(استاد عباد زاده)

Subject:
Date

هوش محاسباتی

«جلسه ۱»

TA هاش از بچه های سال پیش هستن که این درو داشتن

موجود فعلی نمی تونه با محیط تطابق پیدا کنه بلکه نسل های بعدی آن مطابق با آن محیط خواهند بود. مثل درخت

→ از طریق تکامل

انسان هم از طریق تکامل هم از طریق توانایی ساخت وسیله و... می تواند با محیط منطبق شود.

توانایی خود موجود زنده که بتواند شرایط محیط را متناسب با خودش بکند (رأساً)

↓

تنها موجودی که این گونه است: انسان

> تطابق با محیط : عقلانی

> هوش → تعریف : ریاضی

هوش مصنوعی ← تصمیم عاقلانه (عملی رو انتخاب کنه که معیار کارایی را max کنه)

یک معیار کارایی را بهینه کند.

دیرت شده؟ معیار کارایی = زمان ⇐ تصمیم : اسنپ

نشده؟ " = هزینه ⇐ " : مترو

* یک مباحثی از هوش مصنوعی را می گویند هوش محاسباتی : الگوریتم های تکاملی، فازی و شبکه های عصبی
* هوش محاسباتی = هوش مدرن
* هوش سنتی ← الگوریتم های جستجو، منطق، عملگرهای هوشمند

(SPEA ?)

IEEE → انجمن هوش محاسباتی داره. → مهمترین انجمن برق و کامپیوتر دنیا

تکاملی :

تکامل ← بالا رفتن شایستگی فرزندان نسبت به والدین

هدف اصلی تکامل : بقای ژن (نه بقای موجود)

اهداف فرعی : - تنوع ژن (باعث بقای ژن می شه)

تطابق با محیط و به دست آوردن مواد غذایی لازم → - به جای اینکه ویژگی هایی در خودش پیدا کنه تا بچه (زرافه با گراز) بتونه ویژگی هایی پیدا کنه که منبع غذایی اش را عوض کند که نیاز به جنگیدن نباشد.

- تولید مثل

فیزیکی

برگ های پایینی رو می خورد

Scanned by CamScanner

Subject:
Date

* اگر هدف تکامل بهتر شدن بود، دیگه جانور تک سلولی نباید وجود می داشت.
* از نظر تکامل موجودی شایسته تره که بتونه تولید مثل بیشتری داشته باشه. چون تولید مثل باعث انتقال ژن می شه.
* تکامل، طول عمر را بهینه نمی کند. وقتی موجود تولید مثل کرد حتی بهتر است بمیرد تا منابع کمتر مصرف شود!

و فرزندانش را به جایی رساند که بتوانند مستقلاً زندگی کنند

* روش های بهینه سازی ← ۱- تحلیلی و کلاسیک: مشتق = ۰ ⇐ ...

مشکل: تا بیشتر از درجه ۲ نمی تونیم مشتق = ۰ رو حل کنیم

← ۲- روش های جستجو ← ۱ کل فضا را می گردند ⇐ تضمین یافتن جواب بهینه

مشکل: ممکنه فضای جستجو خیلی بزرگ باشه (خصوصاً اگه پیوسته باشه.)

۲ قسمتی از فضا که مطمئن است جواب آنجا نیست را نمی گردند (مثل A^*) ← لازم زیاد وقت می گیره.
↓
تضمینی

۳ جاهایی رو می گردند که امید دارند جواب اون جاست.

عیب: تضمینی نیست

مزیت: فضای جستجوی کوچک
⇓
این روش عملی عه.
↓
الگوریتم های تکاملی جزو این دسته هستند.

نلزوماً بهترین جواب رو نمی ده ولی در زمان قابل قبول، جواب قابل قبولی می ده.

Scanned by CamScanner

Subject:
Date

راه حل و راه حل

* تکامل دنبال بهترین جواب نیست. دنبال جوابیه که قابل قبول باشه

هوش محاسباتی ↔ بهینه کردن معیار کارایی ↔ بهینه سازی ↔ الگوریتم تکاملی

«جلسه ۲»

ژن های مناسب ← ژن هایی که بتوانند به نسل های بعد منتقل شوند (این از طریق تولید مثل ممکنه)

چرا تنوع ژن خوبه؟ چون هرچه مسیرهای بیشتری وجود داشته باشد، شانس اینکه یک ژن منقرض شود ولی یکی بماند، بیشتر می شود.

