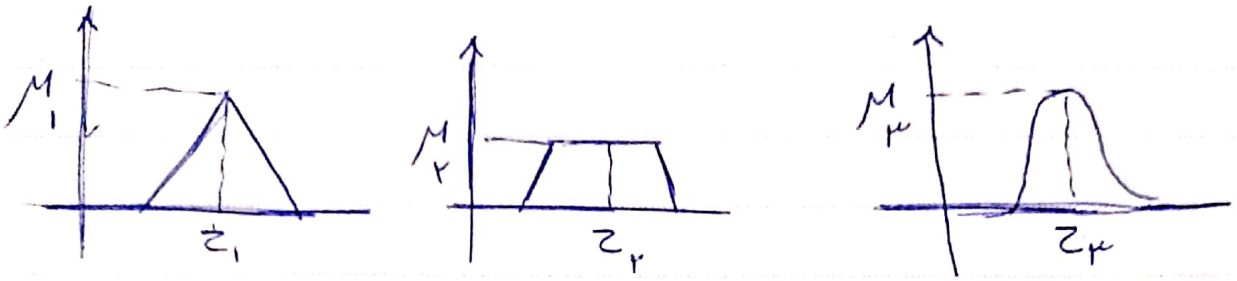
③ متوسط وزنی مراکز

$$z^* = \frac{\sum \bar{z} \cdot \mu_c(\bar{z})}{\sum \mu_c(\bar{z})}$$

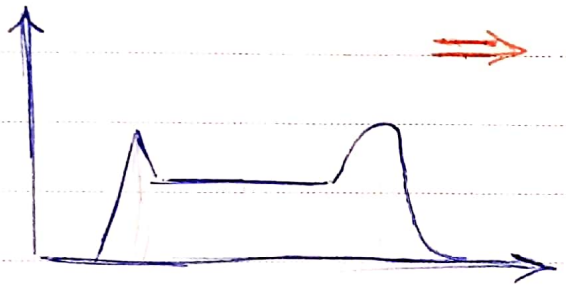
PAPCO

Scanned by CamScanner

Subject
Date



z_1 و z_2 و z_3 (مرکزها) معمولاً یا جایی هستند که Max باشد یا مرکز محدوده Support



⟹ اجتماع شکل‌های بالا

$$z^* = \frac{\mu_1 z_1 + \mu_2 z_2 + \mu_3 z_3}{\mu_1 + \mu_2 + \mu_3}$$

* این دو روش وقتی استفاده می‌شوند که شکل‌های ساده داشته باشیم.

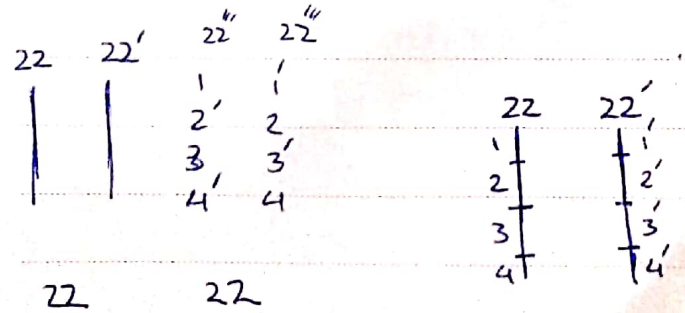
* با توجه به کاربرد از روش‌های مختلفی استفاده می‌کنند

الگوریتم‌های تکاملی

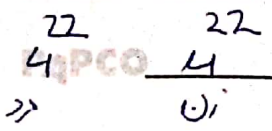
معیار کارایی تکامل این است که نسل‌ها هرچه بیشتر ادامه پیدا کند. هدف تکامل بقای ژن است

(نه بقای موجودات)

ژن کلدی است که خصوصیات فنوتیپی را کد می‌کند.



23 جفت کروموزوم که ۲۲ تا جفتی است و ۲۳ امی فقط برای جنسیت است.



زن	X	X
مرد	X	Y

Scanned by CamScanner

Subject
Date

تکامل

DNA ها به جای ۰ و ۱ (باینری)، A, C, G, T (الفبای ژنتیک) را توجه می‌شوند.
computer

و از سرعت و حافظه بالایی برخوردارند.

شایستگی در زن (تکامل) یعنی بتواند نسل‌های بیشتری بقا پیدا کند.

← دو ویژگی برای تکامل: ۱- تنوع ۲- انتخاب شایسته‌ترین‌ها (انتخاب طبیعی)

یعنی موجوداتی که شایستگی بیشتری دارند، تعدادشان زیاد می‌شود.

گاهاً در ایجاد تنوع تصادف دخالت دارد.

تنوع باعث افزایش کیفیت است و انتخاب کیفیت را کم می‌کند و کیفیت را بالا می‌برد. در نتیجه

این دو ویژگی تکامل با هم در تضاد است. ولی هر کدام هم اگر نباشند، تکامل اتفاق نمی‌افتد!!

تنوع با تولید مثل اتفاق می‌افتد. بین تنوع و انتخاب باید تناسب ایجاد شود.

ویژگی‌های فنوتیپی موجود زنده توسط پروتئین‌ها تعیین می‌شود و پروتئین‌ها از طریق آن

تعیین می‌شود. در واقع می‌توان گفت DNA درخت هستی است. حاصل عمل لقاح،

سلول zigot اولین سلول بنیادی جنینی است که با تقسیم سلولی به مجموعه‌ای از

سلول‌های بنیادی می‌رسد. سلول بنیادی ۲ نوع است: ① جنینی (که می‌تواند به

هر نوع سلولی تبدیل شود، عصبی، قلب و ...) ② بالغ

PAPCO

Scanned by CamScanner

Subject
Date

بنیادی بالغ است به جنینی محدودتر اند یعنی به هر نوع سلولی نمی توانند تبدیل شوند.

برای همین مثلاً وقتی سلول های عصبی از بین می روند، دیگر تولید نمی شوند (حتماً باید سلول بنیادی جنینی اش ذخیره شده باشد)

سلول های بنیادی جنینی بعد از یک مدت از بین می روند و به سلول بنیادی بالغ و سلول های تخصصی تبدیل می شود.

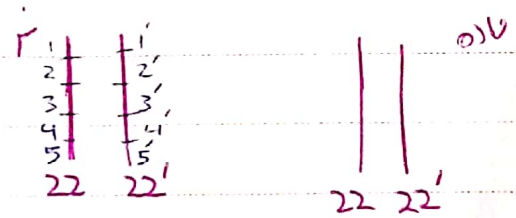
موجودات ۲ نوع اند: ← پلوریپوتنت: از هر کروموزوم ۲ تا دارند. جفت بودن به خاطر این (اکثر موجودات) است که یکی از نر می آید و یکی از ماده. (غیر از جنسیت)

هاپلوئیدی: اکثر تک سلولی ها

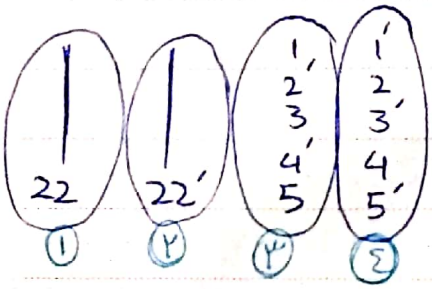
در انسان

* فقط کروموزوم ۲۳ تک است

	X	Y
X	XX	XY
X	XX	XY



وقتی اسپرم تولید می شود (برای مثال) ۴ حالت داریم:



* در همانندسازی و در انسان این هستیم که از همه ی کروموزوم ها حالت ۱ و ۲ را داشته باشیم (یعنی ۱ و ۱' و ۲ و ۲' و ... و ۲۲ و ۲۲')

PAPCO

Scanned by CamScanner

Subject
Date

آزمایش نمونه

	R	R
W	WR	WR
W	WR	WR

نسل اول →

	W	R
W	WW	WR
R	RW	RR

.....

	R	R
R	RR	RR
R	RR	RR

W ← غالب و سفید

R ← مغلوب و قرمز

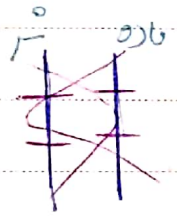
W غالبه پس همه سفید اند

Red، ¼ خود را نشان داد.

اگر ۲ تا RR پیدا کنیم که با هم تولید مثل کنند این نسل را می‌توان گسترش داد.

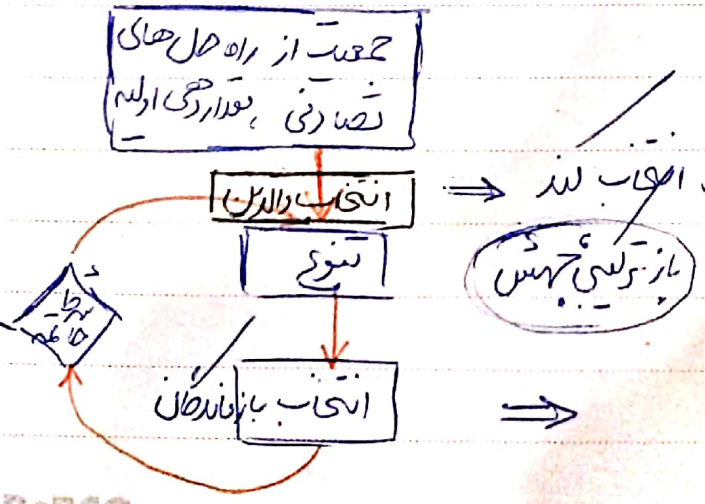
اگر الگوریتم‌های تکاملی هاپلوئیدی اند.

cross_over در هاپلوئیدی‌ها بین کروموزوم پدر و مادر انجام می‌شود (درحالیکه در واقعیت (دیپلوئیدی‌ها)، cross over بین کروموزوم‌های نر یا کروموزوم‌های ماده بود)



RNA poly.. : آنزیمی که می‌تواند از DNA کپی درست کند.

یکی از کاربردهای تئوری تکامل، بدست آوردن Max و Min یک تابع است



PAPCO

Scanned by CamScanner

Subject
Date

مراحل الگوریتم تکاملی :

① بازنمایی — کروموزوم و ژن ها چه باشند ← راه حل

بازنمایی ⟷ کد کردن راه حل

② تولید جمعیت اولیه و ارزیابی آن ها

③ انتخاب والدین ④ بازترکیبی ⑤ جهش

⑥ ارزیابی

⑦ انتخاب بازماندگان ⑧ شرط خاتمه

* در هر مرحله که انتخاب داریم باید یک معیار ارزیابی داشته باشیم که چه کسانی انتخاب شوند.

عملگرهای تنوع: بازترکیبی، جهش

عملگرهای انتخاب: انتخاب والدین، انتخاب بازماندگان

* در الگوریتم های تکاملی این امکان وجود دارد که اکتشاف و استخراج باهم باشند.

* در حالت کلی در الگوریتم تکاملی عملگرهای تنوع برای جستجوی عمومی و عملگرهای انتخاب برای جستجوی محلی اند. اما می توان آن ها را به گونه ای طراحی کرد که آن جستجوی دیگر را هم داشته باشند.

① بازنمایی : از تمام روش های کد کردن در این مرحله می توان استفاده کرد مانند جایگشت:

(عدد int)

۸ وزیر :

1	5	3	8	2	7	6	4

باینری (۱-۰) فضای حالت = 2^{48}

موقعیت وزیر اول

" وزیر دوم

فضای حالت : $8!$

۸ تا ۶ بیت

$2^6 \rightarrow 64$

Scanned by CamScanner

Subject
Date

ویژگی اصلی:

← بازنمایی نمایش راه‌حل‌های مختلف است به طوری که همهٔ حالت‌های ممکن را که می‌تواند جواب باشد، نشان دهد. (در بازنمایی باید تمام راه‌حل‌های ممکن تولید شود)

بازنمایی ای بهتر است که فضای حالت کوچک‌تری داشته باشد.