برای بقای ژن تکامل لازم است. چرا؟ چون منابع غذایی محدود است.
مثلاً یک سلولی تکامل که پیدا کند می تواند از منابع غذایی دیگری هم استفاده کند

چگونه یک موجود زنده می تواند ویژگی های فیزیکی تغییر یافته ی خود را به فرزندانش منتقل کند؟
→ اردید میکروسکوپی → اسلاید
بررسی می کنیم
(خیلی جزئی مهم نیست)

اطلاعات ژنتیکی ← داخل هسته سلول (کروموزوم و DNA داخل هسته)

کروموزوم

تخمک + اسپرم $\xrightarrow{\text{لقاح}}$ زیگوت ⇒ سلول های جنینی → بنیادی

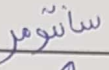
سلول های بالغ

فرایند میتوز (تقسیم سلولی)

Scanned by CamScanner

Subject:
Date

هر جانور یا ... یا ... است.

دیپلوئید ← کروموزوم جفت  سانتومر

هاپلوئید ← " تک.

ژن: خالص / ناخالص (دیپلوئید) — غالب / مغلوب (خودش را نشان نداده)

در مورد مرد کروموزوم جنسی X-Y است. (در مورد زن: XX)
بقیه همه جفت هستند.

نسل اول:

	W	W
R	WR	WR
R	WR	WR

نسل دوم:

	W	R
W	WW	WR
R	WR	RR

زرافه و گراز ← ژن قدبلندی که مغلوب بوده، به دلیل از بین رفتن ژن غالب قدکوتاهی و تولید مثل بیشتر ژن قدکوتاهی، در نسل بعد شده ژن غالب

DNA
↳ خاصیت اسیدی

۴ تا کد ژنتیکی متناظر با ۴ باز شیمیایی است.

A - T
C - G
G - C
T - A
} ۴ حالت پیوند نیتروژنی

یک سمت DNA (نصف نردبان)، کد است (طرف دیگر مکمل آن است)
↳ Codon = هر سه کد پشت سرهم

هر Codon معادل یک آمینو اسید است (۶۴ تا : ۴×۴×۴) ← ولی یک سری شان باهم برابرند.
۲۰ و خورده ای می‌شه.

یک ژن معادل با دنباله ای از آمینو اسیدها است.

(به جای T، U هم می‌گذارند.)
در DNA
در RNA (کپی قسمتی از DNA است.)

PAPCO

Scanned by CamScanner

Subject:
Date

* جهش ← خطا است ← آمینو اسید به خاطر معادل بودن چند مدلش، تا حدودی جلوی جهش را می‌گیرد.

پروتئین ← ویژگی های فیزیکی موجود زنده را تعیین می کند.
← یک تعداد آمینو اسید کنار هم که چون تعدادش ثابت نیست، انواع خیلی زیادی پروتئین داریم

* ۱- پروتئین ها و ۲- شرایط محیطی، می توانند تغییر دهند که یک سلول از کجا برنامه اش را شروع کند و کجا پایان دهد.

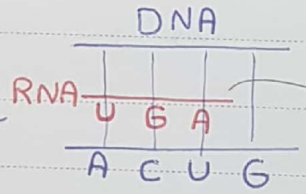
سلول = کارخانه ی پروتئین سازی ← چرا؟

چه جوری RNA از روی DNA کپی برمی دارد؟ از مکمل استفاده می کند.

مکمل که به این صورته:

A → U
C → G
U → A
G → C

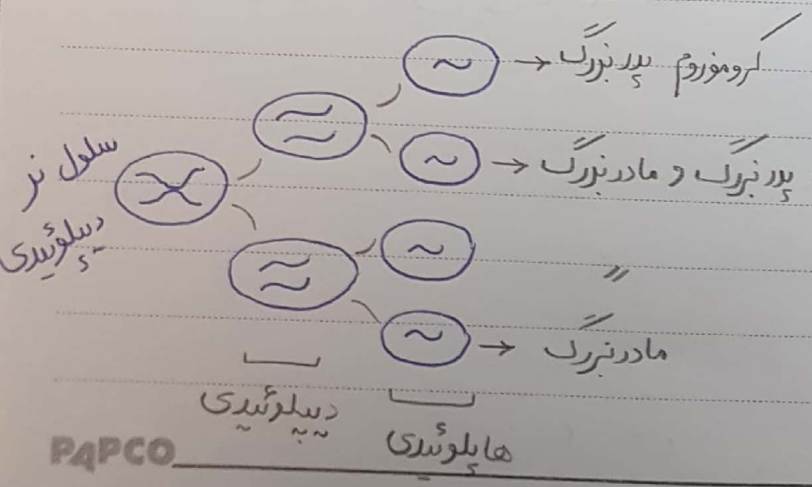
مکمل هر کد می تواند با آن کد match شود و به این ترتیب RNA یک نصف نردبان از قسمتی از DNA را می سازد.