(Gentic algorithm)

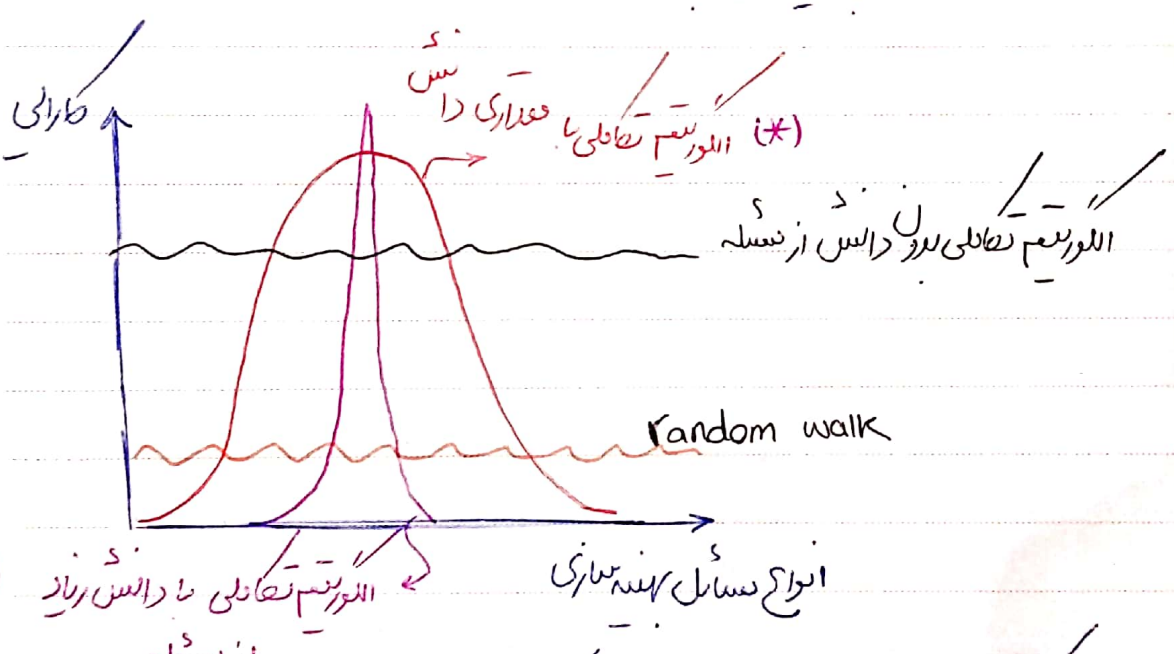
(**) الگوریتم‌های تکاملی:

GA — باینری — Gray Code

ES — اعداد حقیقی — (Evolution Strategy)

EP — گراف — (Evolution programing)

GP — درخت — } ⇒ تولی را بدست می‌آورد که آن تولی بهترین تولی برای آن کار است.



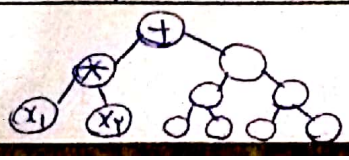
(*) مسائلی که می‌توان با گرفتن اطلاعات فضای حالتشان را کاهش داد. مثلاً تبدیل مسأله به مسأله زنگ اسبری گراف فضای حالت را کم می‌کند. یا بازنمایی با استفاده از جایگشت

مثال //

$$F(x_1, x_2) = a_1 x_1^2 + a_2 x_2^2 + a_3 x_1 x_2 + a_4 x_1 + a_5 x_2 + a_6$$

$$\frac{1}{2}\sum_{l=1}^{N} \left(F(x_1^l, x_2^l) - y\right)^2 \xrightarrow{\text{تولی با اعداد حقیقی}} \boxed{a_1 \mid a_2 \mid \cdots \mid a_6}$$

اگر بخواهیم تولی را هم تحول بده:



Scanned by CamScanner

Subject
Date

Fitness ای خوب است که بتواند برای حالت های متفاوت، تمایز قائل شود.
(شایستگی)

برای جمعیت خیلی اوقات به تعریف پارامترهای ربط پیدا می کند

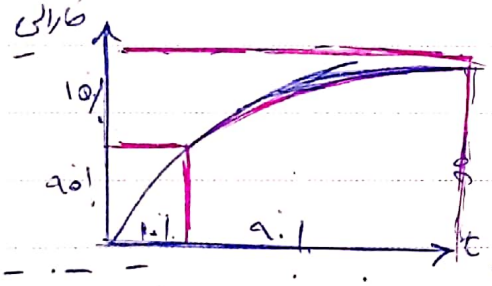
← پارامترهای روی پارتیکل و جمعش توی است. با توجه به پارامترها، متفاوت می شوند.

۲) تولید جمعیت اولیه:

← جمعیت اولیه را به صورت تصادفی یکنواخت انتخاب می کنیم.

سوال: اگر دانش راجع به مسأله داشته باشیم، آیا باید جمعیت را با آن شرط گم کنیم؟

خیر: زیرا ۱) اصل تنوع تکامل زیر سوال می رود. ۲) اگر بدانیم جواب در چه محدوده ای است دیگر احتیاج به الگوریتم تکاملی نخواهیم داشت.



حرکت الگوریتم های تکاملی با زمان اجرا

اگر از جمعیت مشخص و محدود استفاده کنیم فقط ما را به نمودار صورتی می رساند، خیلی در بهترین متفاوت (۱۰٪) نیست. اگر تنوع از بین برود ممکن است توی Min های محلی گیر کنیم اگر حتی آن هایی که مطمئن هستیم جواب ما نیستند، حذف کنیم ضرر کردیم چون می توانند با ترکیب شدن (crossover, mutation) جواب های بهتر به ما بدهند.

۸) شرط خاتمه: ۱- یکی زمانی الگوریتم تکاملی برای با تعداد ارزیابی است (به جای تعداد iteration ها)

۲- در iteration های مختلف چون در هر الگوریتم تکاملی تعداد فرزندان متفاوت است.

۱- تعداد تکرارها
۲- تعداد شایستگی
۳- همگرایی
۴- عدم تنوع

Scanned by CamScanner

Subject
Date

* در هر iteration به تعداد فرزندان ارزیابی داریم.
* تعداد فرزندان در الگوریتم های تکامل مختلف برابر نیست.

نکته
* تعداد فرزندان تابعی از پیچیدگی الگوریتم نیست
اگر به جای ۱۰ فرزند ۱۰۰ تا تولید کند، نمی توان نتیجه گرفت پیچیدگی ۱۰ برابره!

* در مرحله تولید جمعیت، fitness ها محاسبه می شوند.

کم ترین هزینه و ساده ترین روش برای شرط خاتمه ⟸ تعداد ارزیابی

مزیت سادگی و پیچیدگی پایان (محاسبه شرط خاتمه هزینه ندارد).

شرط خاتمه = فرزندان × iteration
تعداد

عیب: تعیین مقدار بهینه کار ساده ای نیست. (ممکن است کم باشد و به جواب نرسیم)

← این شرط خاتمه تنها شرط خاتمه ای است که از حلقه بیرون می آید و در loop گیر نمی کند.

مثلاً نخواهیم برنامه در ۱۰۰ تمام شود.

۲- اگر مقدار بهینه شایستگی را داشته باشیم می توان به عنوان شرط خاتمه آن را به کار برد. روش پایانی ارزیابی

(تعداد ارزیابی ها با مقدار بهینه شایستگی OR می کنیم)

تنها شرطی که اگر از حلقه بیرون آمدیم یعنی بهترین جواب را داریم.

۳- همگرایی: متوسط شایستگی هر نسل اختلاف زیادی با قبلی نداشته باشد. یعنی شیب نمودار کاهشی بر حسب iteration، به صفر میل کند.

عیب: حلقه نامحدود را دارد (یعنی ممکن است این شرط هیچ وقت پیش نیاید)

در هر iteration یک بهترین فرزند است که Max شایستگی هر iteration را دارد. اگر اختلاف این Max ها در

PAPCO

k تا iteration قبلی کمتر از حد آستانه باشد، یعنی به همگرایی رسیدیم.

Scanned by CamScanner

Subject
Date

در ۱ دوتیک را حساب می‌کنیم در ۲ بهترین شاخص‌ها

همگرایی اودرس ممکن است به خاطر نبود تنوع باشد نه به خاطر Min محلی.

اگر با شرط همگرایی بیایم بیرون و هنوز مقدار شاخص اتفاق نیافتاده باشد ⇐ همگرایی اودرس

اگر بخواهیم بفهمیم همگرایی اودرس است یا واقعی، با تابع ارزیابی (دومین روش) And می‌کنیم.

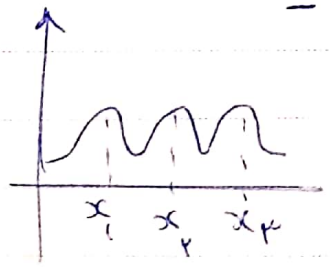
← (همیشه یکسه فهمید اگر مقدار بهینه را داشته باشیم با آن مقایسه می‌کنیم اگر نزدیک بود

یعنی همگرایی واقعی است)

مزیت: چه همگرایی واقعی باشد چه اودرس، از حلقه بیرون می‌آئیم.

۴- عدم تنوع:

عدم تنوع تجربه همگرایی می‌شود ولی همگرایی معنایش لزوماً عدم تنوع نیست.



همگرایی و تنوع ✓

⇐ عدم تنوع یعنی شاخص‌ها با هم برابر است.

در الگوریتم‌های تکاملی، تنوع ژنوتیپی با تنوع فنوتیپی (تنوع شاخص‌ها) با هم یکی است.

← (قله‌ها با هم یکی می‌شوند) همگرایی ⇐ عدم تنوع فنوتیپی

ولی در حل این طور نیست. عدم تنوع ⇐ عدم تنوع ژنوتیپی

Scanned by CamScanner

$\bar{x} = ۲$

Subject

Date

کروموزوم

4	2	3	1
2	3	1	4
3	2	4	1
x_1	x_2	x_3	x_4

ژن

۴ ژن :

محاسبه انحراف معیار ژنوتیپی :

$$\sigma = \frac{\sqrt{(۴-۳)^۲ + (۲-۳)^۲ + (۳-۳)^۲}}{۳}$$

انحراف معیار

* حاصل این انحراف معیار ، تنوع ژنوتیپی است .

اگر صفر باشد یعنی تنوع نداریم .

(هرچه بیشتر یعنی تنوع بیشتر)

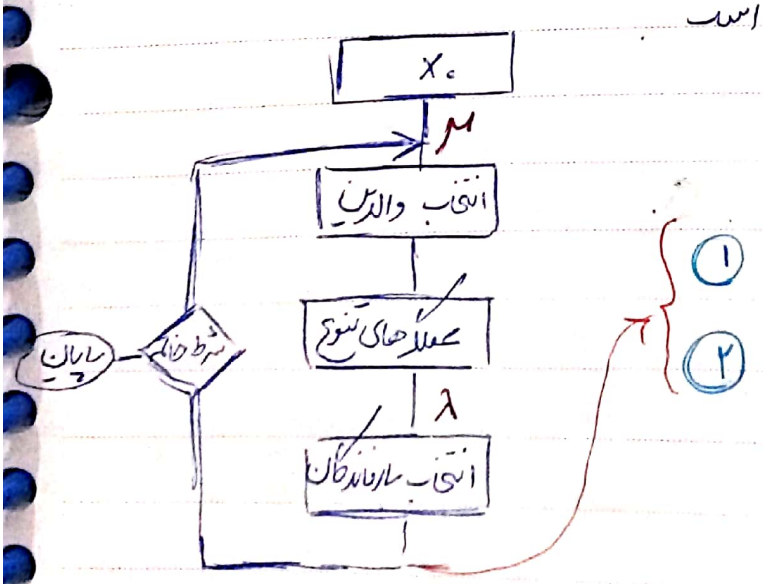
روش های انتخاب :

جمعیت اولیه را با μ نشان می دهیم . λ تعداد فرزندان است .

برای انتخاب والدین فضای امکان μ است .

$\mu \rightarrow$ جمعیت جاری

برای انتخاب سازماندهان فضای امکان $\mu+\lambda$ است .



البته ۲ حالت داریم :

شرط : $\lambda \geq \mu$

① $\lambda \rightarrow (\mu, \lambda) \rightarrow \mu$ از جمعیت μ، λ فرزند ایجاد شده

② $\mu+\lambda \rightarrow (\mu+\lambda) \rightarrow \mu$

PAPCO

Scanned by CamScanner

Subject
Date

عیب μ, λ : ممکنه یک موجود خوب در μ باشد که شانسی برای انتخاب شدن نداشته باشد

← μ, λ خاصیت فرار از Min محلی دارد اما در $\mu+\lambda$ امکان گیر کردن در Min محلی

← (Min محلی که در μ بوده دیگه در λ نیست پس فراموش می‌شود)

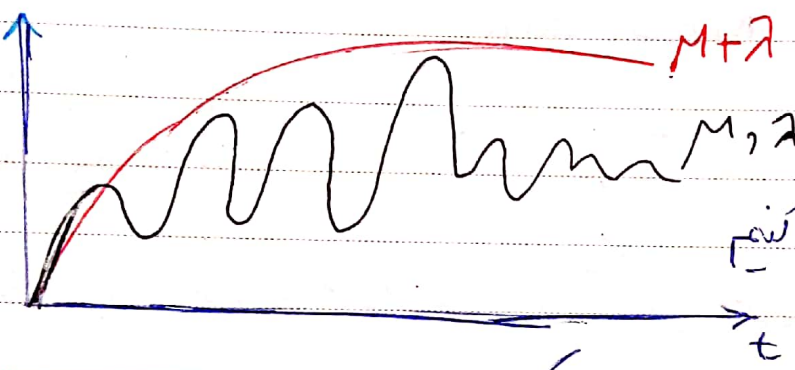
$\mu, \lambda \Longrightarrow$ فراموشکار $\Longrightarrow$ تنوع بیشتر

زیاد است.

$\mu+\lambda \Longrightarrow$ حافظه دارد $\Longrightarrow$ تنوع کمتر $\Longrightarrow$ سرعت همگرایی

بهترین شایسته

هر نسل



$\mu+\lambda \Longrightarrow$ تضمین همگرایی دارد

$\mu, \lambda \longrightarrow$ به تنهایی همگرا نمی‌شود

ما پارامترهای دیگر باید آن را کنترل کنیم

← در $\mu+\lambda$ چون امکان همگرایی اولزمی هست ممکن است به جواب واقعی برسیم.

← μ, λ می‌تواند پارامتر بد را فراموش کند، μ, λ همیشه جای متحرک را دنبال می‌کند.

حذف دائم عوض شود

عیب $\mu+\lambda$ ∞

پس در حالت کلی نمی‌توان گفت کدام خوب و کدام بد است. و بسته به شرایط دارد.

اگر همزمان به جواب برسیم و تنوع دارد خراب می‌شود ⇐ μ, λ

اگر نزدیک جواب هستیم و می‌خواهیم سریع همگرا شویم ⇐ $\mu+\lambda$

PAPCO

Scanned by CamScanner

Subject
Date
الگوریتم تکاملی

$EA(\mu+\lambda)$

$EA(\mu,\lambda)$

$EA(1+1)$

یک موجود داریم از روی یک موجود با جهش یک موجود ایجاد می‌شود و از بین این ۲ یکی انتخاب می‌شود. (شبیه تبه نوردی)

$EA(1,1)$ ⟹ الگوریتم تکاملی نیست

تنوع دارد اما انتخاب نیست پس تکاملی نیست.

Random walk است.

* برای تکاملی بودن هم تنوع و هم انتخاب لازم است.

$EA(1+\lambda)$ یک موجود است λ فرزند ایجاد می‌کند و در نهایت یکی انتخاب می‌شود

$EA(1,\lambda)$ " " و از بین λ فرزند یکی انتخاب می‌شود.

$EA(\mu+1)$ → Steady state

μ تا جمعیت، یک فرزند تولید می‌کند و این فرزند را جایگزین یکی از افراد جمعیت کنیم

Genito ↝ اگر این فرزند جایگزین بدترین در جمعیت μ شود

پس Genito حالت خاص Steady state است.

$EA(\mu,1)$

حتماً $\lambda \geq \mu$ باشد پس این غلط است

PAPCO

Scanned by CamScanner

Subject
Date

Generational Gap ⟹ تعدادی از فرزندان انتخاب نمی‌شوند.

EA(μ+μ) μ موجود داریم، μ فرزند تولید می‌کنیم از بین آن‌ها μ تا را به نسل بعدی می‌دهیم

EA(μ,μ) از بین فرزندان μ تا انتخاب می‌شود. " " "

Generational ← هر نسل جایگزین نسل قبلی می‌شود (یعنی همه فرزندان)

EA(μ+λ) در روش ES معمولاً λ را 7 برابر μ در نظر می‌گیرند.

EA(μ,λ) → λ≥μ

* EP از EA(μ+μ) استفاده می‌کند.

← در این جا ما فقط مشخص می‌کنیم از کجا انتخاب کنیم و هنوز درباره روش‌های انتخاب صحبت نشده است.

انتخاب حتماً نباید متناسب با شایستگی باشد اما بی‌ربط با شایستگی هم نباید باشد.

تنوع را کم می‌کند

روش‌های انتخاب 8

① همه انتخاب شوند
② تصادفی یکنواخت

⟸ شایستگی را در نظر نمی‌گیرند.

* نمی‌شود هم انتخاب والدین و هم انتخاب بازماندگان فقط از این دو روش باشد چون دیگه شایستگی نداریم و در نتیجه تکامل نخواهیم داشت.

Scanned by CamScanner

Subject
Date

$P_1 = \frac{70}{395}$

$\boxed{70 \quad 60 \quad 50 \quad 50 \quad 40} \quad 40 \quad 30 \quad 20 \quad 10 \quad 8 \quad 5 \quad 2$

③ انتخاب برشی

$= \frac{5}{12}$

④ انتخاب برشی متناسب با شایستگی

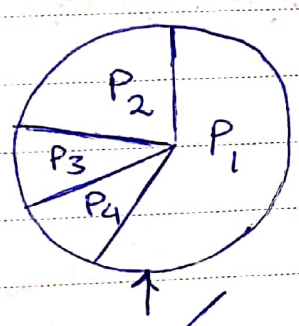
← برای هر موجود یک احتمال تعریف می‌کنیم.

$P_i = \frac{f_i}{\sum_{k=1}^{np} f_k} \qquad \sum_{i=1}^{np} P_i = 1$

انتخاب برشی: افراد را بر اساس شایستگی رتبه می‌کنیم بر اساس این که انتخاب برشی چند درصد باشند از افراد انتخاب می‌کنیم.

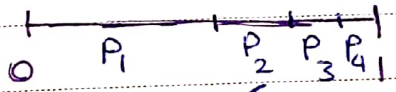
$$n_p = \begin{cases} \mu & \text{انتخاب والدین} \\ \lambda & EA(\mu, \lambda) \\ \mu+\lambda & EA(\mu+\lambda) \end{cases}$$

مثال // roulette wheel



$P_1 = 0.6 \qquad P_3 = 0.1$

$P_2 = 0.2 \qquad P_4 = 0.1$



* به تعداد موجوداتی که می‌خواهیم انتخاب کنیم عدد تصادفی یکنواخت تولید می‌کنیم و می‌بینیم که در کدام بازه می‌افتد.

$n_i \rightarrow P_i$ تعداد موجوداتی که با شایستگی P_i باید انتخاب شوند

از نظر ریاضی: $n_i = N \times P_i$

تعداد کل افرادی که می‌خواهیم انتخاب کنیم

عیب این روش: اگر می‌خواهیم دقیقاً انتخاب متناسب با شایستگی باشد باید تعداد افراد با شایستگی P_i در جمعیت انتخاب شده دقیق برابر با $N \times P_i$ باشد ولی در عمل این اتفاق نمی‌افتد

تعداد کل افرادی که می‌خواهیم انتخاب شوند

*وقتی تعداد موجودات کم است این فرمول تضمین نمی‌شه

Scanned by CamScanner

Subject
Date

انتخاب

باید شوند $ns = \begin{cases} \lambda & \text{انتخاب والدین} \\ \mu & EA(\mu,\lambda) \\ \mu & EA(\mu+\lambda) \end{cases}$

می خواهیم λ تا فرزند داشته باشیم پس
همانطور λ تا والد می خواهیم.

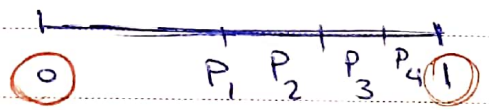
$np \leftarrow$ موجودیتی که داریم و از بین آن ها می خواهیم انتخاب کنیم. $\sum_{i=1}^{np} P_i = 1$

روش دوم: طول خط کش این دفعه بین ۰ تا $1-\frac{1}{ns}$ است.

SUS

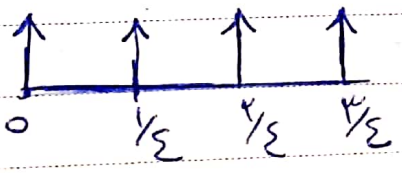
① تعیین P_i $\quad P_i = \frac{f_i}{\sum_{k=1}^{np} f_k}$ $\quad \sum_{i=1}^{np} P_i = 1$

② گذاشتن P_i ها روی خط کش اولیه



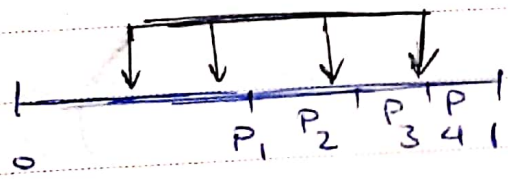
③ انتخاب خط کشی به طول $1-\frac{1}{ns}$ و قسمت بندی آن به صورتی که هر قسمت $\frac{1}{ns}$ باشد.

$ns = 4 \Longrightarrow 1-\frac{1}{ns} = 1-\frac{1}{4} = 3/4$



$\Longrightarrow$ همیشه n_s تا فلش داریم

④ شروع خط کش دوم روی خط کش اول با یک عدد تصادفی یکنواخت بین ۰ و $\frac{1}{ns}$ انتخاب می شود.



$n_1 = [4 \times 0.6] = 2 - 3$

$n_2 = [4 \times 0.2] = 0 - 1$

$n_3 = [4 \times 0.1] = 0 - 1$

$n_4 = [4 \times 0.1] = 0 - 1$

در جدول NP_i در روش SUS تضمین می شود.

PAPCO

Scanned by CamScanner

Subject
Date

← وقتی تعداد زیاد باشد، دو روش با هم فرقی ندارند.

← روش SUS تضمین می‌دهد که در تعداد کم هم $n_i = N \times P_i$ برقرار است.

ولی در روش Roulette wheel چون تعداد کم است این اتفاق نمی‌افتد.

✓ روش متناسب با شایستگی ۲ ایراد دارد: (هم‌چنین انتخاب برتر)

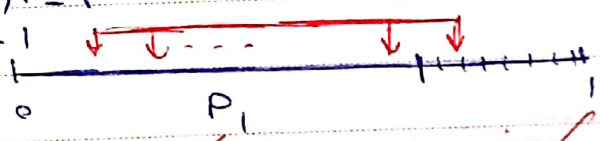
① همگرایی زودرس ② سالم شدن

90 1.1 0.9 1.05 0.95 1.06 0.44 0.98 1.02 1

حداکثر

$P_1 = \frac{90}{100} = 0.9$ $n_1 = 10 \times 0.9 = 9$

$n_{2-10} = 10 \times 0.01 = 1$



$P_{2-10} \approx 0.01$

* مثلاً اگر بخواهیم ۱۰ هم انتخاب کنیم، ۹ تا از P_1 انتخاب می‌شود و یکی از بقیه

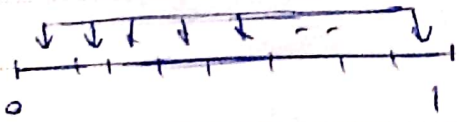
وقتی یک موجود داریم که مقدار شایستگی آن با بقیه خیلی اختلاف دارد ← همگرایی زودرس

$\bar{f} \ll f_{best} \ll f_{optimom}$

جمعیت به آن موجود همگرا می‌شود.

10 10.01 9.99 ----

اختلاف‌ها خیلی کم، همه موجودات نزدیک هم‌اند.



از همه یکی انتخاب می‌شود پس همان جمعیت قبلی

دوباره تکرار می‌شود ← سالم شدن

$\bar{f} \approx f_b \ll f_{optimom}$

* از نظر شایستگی تنوع ندارند.