بعد از غشاء خارج می شه چون کوتاهه.

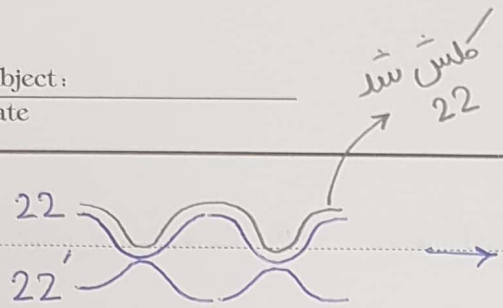
- ژن ها چگونه به فرزندان منتقل می شوند؟

۲۳ نوع جفت کروموزوم داریم

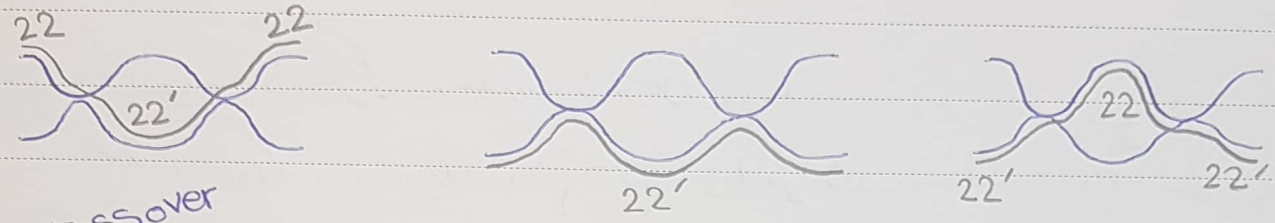


تقسیم سلولی
- معمولی
- جنسی ← بهش می گن میوز

Scanned by CamScanner



در همانند سازی ۴ حالت اتفاق می افتد.

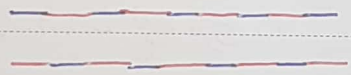


crossover رخ می دهد.

4^{23} حالت برای نر و 4^{23} حالت برای ماده ⟸ $4^{23} \times 4^{23}$ حالت برای فرزند

شانس اینکه فرزند هیچ کدام از کروموزوم هایش ترکیبی نباشد، نزدیک به صفر است!

حالت هاپلوئیدی:



ژن ها در چند نقطه جابه جا می شوند. (؟)

جهش ← در حد کودون
" ژن
" کروموزوم ← مثلاً به جای جفت، سه تایی بشه!

تکامل موفقیتش به خاطر تدریجی بودن است. ← جهش و شرایط محیطی باعث افزایش سرعت تکامل می شود.

Scanned by CamScanner

Subject:
Date

اولش چی گفته؟ «جلسه ۳» در تضاد هستند.

برای اینکه تکامل بتونه اتفاق بیفتد، تنوع و انتخاب طبیعی باید وجود داشته باشد همزمان و هم وزن نگه داشته شوند.

↓ می‌خواد بهترین‌ها رو انتخاب کنه

تنوع به تنهایی ← به خاطر محدودیت منابع نمی‌شه تنوع مدام بیشتر و بیشتر شه.

موجب تکامل نمی‌شه!

متوسط شایستگی را پایین می‌آورد.

یک اسلاید داشت

تنوع انتخاب رو کنترل می‌کنه ← تنوع نمی‌ذاره انتخاب همه‌ی بهترین‌ها رو انتخاب کنه.

انتخاب تنوع " " " ← انتخاب نمی‌ذاره تنوع مدام بیشتر و بیشتر شه از همه جهت. انواع خاصی رو به نسل بعد انتقال می‌ده.