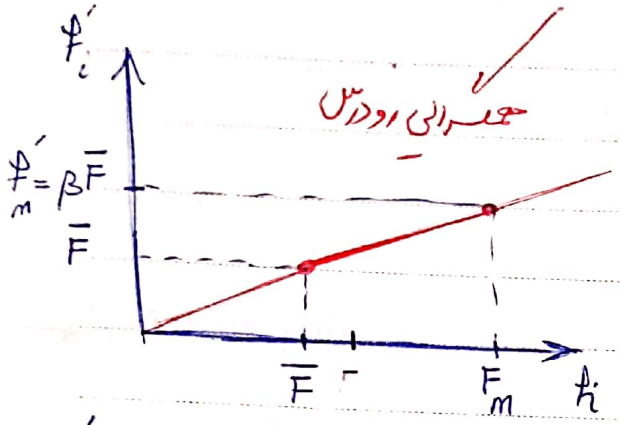
PAPCO

Scanned by CamScanner

Subject
Date

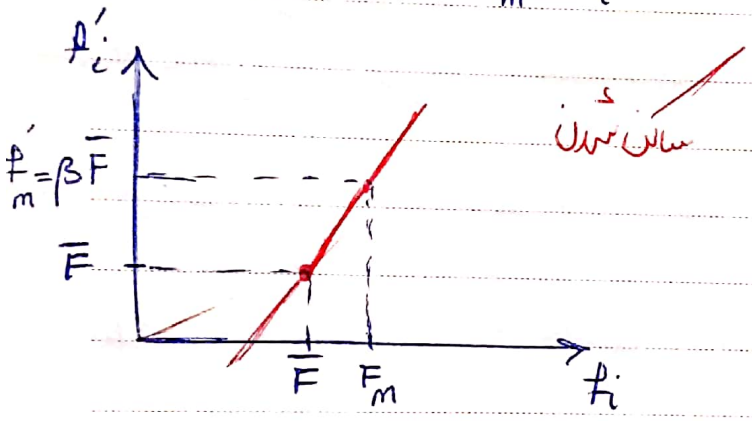
یکی از روش های اینکه شکل همگرایی اودرس و سالن شدن را از بین ببرد، این است

که شایستگی را عوض کنیم.



← شایستگی مجازی به نام f' تعریف می‌کنیم.

$\bar{F}$ همان $\bar{F}$ بماند ولی F_m فاصله‌اش تا $\bar{F}$ کم شود.



F_m فاصله‌اش تا $\bar{F}$ زیاد شود.

$x_1 = \bar{f}$ $\quad y_1 = \bar{f}$

$x_2 = f_m$ $\quad y_2 = \beta \bar{f}$

معادله خط $\Longrightarrow$ $y - \bar{f} = \dfrac{\beta \bar{F} - \bar{F}}{f_m - \bar{f}} (x - \bar{f})$ (y: f_i'، x: f_i)

$$\Longrightarrow f_i' = \bar{f} + \frac{(\beta - 1)}{f_m - \bar{f}} (f_i - \bar{f})$$

← P_i جدید را در روش SUS می‌گذاریم، مشکل حل می‌شود.

$$P_i \text{ جدید} = \frac{f_i'}{\sum_{k=1}^{np} f_k'}$$

۵) انتخاب برشی متناسب با شایستگی تعدیل شده خطی (↑)

PAPCO

Scanned by CamScanner

Subject
Date

۶ متناسب با شایستگی رتبه بندی شده

70 70 60 60 50 45 8 7 2 1

10 9 8 7 6 5 4 3 2 1 ← رتبه ها

به ترتیب Sort می کنیم

$F_i = i$

← بدترین رتبه 1

$$P_i = \frac{f_i}{\sum_{k=1}^{np} f_k} = \frac{f_i}{\sum_{k=1}^{np} k} = \frac{2i}{np(np+1)}$$

← بهترین رتبه np

← عیب این روش هزینه ی Sort است.

← یک فرق مهم که رتبه بندی نسبت به بقیه دارد این است که لازم نیست شایستگی موجودات را به طور دقیق بدانیم، کافی است فقط بدانیم کدام یک شایستگی بیشتری دارند.

روش ۵ و ۶ را می توان به گونه ای تغییر داد که فشار انتخاب را کنترل کند.

فشار انتخاب: $SP = E[\underbrace{nb}_{P_b}] = n_s \cdot P_b$

امید ریاضی تعداد موجوداتی که از بهترین موجودات (best) انتخاب می کنیم. (برای والدین یا نسل بعدی)

n_s ← تعداد موجوداتی که می خواهیم انتخاب کنیم.

فشار انتخاب پارامتر خوبی برای روش های انتخاب است.

SP↓ : سرعت همگرایی کمتر و تنوع بیشتر

SP↑ : سرعت همگرایی بیشتر و تنوع کمتر

PAPCO

Scanned by CamScanner

Subject
Date

شاخص

(۷) $P_i = a P_i + b$ در روش (۵):

(۸) $P_i = a(i) + b$ در روش (۶):

دو معادله برای بدست آوردن a و b می‌خواهیم:

۱) $\sum_{i=1}^{np} P_i = 1$ ۲) $SP = ns \times P_b$

(۷) متناسب با شاخص مقیاس شده خطی با کنترل فشار انتخاب

(۸) متناسب با شاخص رتبه‌بندی شده با کنترل فشار انتخاب

روش ۸:

$$\sum_{i=1}^{np} P_i = \sum_{i=1}^{np} (ai+b) = 1$$

$$\Longrightarrow \left(a \sum_{i=1}^{np} i\right) + b\left(\sum_{i=1}^{np} 1\right) = 1 \Rightarrow a\,\frac{np(np+1)}{2} + b.np = 1$$

$$\frac{np(np+1)}{2} \qquad np \xrightarrow{\text{تقسیم بر } np} a\,\frac{np+1}{2} + b = \frac{1}{np} \quad (1)$$

$$SP = ns.P_b = ns(a.np + b) \Longrightarrow anp + b = \frac{SP}{ns} \quad (2)$$

$$(2) - (1) \longrightarrow a = 2\,\frac{\frac{SP}{ns} - \frac{1}{np}}{np - 1}$$

جایگذاری در رابطه (۱) $\longrightarrow b = \frac{1}{np} - a\,\frac{np+1}{2}$

$$P_i = 2i\,\frac{\frac{SP}{ns} - \frac{1}{np}}{np - 1} + \left(\frac{1}{np} - a\,\frac{np+1}{2}\right)$$

$$P_i = i(a) + \frac{1}{np} - a\,\frac{np+1}{2} = a\left(i - \frac{np+1}{2}\right) + \frac{1}{np}$$

PAPCO

Scanned by CamScanner

Subject
Date

$$P_i = 2\,\frac{\frac{SP}{nS} - \frac{1}{np}}{np-1}\left(i - \frac{np+1}{2}\right) + \frac{1}{np} \qquad 1 < SP < 2$$

← ارش ۷ :

$$\sum_{i=1}^{np} P_i = a\sum_{i=1}^{np} f_i + b(np) = 1$$

$$\frac{a\sum_{i=1}^{np} f_i}{np} + b = \frac{1}{np} \Longrightarrow a\bar{f} + b = \frac{1}{np} \quad (1)$$

$\bar{f}$

$$SP = nS \cdot P_b = (a f_b + b)\,nS \qquad a f_b + b = \frac{SP}{nS} \quad (2)$$

$$(2) - (1) \longrightarrow a(f_b - \bar{f}) = \frac{SP}{nS} - \frac{1}{np}$$

$$P_i = a f_i + \frac{1}{np} - a\bar{f} = a(f_i - \bar{f}) + \frac{1}{np}$$

$$(1),(2) \Longrightarrow b = \frac{1}{np} - a\bar{f}$$

$$\Longrightarrow P_i = \frac{\frac{SP}{nS} - \frac{1}{np}}{f_b - \bar{f}}(f_i - \bar{f}) + \frac{1}{np}$$

Q... tournament (۹)

از بین جمعیت Q تا به صورت تصادفی انتخاب می‌کنیم و بهترین را از بین آن‌ها انتخاب می‌کنیم و به تعداد افرادی که می‌خواهیم این کار را تکرار می‌کنیم

Q↑ : سرعت همگرایی بیشتر و تنوع کمتر

Q↓ : تنوع ↑ و سرعت همگرایی ↓

۲ نوع:
- با جای‌گذاری (اونی که انتخاب کردیم را به جمعیت برگردانیم)
- بدون جای‌گذاری

PAPCO

Scanned by CamScanner

Subject
Date

بدون جایگذاری ← تعادل بیشتر — SP = 1 (فقط بهترین را بسیار انتخاب می‌کنیم)

با جایگذاری ← ns تا از بهترین برمی‌داریم. — SP

Q = 1 ⇐ تصادفی نمی‌خواهد ⇐ Binary tournament ⇐ Q = ۲

۲ < Q < ۵

در این روش هم لازم نیست شایستگی را بدانیم

مزیت آن نسبت به رتبه‌بندی این است که Sort نداریم و فقط Max را می‌خواهیم. (از جمعیت ۲ تا ۵ تایی)، همهٔ مزایای روش‌های قبل را دارد و عیب آن‌ها را ندارد.

اگر بخواهیم تنوع حفظ شود، بدون جایگذاری بهتر است.

۳ روش کنترل پارامتر:

Deterministic : (قطعی)

Adaptive : (تطبیقی)

Self-Adaptive : (خود تطبیقی)

در ES هر ۳ نوع وجود دارد.

Genetic Algorithm

شناسنامه‌ی یک الگوریتم تکاملی، یعنی در هر قسمت از چه الگوریتمی استفاده می‌کنیم.

PAPCO

Scanned by CamScanner

Subject
Date

سانسانه :

GA

بازنمایی	باینری (در بعضی موارد Gray Code)
جمعیت اولیه	تصادفی یکنواخت $\mu = 4$
انتخاب والدین	متناسب با شایستگی P_i SUS & RW $\lambda = \mu$
بازترکیبی	$P_c \geq 0.8$ ← احتمال اینکه بازترکیبی انجام بشه
جهش	$\frac{1}{\mu \times \text{تعداد ژن}} < P_m < \frac{1}{\mu}$
ارزیابی	$B \rightarrow x \rightarrow f$
انتخاب بازماندگان	λ , μ همه انتخاب شوند
شرط خاتمه	تعداد ارزیابی

PAPCO

Scanned by CamScanner

Subject
Date

EXP // ۲۰ تا شهر داریم ، ۴ تا فرودگاه ، فرودگاه ها چطوری باشند تا فاصله ی هر شهر تا نزدیک ترین فرودگاه اش Min باشد ؟

شهرها x_i y_i nc

فرودگاه ها x_{aj} y_{aj} na تعداد

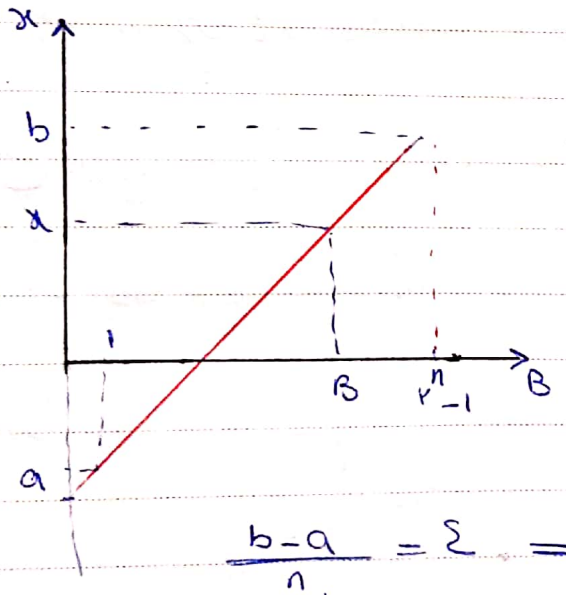
$[0 - 2^{nb} - 1]$ بیت nb

کروموزوم A این داره ⟷

x_{a1}	y_{a1}	x_{a2}	y_{a2}	x_{a3}	y_{a3}	x_{a4}	y_{a4}

تابع هدف :

$$\sum_{i=1}^{nc} \min_{j=1}^{na} \left\{ (x_i - x_{aj})^2 + (y_i - y_{aj})^2 \right\} \rightarrow \text{Min شود}$$



$$x = a + \frac{b-a}{2^n - 1} B$$

$$\text{if } B = 1 \Rightarrow x = a + \frac{b-a}{2^n - 1}$$

$$\frac{b-a}{2^n - 1} = 4 \Longrightarrow \frac{b-a}{4} = 2^n - 1$$

$$\Longrightarrow \frac{b-a}{4} + 1 = 2^n \Longrightarrow n \ln(2) = \ln\left(\frac{b-a}{4} + 1\right)$$

تعداد بیت هایی که برای مدل کردن لازم داریم

PAPCO

Scanned by CamScanner

Subject
Date

جمعیت اولیه : برای هر x_{a_i} ، y_{a_i} ، ۱۰ بیت در نظر می‌گیریم.