برای همین تکامل تدریجی است. (وگرنه، موفق نیست)

تنوع می‌خواد جمعیت رو واگرا کنه، انتخاب همگرا

تنوع عملکردش از طریق تولیدمثله، انتخاب از طریق انتخاب طبیعت

تنوع کمیت رو افزایش می‌ده و انتخاب کیفیت رو.

اگر یک ژن در پدر غالب باشد، ممکن است در فرزند مغلوب شود یا برعکس

یک کروموزم را معادل با یک موجودی گیریم (چون ویژگی‌های موجود را کروموزم‌ها تعیین می‌کنند)

یک راه حل برای یک مسئله‌ی بهینه‌سازی متناظر می‌توان گرفت با یک روش بازنمایی coding در تکامل

مثلاً ۲۰ شهر داریم. ۱۰ فرودگاه می‌خواهیم بزنیم تا فاصله هر شهر تا نزدیک‌ترین فرودگاه min باشد.

← یک راه حل متناظر با $(x_1, y_1, \dots, x_{10}, y_{10})$ است.

مختصات شهرها $= (x_{c_i}, y_{c_i})$

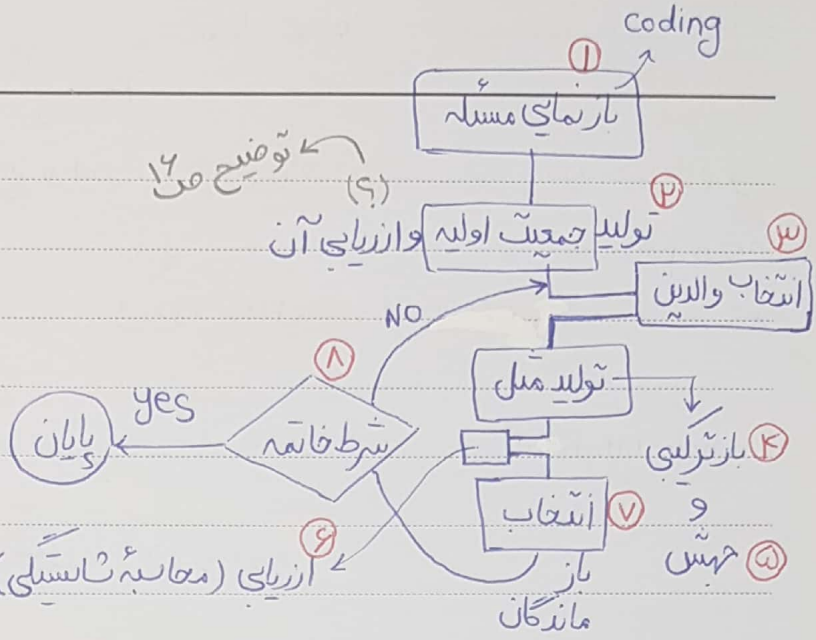
تابع هدف:
$$\sum_{i=1}^{20} \min_{j=1}^{10} \left((x_j - x_{c_i})^2 + (y_j - y_{c_j})^2\right)$$

PAPCO

راه حل ← پارامترهایی از مسئله که اگر تغییر کند، تابع هدف تغییر کند.

Scanned by CamScanner

Subject:
Date



معادل
راه حل ⟷ موجود
تابع هدف ⟷ شایستگی
راه حل های اولیه ⟷ جمعیت اولیه

④ و ⑤ باهم باعث تولید مثل و تولید فرزندان می شه (عملگرهای تنوع) هستند

انتخاب والدین ← چه موجوداتی تولید مثل کنند
انتخاب بازماندگان ← چه موجوداتی برونده به نسل بعد
} عملگرهای انتخاب هستند

- حل مسئله‌ی ۴ وزیر با الگوریتم تکاملی:

* شایستگی ۰ و ۱ ای نیست
* جمعیت اولیه معمولاً به صورت تصادفی است.

$0+2+2+2 \Rightarrow 1/6$

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

مکان وزیر ۴
مکان وزیر ۳
مثل A و T و G و C

* تابع شایستگی‌ای بهتر است که برای هر دو موجود بتواند تمایز قائل شود و هیچ دو موجود متفاوتی تابع شایستگی یکسان نداشته باشند تا بتوان ارجحیت داد یکی را بر دیگری.

* تابع شایستگی برای مثال بالا ← تعداد وزیرهایی که هر وزیر را تهدید می کنند ← مجموعش / معکوسش

* معمولاً فقط یا انتخاب والدین یا انتخاب بازماندگان را با شایستگی درگیر می کنند.