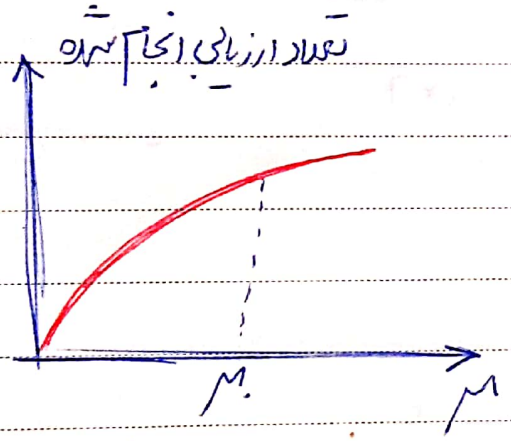
فرض : $n_a = 2$ ۲ فروشگاه ، برای هر کدام یک x و یک y داریم ⟸ ۴۰ بیت

$M = 3$

x_{1a}	y_{1a}	x_{2a}	y_{2a}

برای تولید جمعیت اولیه به طور تصادفی ۰ و ۱ بگذاریم.

برای تعیین موقعیت شهر یا خط کنیم



ایجاد تنوع از دو طریق : ① iteration ها ② جمعیت

ایجاد تنوع از طریق جمعیت تا یک جای بهبودی دهد.

← فقط برای ES اثبات شده که تا $\lambda = 7M$ برای تنوع خوب است. بعد از آن دیگر تنوع نداریم.

تابع هدف : $$\text{Min} \sum_{i=1}^{N_c} \min_{j=1}^{N_a} \left[(x_i - x_{ja})^2 + (y_i - y_{ja})^2 \right]$$

x_i: x شهر ، x_{ja}: x فروشگاه ، y_i: y شهر ، y_{ja}: y فروشگاه

$$B = b_0 \times 2^0 + \dots + b_{n-1} 2^{n-1}$$

مرحله انتخاب والدین :

	20	15	25
شانسیلی			

PAPCO

Scanned by CamScanner

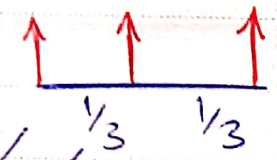
Subject
Date

$$P_1 = \frac{\frac{1}{20}}{\frac{1}{20}+\frac{1}{15}+\frac{1}{25}} \qquad P_2 = \frac{\frac{1}{15}}{\frac{1}{20}+\frac{1}{15}+\frac{1}{25}}$$

$$P_3 = \frac{\frac{1}{25}}{\frac{1}{20}+\frac{1}{15}+\frac{1}{25}}$$

** در اللوسم ژنتیک M باید زوج باشد

خط لش :



یک عدد تصادفی بین ۰ و ۱ تولید می‌کنیم اگر از ۰/۸ کمتر بود بازترکیبی انجام می‌دهیم اگر بزرگتر بود انجام نمی‌دهیم.

M	P_1	P_2	P_3	P_4
والدین	P_2	P_1	P_2	P_3
بازترکیبی	P_2	P_1	P_2	P_3
جمعیت λ	P_2''	P_1'	P_2	P_3'

مثلاً بازترکیبی P_1 و P_2 :

* اگر برش از وسط ژن باشد، ژن تغییر می‌کند ولی اگر از محل تمام شدنش باشد، ژن جابه‌جا می‌شود.

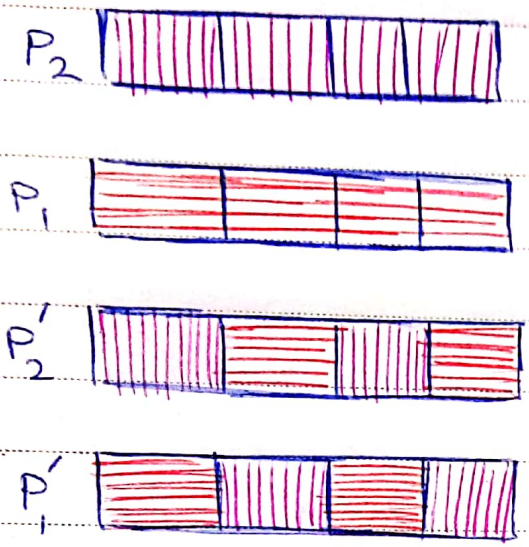
← در صورت تغییر ژن، ژنگی جدید ایجاد می‌شود ولی در صورت جابه‌جایی فقط ژنگی‌ها جابه‌جا می‌شوند

← در شکل اول از صفحه بعد برش‌ها دقیقاً از محل اتمام ژن ها صورت گرفته و ژن‌ها فقط جابه‌جا شده‌اند.

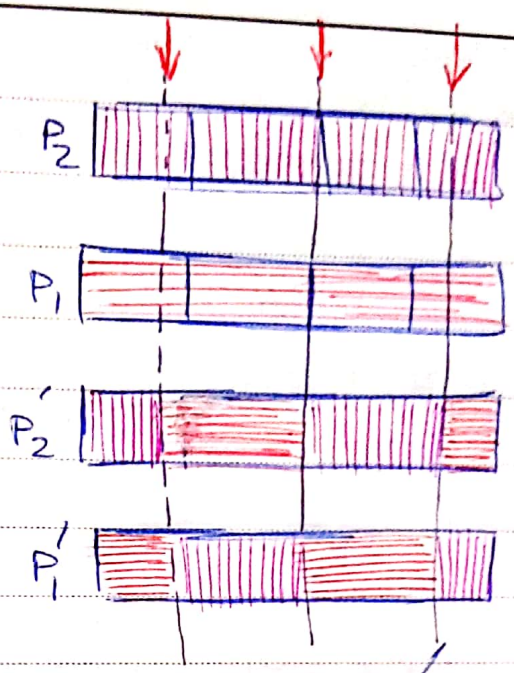
PAPCO

Scanned by CamScanner

Subject
Date



شکل (۱)



شکل (۲)

حدس: $P_m = 0.05$ $\quad \frac{1}{40} < P_m < \frac{1}{4}$

۱۶۰ تا عدد تصادفی بین ۰ و ۱ تولید می‌کنیم (۱۶۰ بیت داریم) ← اگر عدد کمتر از ۰.۰۵ بود

اون بیت رو toggle می‌کنیم (اگر صفر بود یک می‌کنیم و برعکس)

سوال: برای بازنمایی، چرا Grey code از binary بهتر است؟

چون تغییر یک بیت در باینری، عدد را خیلی تغییر می‌دهد ولی تغییر یک بیت در grey code عدد را تغییر زیادی نمی‌دهد ← در همسایگی جابه‌جا می‌شویم.

Grey code ← جهت و محلی (جهش)

Binary ← جابه‌جایی عمومی

بازرگشی ← جست و جوی عمومی

← چرا بازرگشی می‌خواهد بهتر از نقطه‌ای است؟ در بازرگشی چند نقطه‌ای احتمال اینکه ابتدا و انتهای کروموزوم تغییر کند، خیلی کم است.

PAPCO

Scanned by CamScanner

Subject
Date

الگوریتم ES :

پارامترها	اعداد حقیقی
جمعیت اولیه	تصادفی یکنواخت μ
انتخاب والدین	تصادفی یکنواخت $\lambda = 7\mu$
جهش (نویز نرمال)	$P_m = 1$ روی همه‌ی موجودات جهش انجام می‌شود. (نویز نرمال)
بازترکیبی	$P_c < 0.4$ متوسط‌گیری
ارزیابی	$x \to C(x) \to 1/C(x)$
انتخاب بازماندگان	متناسب با شایستگی λ و μ $\mu+\lambda$ تورنومنت Q
شرط خاتمه	تعداد ارزیابی

* جهش قبل از بازترکیبی انجام می‌شود، چون خیلی به هم شبیه‌اند.

x_{1a}	y_{1a}	x_{2a}	y_{2a}	x_{3a}	y_{3a}	x_{4a}	y_{4a}

$x_{ja} = a + (b-a)\,R[0,1]$ → عدد تصادفی بین a, b

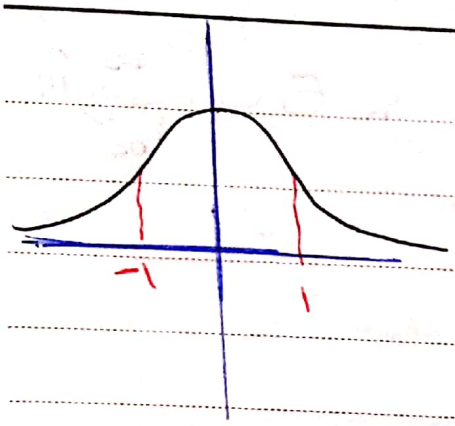
$y_{ja} = c + (d-c)\,R[0,1]$

تصادفی یکنواخت ← شانس انتخاب موجودات با هم برابره

PAPCO

Scanned by CamScanner

Subject
Date



تصادفی نرمال ← با احتمال ۶۳٫۳ بین ۱- و ۱

با احتمال حدود ۹۰ بین ۲- و ۲

← متوسط قدم‌ها در جهش : $|x'_{ia} - x_{ia}|$

← با تعداد تکرار زیاد به طور متوسط σ می‌شود

(یعنی)

$$x'_{ia} = x_{ia} + \sigma N(0,1)$$

x_{ia} جهش یافته

چون متوسط توزیع نرمال صفر است و در یک σ ضرب شده، در نتیجه متوسط آن σ می‌شود.

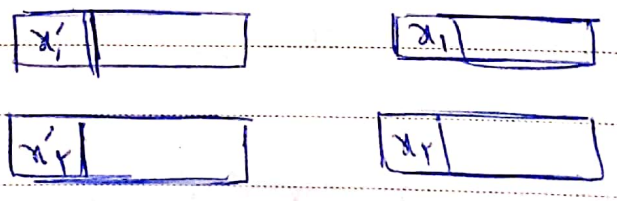
در ES به اندازه σ اندازه جهش می‌کنیم ← چه σ بزرگتر باشد قدم‌های بزرگتری برمی‌داریم.

بازترکیبی :

$$x'_1 = \alpha x_1 + (1-\alpha) x_2$$

$$x'_2 = (1-\alpha) x_1 + \alpha x_2$$

$$0 < \alpha < 1$$

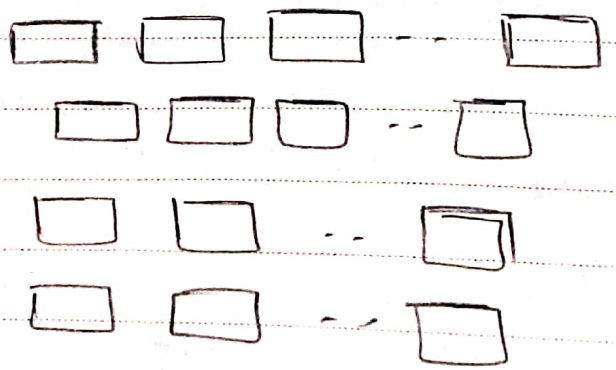


۱) ۳ والد

۲) ۷ والد

۳)

۴)



۱۰ تا

۷۰ تا

۷۰ تا جهش یافته

بازترکیبی

PAPCO

Scanned by CamScanner

Subject
Date

$x_1' = \alpha x_1 + (1-\alpha) x_2$

$x_2' = (1-\alpha) x_1 + \alpha x_2$

$0 < \alpha < 1$

۳ روش وجود دارد :

① برای یک یک آن ها این کار انجام بشه

② از یک جایی به بعد انجام شود

③ نقطه برای یک آن

علت اینکه در $\mu+\lambda$ از Q turnoment استفاده می‌شود : تعادل بین تنوع و همگرایی

$\mu+\lambda$ سرعت همگرایی دارد ، Q تورنمنت تنوع دارد ⟸ باید هر دو را داشته باشیم و یک جمعیت

کار نکنیم .

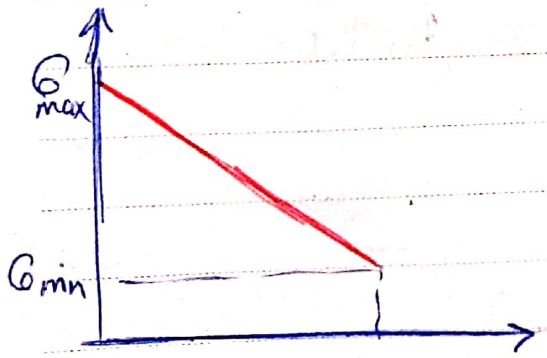
۳ روش برای یافتن σ : ① deterministic ② تطبیقی Adaptive ③ خود تطبیقی

$$\sigma_t = \sigma_{max} + \frac{\sigma_{min} - \sigma_{max}}{N} \times t \quad : \text{deterministic}$$

σ_{min} ← با چه دقتی جواب را می‌خواهی ؟

σ_{max} ← معمولاً ضریبی از بیشترین فاصله را در نظر می‌گیرند .

تعیین σ_{min} و σ_{max} راحت است .