PAPCO

Scanned by CamScanner

Subject:
Date

بهترین حالت

انتخاب به این صورت است: $\frac{1}{6}$ $\frac{1}{10}$ $\frac{1}{4}$ $\frac{1}{2}$

جهش ← به طور تصادفی جای یکی از 1 ها را عوض کنیم.
در این مسئله

بازترکیبی ← cross over

* معمولاً جمعیت را در نسل بعد با قبلی اندازه اش را یکی می کنند.

دو اصطلاح اکتشاف و استخراج در الگوریتم های جستجو:

جستجوی عمومی (محل وجود جواب رو مشخص می کنیم) ← جستجو رو دقیق می کنیم در اون محل

* مثل استخراج طلا می مونه.
* باید تعادل ایجاد کرد بین اکتشاف و استخراج. اگه تمام عمرمون رو بذاریم روی پیدا کردن محل جستجو، همش ضرره. ← اینکه یجا دست بکشیم از اکتشاف و بریم سراغ استخراج از همه بهتره؟ ← مسئله ی باز است.

« جلسه ۴ »

قسمت ⑥ در ص بعد ← برای اینکه ببینیم کدام موجودات به نسل بعد منتقل شوند.

① بازنمایی مسئله: ۱- راه حل ها باید کل فضای ممکن را نشان دهد. روش کد کردن باید بتواند تمام حالات ممکن را که پتانسیل جواب بودن را دارند، پوشش دهد.

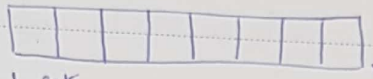
تمام راه حل های موجود ⊆ فضای جستجوی روش coding

PAPCO

Scanned by CamScanner

* امّا هر چه فضا کوچکتر باشد، سرعت و دقت پیدا کردن جواب بالا می رود. پس فضای جستجوی روش coding ما خیلی هم نباید بزرگتر از مجموعه راه حل های ممکن باشد.

برای مثال: در مسئله‌ی ۸ وزیر:



$8^8 \leftarrow 64$

$10^{15} \leftarrow 2^{48}$

۱-۶۴

هر وزیر در یکی از ۶۴ خانه قرار می گیرد.

در نظر گرفتن ۸ وزن (ساده‌ترین حالت)

$10^{15} \times 10^{-9} = 10^{6}\,s$

± نانوثانیه طول بکشد کنترل هر حالت

۱۰ روز

امّا اگر این طور در نظر بگیریم که هیچ دو وزیری در یک ستون قرار نمی گیرند ⇐ فضای حالت = 8^8 ← در ۰٫۱ نانوثانیه می توان کل فضا را گشت.

در یک سطر و ستون نمی توانند ... ⇐ $8!$ ← حالت ایده آل

۲- تا حد ممکن فضای جستجو کوچک شود.

۳- coding طوری باشد که بتوان بازترکیبی و جهش را بهتر تعریف کرد.

۴- " " " " " تابع شایستگی " " " ← شایستگی دقیق مهم نیست ولی نباید بین هر دو موجودی اولویت بندی کند.

به عبارتی: در تکامل به مقدار شایستگی وابسته نیستیم ← بین هر دو موجودی اولویت بندی کند.

- روش‌های بازنمایی از دیدگاه ریاضی:

۱- باینری: مثلاً در مسئله‌ی فرودگاه: 2^{21} (m) $= 2 \times 10^{6} = 2 \times 10^{3} \times 10^{3}$ ← هر فرودگاه یک مختصات x و y و هر x یا y عددی بین ۰ تا ۲۰۰۰ km

هر ژن: ۲۱ بیت و ۲۰ تا ژن ← $\underbrace{10 \times 2}_{x,y} = 20$ ← یک کروموزوم

باینری

مشکل: باعث مشکل بایاس در بازترکیبی و جهش می شود ⇐ استفاده از grey code به جای

PAPCO

Scanned by CamScanner

Subject اگر بیت با ارزش باشد، تغییرات بسیار زیاد خواهد بود.
Date →

روش

در باینری نمی توان گفت بازترکیبی وجهش دارند جستجوی عمومی انجام می دهند یا محلّی
چون بیتی که تغییر می کند برای تمام ۰ و ۱ ها با ارزش های مکانی متفاوت یکسان است.
احتمال