PAPCO

Scanned by CamScanner

Subject
Date

روش Adaptive برای ES(1+1) ← یک دالد تحقیق می‌دهیم، هر کدوم بهتر بود

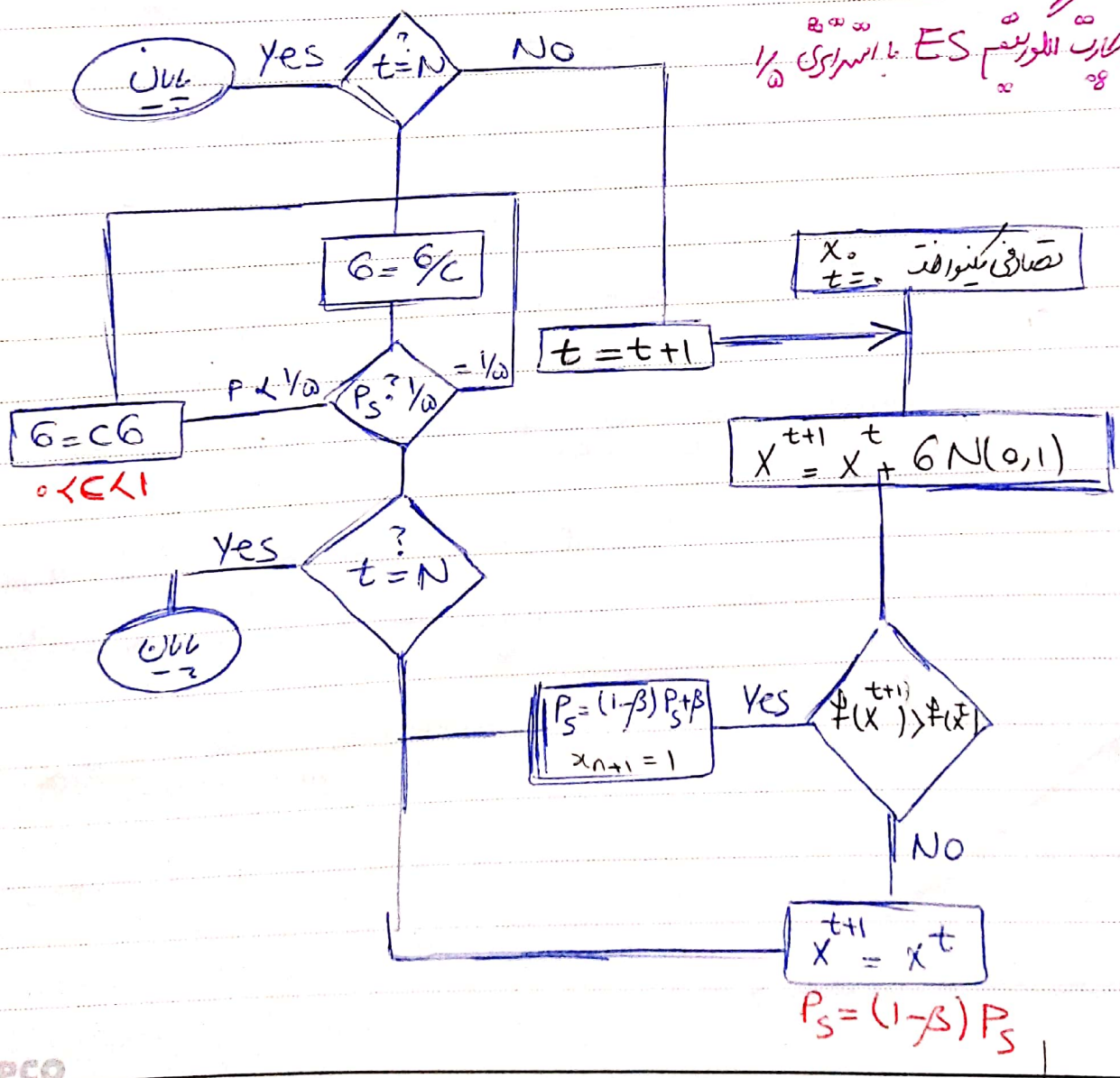
برای مرحله بعد می‌رود. مثل تپه نوردی فقط یک استراتژی برای σ دارد.

(استراتژی $\frac{1}{5}$ فقط برای ES(1+1) اثبات شده است)

اثبات شده که اگر تعداد موفقیت‌ها برابر $\frac{1}{5}$ باشد به همگرایی Max می‌شود.

تعداد موفقیت $\frac{1}{5}$ یعنی ۵ بار ۱ فرزند انتخاب شود.

فلوچارت الگوریتم ES با استراتژی $1/5$



Scanned by CamScanner

Subject
Date

← چگونه می‌توان تعداد موفقیت‌ها را روی σ نگه داشت؟ با اندازه قدم‌ها می‌توانیم موفقیت‌ها را کم یا زیاد کنیم

هرچه قدم‌ها بزرگتر باشند تعداد موفقیت‌ها کم‌تر می‌شود و برعکس.

(وقتی از جواب دور هستی، هر قدمی برداری می‌تونه موفقیت محسوب شود)

احتمال موفقیت را چگونه حساب کنیم؟ P_s احتمال موفقیت

$$x_1 \quad x_2 \quad \cdots \quad x_n \qquad \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$x_1 \quad x_2 \quad \cdots \quad x_n \quad x_{n+1} \qquad \bar{x}_2 = \frac{n\bar{x} - \overbrace{x_1}^{\simeq \bar{x}} + x_{n+1}}{n}$$

$$\Longrightarrow \bar{x}_2 = \frac{n\bar{x} - \bar{x} + x_{n+1}}{n} = \frac{n-1}{n}\bar{x} + \underbrace{\frac{1}{n}}_{\beta} x_{n+1}$$

$$= (1-\beta)\bar{x} + \beta x_{n+1}$$

< متوسط‌گیری پنجره‌ای >

← وقتی موفقیت را یک می‌گیریم، عدم موفقیت را صفر می‌گیریم. این فرمول در واقع احتمال موفقیت را حساب می‌کند.

استراتژی خود تطبیقی

x_{1a}	- - - -	y_{4a}	σ

σ را داخل کروموزم بگذاریم.

روی σ بازترکیبی انجام نمی‌دهیم، برای جهش داریم:

$$\sigma' = \sigma e^{\tau N(0,1)}$$

ضریب یادگیری

PAPCO

Scanned by CamScanner

سوال 8 اول σ را جهش بدهیم بعد x یا برعکس؟

سوال 9 چرا x را قطعی جهش می‌دهیم، σ را نمی‌ای؟ یک دلیل: σ نباید منفی شود چون

اندازه قدم است $e^{\tau N(0,1)}$ عددی که تقریباً کمی از 1 بزرگتر یا کوچکتر است.

در نتیجه باعث می‌شود σ در هر Step تغییرات زیادی نکند.

ضریب یادگیری هم کوچک است تا یادگیری تدریجی انجام شود.

← اگر اول x را جهش بدهیم، اثر σ جهش یافته را در نسل بعد خواهیم دید، اما اگر اول

σ را جهش دهیم، اثر آن در همین نسل دیده می‌شود.

برای اینکه بفهمیم σ خوب یا نه باید ابتدا قدم را برداریم بعد ببینیم x تغییر خوبی دارد یا نه ←

باید اول σ را جهش داد. خوب و بد بودن σ و x هر دو از روی x مشخص می‌شود.

اگر اول x را جهش بدهیم، این نسلی که داریم ارزیابی می‌کنیم با σ ای است که دیگر

بیرون انداختیمش!!

EA

بازترکیبی ← جایگشت: باید جهش و بازترکیبی به گونه ای باشد که جایگشت بهم نخورد.

2	1	3	4

PAPCO

Scanned by CamScanner

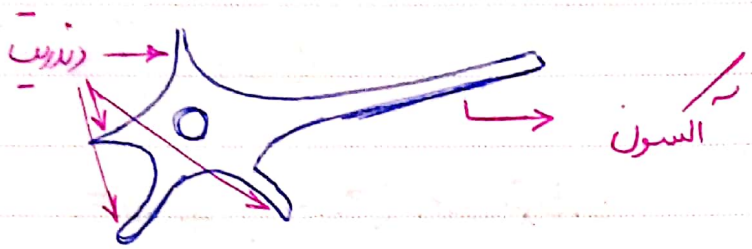
Subject
Date

حسیت وضعی عمومی ← یک زیر مجموعه ای از حسیت انتخاب کن های همه رو تغییر بده

حسیت وضعی محلی ← برای حسش یک نقطه رو انتخاب کن بچاشو با همسایه اش عوض کن

شبکه های عصبی

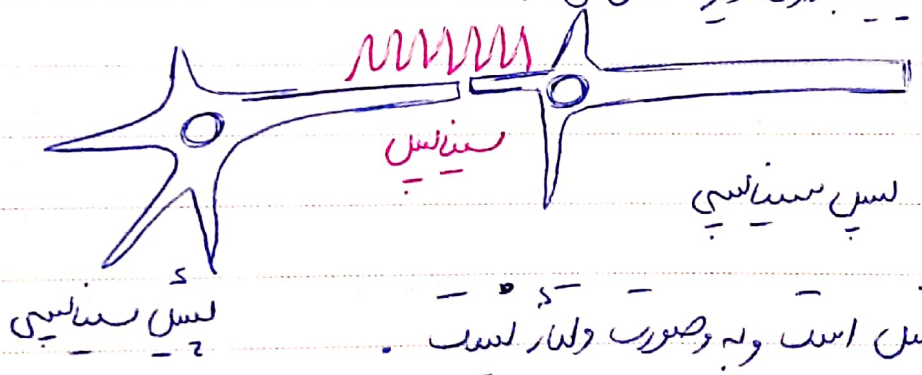
نورون ها ← سلول های عصبی



خروجی ← آکسون

ورودی ها ← دندریت

← آکسون از طریق اتصال شیمیایی به نورون دیگر متصل می شود.



خروجی یک نورون به صورت فرکانس است و به صورت ولتاژ نیست.

و تعداد این ها دارای اهمیت است و دامنه آن ثابت است.

یادگیری یعنی تغییر تعداد سوراخ ها در سیناپس.

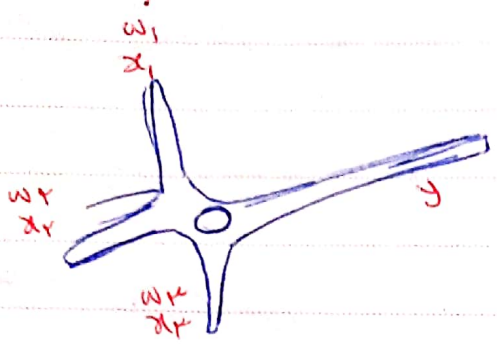
ورودی تبدیل به ولتاژ شده خروجی فرکانسه !!

PAPCO

Scanned by CamScanner

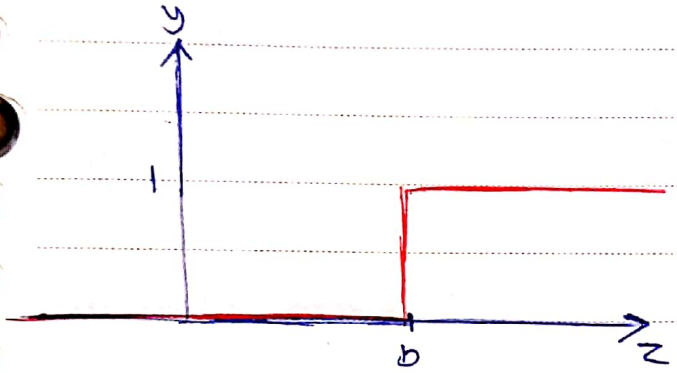
Subject
Date

وزن هر ورودی



ساده‌ترین مدلی که به صورت ریاضی برای نورون وجود دارد :

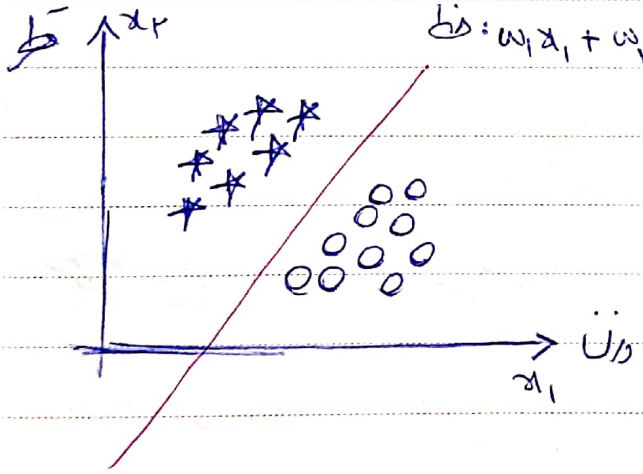
$$\begin{cases} y=1 & \overbrace{\sum_{i=1}^{n} w_i x_i}^{z} > b \\ y=0 & \sum_{i=1}^{n} w_i x_i < b \end{cases}$$



< تابع فعالیت نورون >

$$0.5 \times Sgn(z-b) + 0.5$$

فرض کنید مقداری سیب و پرتقال داریم، می‌خواهیم وزن و قطر آن‌ها را نشان دهیم.