امّا در gray code می توان گفت بازترکیبی جستجوی عمومی انجام می دهد و جهش جستجوی محلّی

۲- gray code : ؟ ← الگوریتم آن به این صورت است که هر عدد نسبت به عدد قبلی تنها ۱ بیت، صفر، یک شده یا برعکس. به این صورت :

۰	۰۰۰
۱	۰۰۱
۲	۰۱۱
۳	۰۱۰
۴	۱۱۰
۵	۱۱۱
۶	۱۰۱
۷	۱۰۰

هرچه صفر و یک عوض شود، ۹

۳- اعداد حقیقی : ؟ (از این جا به بعد ← voice)

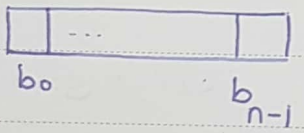
فرق با باینری در cross over : موجود جدید نمی تونه تولید شه.
مثلاً رنگ چشم جدید تولید نمی شه! ← از وسط ژن ها نمی شه ترکیب کرد.
فقط جابه جایی ژن ها را داریم!

استراتژی تکاملی

ES ← ۳ GA ← ۱ و ۲

* فرمول محاسبه ی عدد در binary :

$$\sum_{i=0}^{n-1} b_i 2^i$$



GP ←

۵- درخت : جواب پیدا کردن بهترین مدل است.

۶- گراف

EP ←

۴- جایگشت :

* بازترکیبی استاندارد فقط در روش binary می تواند پیاده سازی شود.

PAPCO

Scanned by CamScanner

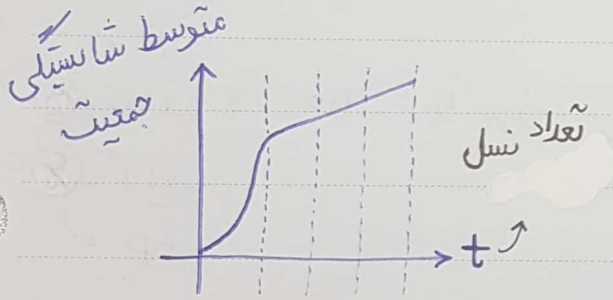
* الگوریتم های تکاملی، وابسته به مسئله نیستند.
به دو روش binary و gray code ، GA (الگوریتم ژنتیک) گفته می شود.

«جلسه ۵»

* جایگشت از جمله روش هایی که فضای جستجو رو می تواند کوچک کند

② تولید جمعیت اولیه : ← در الگوریتم های تکاملی تعداد جمعیت ثابت است : μ (تعداد بازماندگان)
λ : تعداد فرزندان که در هر حلقه‌ی تکامل تولید می شود.

منحنی سرعت همگرایی

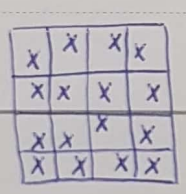


توزیع تصادفی یکنواخت

- جمعیت اولیه باید متنوع انتخاب شود نه بایاس شده
چون به هر حال طبق نمودار مقابل، الگوریتم تکاملی سرعت همگرایی اش در ابتدا بسیار بالاست پس ارزش ندارد بخواهیم این زمان را save کنیم و ریسک بایاس انتخاب کردن جمعیت را بپذیریم (که در min محلی گیر کنیم)

- توزیع یکنواخت چیست؟

کل فضا را تقسیم کنیم به تعداد μ قسمت مساوی و بعد هر موجود را به طور تصادفی از یکی از قسمت ها انتخاب می کنیم.



مثلاً $\mu=16$ ←

Subject
Date

با احتمال ۷۰٪ فردا باران می‌آید یعنی: اگر شرایط آب و هوای فردا ۱۰۰۰ دفعه تکرار شود، در ۷۰۰ بار آن، باران خواهد آمد.

* آمار و احتمال با تکرار ممکن می‌شود (و تعداد داده‌های زیاد)

* تعداد دفعاتی که تابع شایستگی را محاسبه می‌کنیم چقدر است؟

مرحله ۶ ص ۸ — مرحله ۲ ص ۸

$\lambda N + \mu$

N: تعداد نسل

چرا فقط بار اول ارزیابی انتخاب والدین رو حساب می‌کنه؟

توضیح ص ۱۶

مثلاً در عدد λ باهم تفاوت دارند.

* برای مقایسه‌ی دو الگوریتم تکامل باید منحنی‌ها را برحسب تعداد ارزیابی - متوسط شایستگی جمعیت رسم کرد - چون سرعت الگوریتم تکاملی برحسب تعداد ارزیابی‌ها مشخص می‌شود.

دفعات

۳ انتخاب

والدین ← از جمعیت μ انتخاب می‌شوند ← روش‌های چندجمعیتی (چندهمسایگی) ← فقط از نخش‌های خاصی از μ انتخاب می‌کند (زیرجمعیت‌ها)

محل انتخاب: ← بازماندگان

$\mu + \lambda$

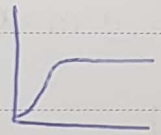
μ و λ

بازماندگان فقط از λ انتخاب می‌شوند.

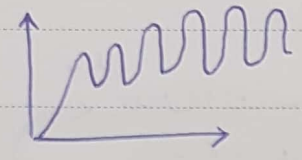
حتماً باید $\lambda > \mu$ باشد

بازماندگان از هر دوی μ و λ انتخاب می‌شوند

اگر به شایستگی خیلی بیشتر از تنوع اهمیت داده شود:



و اگر به تنوع خیلی بیشتر از شایستگی اهمیت داده شود:



• EA(1+1) → فقط یک موجود داریم ← بازترکیبی نمی‌تونه بکنه. اما جهش پیدا می‌کنه

اگه بهتر شد شایستگی، جهش یافته می‌ره مرحله بعد وگرنه جهش نیافته همی‌ره مرحله بعد

Evolutionary Algorithm

مانند الگوریتم تپه نوردی است (Hill Climbing)

Scanned by CamScanner

← مجبوریم فرزند جهش یافته را ببریم به نسل بعد
(انتخاب دیگری نداریم!)

- EA(1و1) → random walk است. مانند الگوریتم
الگوریتم تکاملی نیست! بلکه تصادفی است.

- EA(1+λ) (λ>1) → یک موجود داریم. با جهش λ تا فرزند تولید می‌کنیم. از بین این 1+λ تا موجود یکی را به نسل بعد می‌بریم. (مانند تپه نوردی)

- EA(1,λ) (λ>1) → از روی 1 موجود λ تا فرزند تولید و یکی از λ به نسل بعد

- EA(μ+1) → یک فرزند تولید می‌کنیم و از بین μ+1 ای انتخاب می‌کنیم برای نسل بعد. μ تا موجود داریم.
  - اگر اونی که حذف می‌شه بدترین باشه: Genito
  - Steady state: اگر اون یکی که حذف می‌شه از بین فرزندان باشه (؟)

- ~~EA(μ و 1)~~ → چون λ اش از μ بزرگتر نیست پس امکان پذیر نیست.

- EA(μ+μ) → μ تا موجود داشته باشیم. μ تا فرزند داشته باشیم. از بین این 2μ، μ تا برن به مرحله بعد.

- EA(μ و μ) → Genetic algorithm
μ تا موجود داریم. μ تا فرزند. همه‌ی فرزندان رو می‌بریم به نسل بعد و والدین می‌میرند.

* در ES از EA(μ+λ) و EA(μ و λ) با شرایط $\lambda \approx 7\mu$ استفاده می‌شود.
با تجربه ثابت شده در این صورت سرعت max می‌شود.

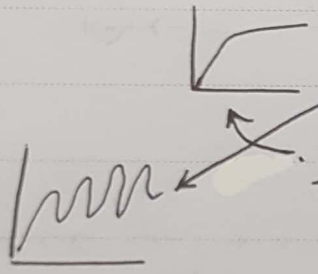
* ES های خاص
  - ES(μ+λ)
  - ES(1+1)
  - ES(1+λ)

باید روش انتخاب بر اساس شایستگی باشد.

μ و λ ← فراموش کار است. ← تنوع را زیاد می‌کند!

μ+λ ← حافظه دار است. ← سرعت همگرایی را بالا می‌برد.

باید در کنار روشی باشد که تنوع داشته باشد.



از انتخاب

تعادل برقرار بماند.