پرتقال $w_1x_1 + w_2x_2 > b$

سیب $w_1x_1 + w_2x_2 < b$

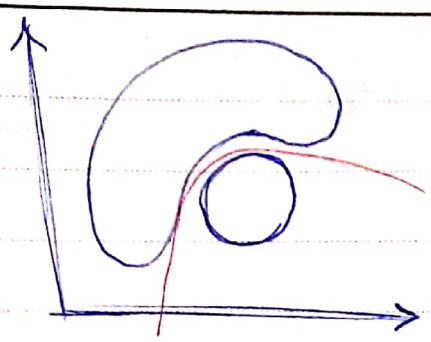
پس یک نورون به تنهایی تصمیم می‌گیرد که سیب یا پرتقال!

یک نورون در صورتی که خطی جدا شده باشند، می‌تواند تصمیم بگیرد. مثلاً در صورتی که به شکل زیر باشند، یک نورون نمی‌تواند. اما ثابت شده شبکه‌ای از نورون‌ها می‌تواند

PAPCO

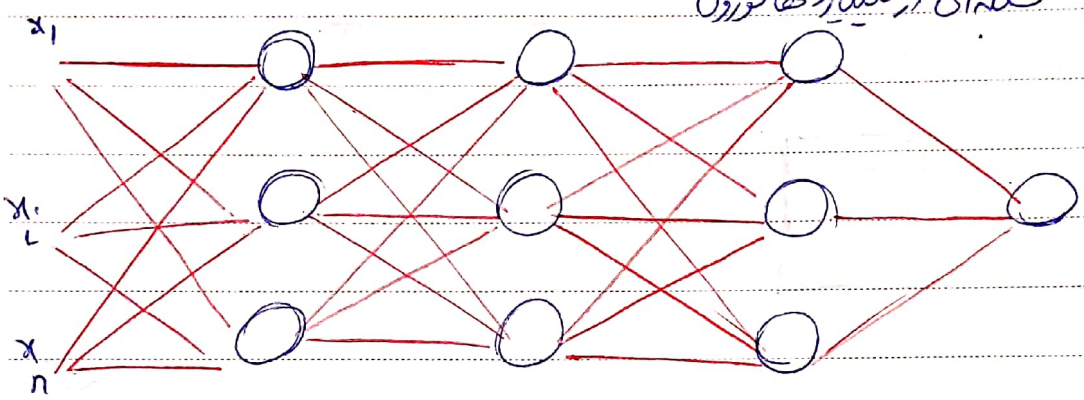
Scanned by CamScanner

Subject
Date

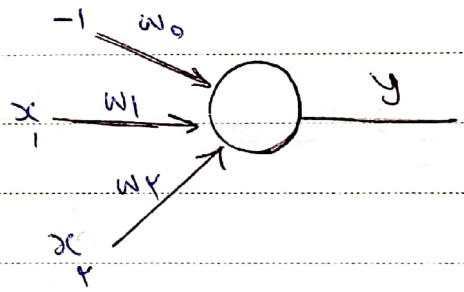


تابع غیرخطی را ادل کند.

شبکه نورون‌ها: شبکه‌ای از میلیاردها نورون



$$-w_0 + w_1 x_1^1 + w_2 x_2^1 = 1$$
$$-w_0 + w_1 x_1^l + w_2 x_2^l = 0$$
$$-w_0 + w_1 x_1^L + w_2 x_2^L = 1$$



$$\begin{bmatrix} -1 & x_1^1 & x_2^1 \\ \vdots & \vdots & \vdots \\ -1 & x_1^l & x_2^l \\ \vdots & \vdots & \vdots \\ -1 & x_1^L & x_2^L \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_1 \\ \vdots \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$\underset{L\times 3}{X} \times \underset{3\times 1}{W} = \underset{L\times 1}{Y}$$

$$XW = Y$$

$$E = \frac{1}{2}(XW - Y)^T (XW - Y)$$

$$e = \begin{bmatrix} -w_0 + w_1 x_1^1 + w_2 x_2^1 - 1 \\ \vdots \\ -w_0 + w_1 x_1^l + w_2 x_2^l - 0 \\ \vdots \\ -w_0 + w_1 x_1^L + w_2 x_2^L - 1 \end{bmatrix}$$

PAPCO

Scanned by CamScanner

Subject
Date

E تابع هدف ای است که می خواهیم آن را Min کنیم.

$$E = \frac{1}{2}(XW-y)^T(XW-y)$$

$$\frac{\partial E}{\partial W} = \frac{1}{2}(2)\left(X^T(XW-y)\right) = 0$$

$$X^T(XW-y) = 0 \longrightarrow X^TXW = X^Ty$$

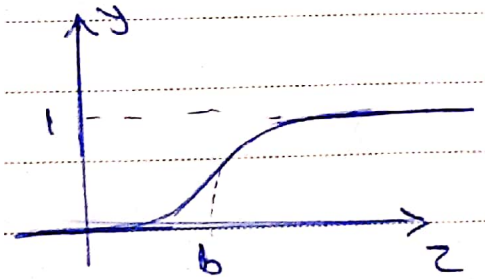
$$\Longrightarrow \boxed{W = (X^TX)^{-1}X^Ty}$$

W را به گونه ای بدست آوردیم که مجموع مربعات خطا را Min کند.

مشتق ماتریسی: $y = (\underset{1\times m}{X}\underset{m\times 1}{A})^T(XA)$

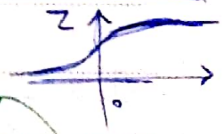
$$\frac{dy}{dA} = 2X^TXA$$

* برای اینکه تابع فعالیت لوژیستیک مشتق پذیر باشد، آن را به شکل زیر در می آوریم:



سیگموئید:

$$y = \frac{1}{1+e^{-(z-b)}} \xrightarrow{\text{از صفر شد}} \boxed{\frac{1}{1+e^{-z}}}$$



$$\sum_{i=1}^{n} w_i x_i > b \overset{\text{برای سادگی}}{\equiv} \sum_{i=0}^{n} w_i x_i > 0$$

$$\boxed{\begin{array}{l} w_0 = b \\ x_0 = -1 \end{array}}$$

Scanned by CamScanner

Subject
Date

تابع سیگموید

لایه دوم

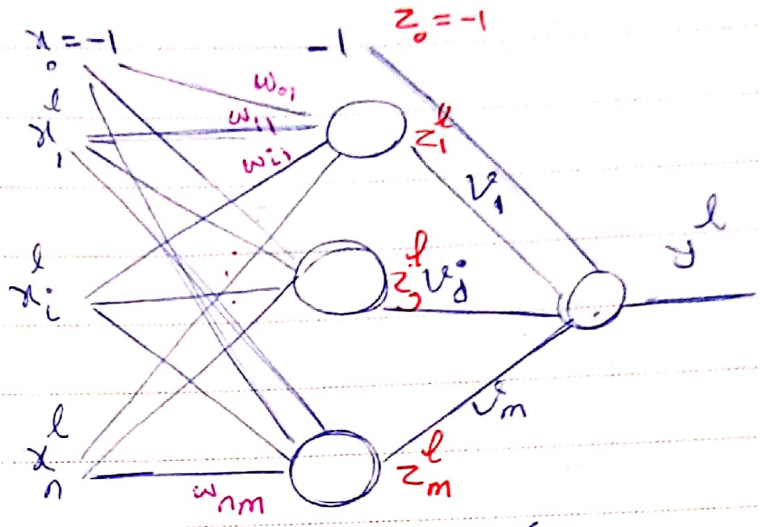
$$z_j^l = S\left(\sum_{i=0}^{n} w_{ij}^{2} x_i^l\right)$$

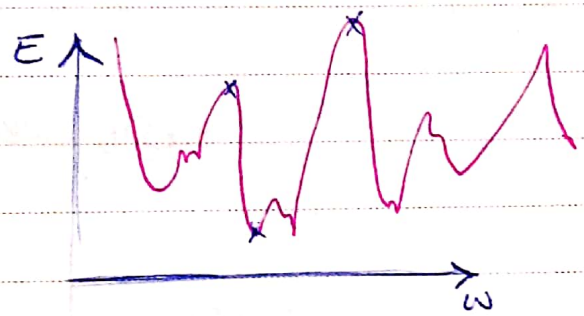
$$y^l = S\left(\sum_{j=0}^{m} v_j z_j^l\right)$$

$$E = \frac{1}{2}\sum_{l=1}^{L}\left(y^l - y^{*l}\right)^2$$

خطا

مجموع مربعات





یکی از روش های حل گرادیان است.

← شکل: در Min محلی گیر می کند.

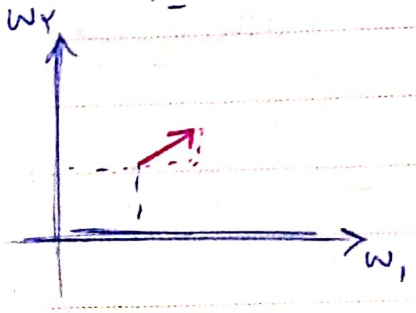
$$\nabla_w E = \begin{bmatrix} \frac{\partial E}{\partial w_{01}} \\ \vdots \\ \frac{\partial E}{\partial w_{nm}} \end{bmatrix}$$

بردار گرادیان

جهت گرادیان ← جهت بیشترین شیب (بیشتر شدن) است.

پس مثلاً وقتی می خواهیم به قله برسیم، در جهت گرادیان حرکت می کنیم.

پس برای اینکه در جهت کم تر شدن حرکت کنیم، باید در جهت عکس بردار گرادیان حرکت کنیم.



PAPCO

Scanned by CamScanner

Subject
Date

روش گرادیان :

اگر بردار w به شکل زیر تعریف کنیم :

$$w = \begin{bmatrix} w_{01} \\ w_{11} \\ \vdots \\ w_{mn} \\ v_0 \\ \vdots \\ v_m \end{bmatrix} \longrightarrow \nabla_w E = \begin{bmatrix} \frac{\partial E}{\partial w_{01}} \\ \\ \frac{\partial E}{\partial w_{mn}} \\ \vdots \\ \frac{\partial E_m}{\partial v_m} \end{bmatrix}$$

در حالت برداری :

$$\overset{new}{w^{(t+1)}} = \overset{old}{w^{(t)}} - \eta \nabla_w E \Big|_{w^t}$$

در حالت غیر برداری :

$$w_{ij}^{new} = w_{ij}^{t} - \eta \frac{\partial E}{\partial w_{ij}}$$

برای یک یک حساب کنیم

$\eta \leftarrow$ یک ضریب که بستگی دارد طول قدم ها چه قدر باشد.

یادگیری آنلاین $\leftarrow$ یعنی دونه دونه انجام بدیم . درست : سیستم همیشه در حال یادگیری است

یادگیری batch $\leftarrow$ از اول کل را قرار می دهیم . (Σ داریم)

(اگر batch داریم)

$$\frac{\partial E}{\partial v_j} = \sum_{l=1}^{L} (y^l - y^{*l}) \frac{\partial y^l}{\partial v_j}$$

$$y^l = S(u) = \frac{1}{1+e^{-u}} \Longrightarrow S'(u) = \frac{e^{-u}}{(1+e^{-u})^2} = (1-S(u))\,S(u)$$

PAPCO

Scanned by CamScanner

Subject
Date

$$\rightarrow \frac{\partial y^l}{\partial v_j} = \frac{\partial y^l}{\partial u} \times \frac{\partial u}{\partial v_j}$$

$$\Longrightarrow \frac{\partial E}{\partial v_j} = \sum_{l=1}^{L} (y^l - y^{*l})\, s'\left(\sum_{j=0}^{m} v_j z_j^l\right) z_j^l$$

$$\rightarrow v_j^{t+1} = v_j^t - \eta \underbrace{\sum_{l=1}^{L} (y^l - y^{*l})\, s'\left(\sum_{p=0}^{m} v_p z_p^l\right)}_{e} z_j^l$$

وزن های لایه آخر با روش گرادیان

$$\frac{\partial E}{\partial w_{ij}} = \sum_{l=1}^{L} (y^l - y^{*l}) \frac{\partial y^l}{\partial w_{ij}}$$

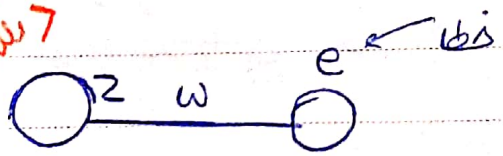
① $$\rightarrow \frac{\partial y^l}{\partial w_{ij}} = s'\left(\sum_{p=0}^{m} v_p z_p^l\right) v_j \frac{\partial z_j^l}{\partial w_{ij}}$$

② $$\rightarrow \frac{\partial z_j^l}{\partial w_{ij}} = s'\left(\sum_{s=0}^{n} w_{sj} x_s^l\right) x_i^l$$

②, ① $$\rightarrow w_{ij}^{new} = w_{ij}^{old} - \eta \overbrace{\sum_{l=1}^{L} (y^l - y^{*l})\, s'\left(\sum_{p=0}^{m} v_p z_p^l\right) v_j\, s'\left(\sum_{s=0}^{n} w_{sj} x_s^l\right)}^{e} x_i^l$$

فرمول آپدیت w

مدل ∞ لایه

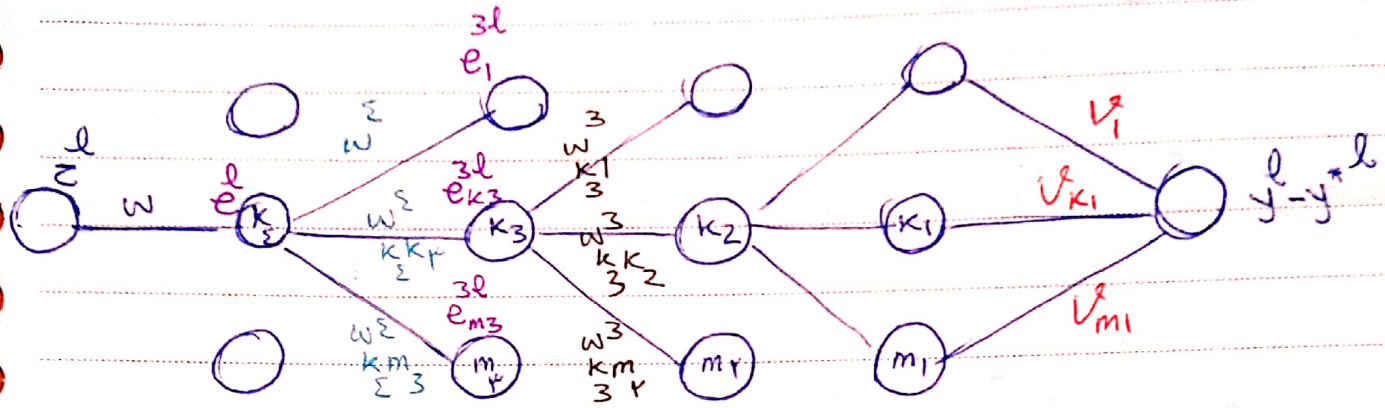


$$w^{t+1} = w^t - \eta z e$$

PAPCO

Scanned by CamScanner

Subject
Date



: back propagation

$$\boxed{w^{new} = w^{old} - \eta \sum_{l=1}^{L} z^l e^l}$$

$$e^l = \sum_{k_3=1}^{m_3} w_{k_4k_3} e^{3l}_{k_3} S'_{k_4}$$

مشتق در همان نقطه e^l

به همین ترتیب .....

$$e^{3l} = \sum_{k_2=1}^{m_2} w^3_{k_3k_2} e^{2l}_{k_2} S'_{k_3}$$

$$E = \frac{1}{2}(y-y^*)^T(y-y^*) = \frac{1}{2}\sum_{l=1}^{L}(y^l - y^{*l})^2$$

$$y = \begin{bmatrix} y^1 \\ \vdots \\ y^L \end{bmatrix}_{L\times 1} \qquad y^* = \begin{bmatrix} y^{*1} \\ \vdots \\ y^{*L} \end{bmatrix}$$

خروجی شبکه عصبی ← y

PAPCO

Scanned by CamScanner

Subject
Date

روش ۱ لوس نیوتن :

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_{1000} \end{bmatrix}$$

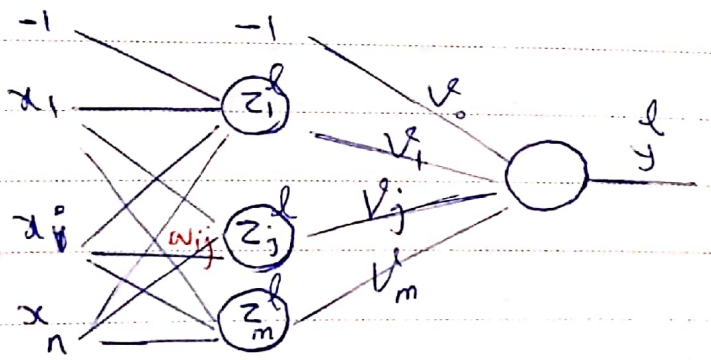
ماتریس ژاکوبین

$$\frac{dy}{dw} = \begin{bmatrix} \frac{\partial y^1}{\partial w_1} & \cdots & \frac{\partial y^1}{\partial w_i} & \cdots & \frac{\partial y^1}{\partial w_{1000}} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial y^l}{\partial w_1} & \cdots & \frac{\partial y^l}{\partial w_i} & \cdots & \frac{\partial y^l}{\partial w_{1000}} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial y^L}{\partial w_1} & \cdots & \frac{\partial y^L}{\partial w_i} & \cdots & \frac{\partial y^L}{\partial w_{1000}} \end{bmatrix}$$

بسط تیلور خطی : $y - y^t = J_{L\times 100}(w - w^t)$

ماتریس ژاکوبین

EXP



$$w = \begin{bmatrix} w_{01} \\ w_{11} \\ \vdots \\ w_{i1} \\ \vdots \\ w_{im} \\ \vdots \\ w_{n1} \\ \vdots \\ w_{nm} \\ v_0 \\ v_1 \\ \vdots \\ v_m \end{bmatrix}$$

$$y^l = S\left(\sum_{j=0}^{m} v_j z_j^l\right)$$

$$z_j^l = S\left(\sum_{i=0}^{n} w_{ij} x_i^l\right)$$

PAPCO

Scanned by CamScanner

Subject
Date

$$\frac{\partial y^l}{\partial w_{ij}} = S'\left(\sum_{j=0}^{m} v_j z_j^l\right) v_j \, S'\left(\sum_{i=0}^{n} w_{ij} x_i^l\right) x_i^l$$

$$\frac{\partial y^l}{\partial v_j} = S'\left(\sum_{j=0}^{m} v_j z_j^l\right) z_j^l \qquad j=0,\dots,m$$

$i=0,\dots,n \qquad l=1,\dots,L$
$j=1,\dots,m$

y در لحظه t

طبق فرمول : $Y - Y(t) = J(t)\,(W - W(t))$

$$Y = J(t)\underbrace{(W - W(t))}_{\Delta W} + Y(t)$$

$$E = \frac{1}{2}\left(Y(t) - Y^* + J\Delta W\right)^T \left(Y(t) - Y^* + J\Delta W\right)$$

$$E = \frac{1}{2}\left(\Delta Y + J\Delta W\right)^T \left(\Delta Y + J\Delta W\right)$$

$$\frac{\partial E}{\partial \Delta W} = J^T(\Delta Y + J\Delta W) = 0 \rightsquigarrow J^T \Delta Y = -J^T J \Delta W$$

تا این جا روش گوس نیوتن بود ولی ایرادی دارد که روش LM آمد.

$$\boxed{\Delta W = -(J^T J)^{-1} J^T \Delta Y}$$

← نقطه ای که خطی کردن از آن شروع می شود، یک نقطه تصادفی است.

← بسط روش : ماتریس $J^T J$ معکوس پذیر نیست لزوماً.

PAPCO

Scanned by CamScanner

Subject
Date

روش LM :

$$E = \frac{1}{2}(\Delta y + J\Delta w)^T(\Delta y + J\Delta w) + \frac{\lambda}{2}\Delta w^T \Delta w$$

خطی سازی کردیم ولی با این شرط که نگذاریم خیلی دور شود (با اضافه کردن این ترم)

← باعث می‌شود دقت خطی سازی بالا برود و ماتریس‌ها نیز معکوس‌پذیر اند.

$$\frac{\partial E}{\partial \Delta w} = J^T(\Delta y + J\Delta w) + \lambda \Delta w = 0$$

$$J^T\Delta y + (J^TJ + \lambda I)\Delta w = 0$$

$$\Delta w = (J^TJ + \lambda I)^{-1} J^T \Delta y$$

درصورتی که $\lambda > 0$ باشد.

مقادیر ویژه با λ جمع می‌شود درنتیجه دیگر مقدار ویژه صفر نداریم.

ماتریس J^TJ مقادیر ویژه اش یا صفر است یا بیشتر. (Semi positive)

Semi positive : $\alpha^T_{1\times n} \cdot J^TJ_{n\times n} \cdot \alpha_{n\times 1} \geq 0$

$$(J\alpha)^T J\alpha = \sum \beta_i^2 \geq 0 \quad ✓$$

در روش LM، نکته همان انتخاب λ است که اندازه قدم ها را نشان می‌دهد.

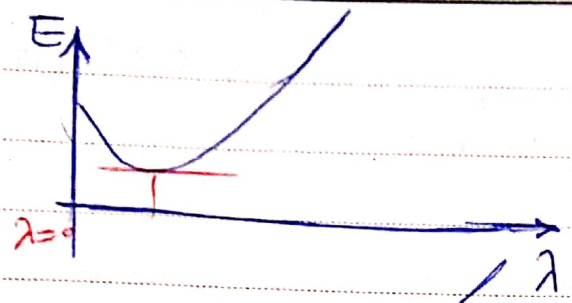
هرچه λ بزرگتر ← اندازه قدم ها کوچکتر

در $\lambda = 0$ ← بزرگترین اندازه قدم ← روش نیوتن

PAPCO

Scanned by CamScanner

Subject
Date



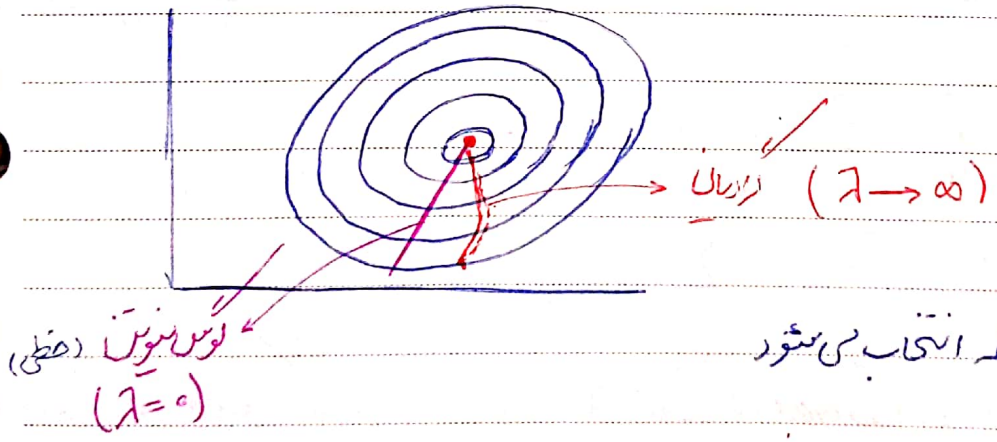
مقدار λ را با validation cross تشخیص می‌دهند.

حرفه الگوریتم زیاد است چون برای λ های مختلف E را حساب کنیم

هرچه λ زیاد شود، اندازه قدم ها کم می‌شود، خطای آموزشی افزایش می‌یابد، خطای validation کم می‌شود ⟸ نقطه ای که این ۲ خطا به هم برسند، بهترین مقدار λ است.

⟵ در روش LM و گوس نیوتن، اندازه قدم با روش گرادیان فرق دارد و ما اصلاً در جهت گرادیان حرکت نداریم.

کانتور ← منحنی هایی که روی منحنی خطا با هم برابر است.



روش LM با توجه به λ ای که انتخاب می‌شود بین این دو روش است.

overfit ⟸ الگوریتم روی داده های آموزشی خوب یاد گرفته، اگر داده جدید بدهیم نمی‌تونه انجام بده

PAPCO

Scanned by CamScanner