Scanned by CamScanner

Subject
Date

احتمال بهترین موجود — تعداد موجوداتی که می‌خواهیم انتخاب کنیم

- فشار انتخاب (PS) : $E[n_b] = SP = N_2 * P_b$ → best

احتمال انتخاب هر موجود : P_i (چه برای انتخاب والدین چه بازماندگان)

① در انتخاب والدین : $N_2 = \lambda$ $N_1 = \mu$ (تعداد موجوداتی که داریم. همون که دوص قبل در موردش صحبت می‌کردیم)

② در انتخاب بازماندگان : $N_2 = \mu$ $N_1 = \mu + \lambda$
با روش $\mu + \lambda$

③ " " " μ و λ : $N_2 = \mu$ $N_1 = \lambda$

* $\sum_{i \in N_1} P_i = 1$ همواره باید :

$P_i = \frac{f_i}{\sum_{j=1}^{N_1} f_j}$ → برای وقتی که P_i ها بر اساس و متناسب با شایستگی تعریف شوند.
! (گاهی فقط یک ربطی به شایستگی دارد.) !

$E[n_i] = N_2 \cdot P_i$

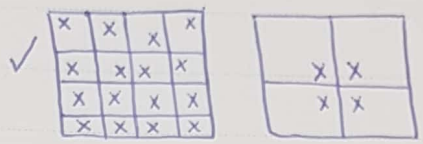
امید ریاضی

اگر فشار انتخاب کوچکتر از ۱ باشد ← یعنی داریم بهترین موجودمان را فراموش می‌کنیم.
" " " بزرگتر از ۱ باشد ← سرعت همگرایی خیلی بالا. فقط از روی بهترین copy می‌کند

PAPCO

Scanned by CamScanner

* بازترکیبی و جهش باید این ویژگی را داشته باشند که گراف یا جایگشت را خراب نکنند.
* در نمودار ص ۱۴ ← بیشترین شیب در ۱/۴ اول نمودار است. و از آن به بعد برای بهتر شدن جواب ها به زمان زیادی نیاز خواهیم داشت.
* توزیع یکنواخت هرچه تکرار بیشتر باشد، بهتر خواهد شد.



۱- ←

۲- ← هرچه تعداد دفعات بیشتری تاس بیندازیم، دقیق تر می توان گفت در ۱/۶ موارد عدد ۴ ظاهر می شود.

* توضیح ص ۸: برای تکامل باید "انتخاب" داشته باشیم امّا نه هر انتخابی! انتخابی که مبتنی بر شایستگی باشد. پس در الگوریتم تکامل، قبل از هر انتخاب (والدین و بازماندگان) باید یک بار ارزیابی صورت بگیرد.

* در $N8+\mu$ ، μ در برابر $N8$ قابل صرف نظر است.
* معمولاً μ بین ۱۰ و ۱۰۰ می گیرند.

* توضیح ص ۱۳: چون آن چیزی که در یک الگوریتم تکامل در هر بار اجرا وقت گیر است، محاسبه ی شایستگی تمام جمعیتی است که بعد از روی آن انتخاب را انجام می دهیم.

* مقایسه از آن لحاظ که بفهمیم کدام سرعت همگرایی بیشتری دارد.
الگوریتم تکاملی

* وقتی یک موجود جهش پیدا کند، هم خودش و هم جهش یافته اش به عنوان ۲ موجود جدا در جمعیت خواهند بود! ← چرا؟

Scanned by CamScanner

Subject
Date

* مزیت حافظه دار بودن این است که بهترین ها را فراموش نمی کند اما در عوض ممکن است در min محلی گیر کنیم.

«جلسه ۶»

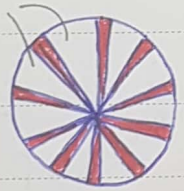
f_i ← شایستگی موجود i ام

P_i ← احتمال انتخاب موجود i ام

فرض: $P_i \propto f_i \Longrightarrow P_i = \frac{f_i}{\sum_{j=1}^{N_1} f_j}$

از بین N_1 موجود، N_2 موجود انتخاب می‌شوند. (یعنی اونی که انتخاب کردیم دوباره می گذاریمش داخل جمعیت و می‌توانیم دوباره انتخابش کنیم.) با جایگزینی

مثال: $N_1 = 3$, $N_2 = 2$, $f_1 = 5$, $f_2 = 3$, $f_3 = 2$ ⟸ $P_1 = 0.5$, $P_2 = 0.3$, $P_3 = 0.2$



اکثراً خالیه؟

(هرچی پول بیشتر بذاری احتمالش بیشتره برنده شی.)

۲ روش انتخاب ← RW (Rooled wheel) ؟ roulette

SUS: Spotastic, Universal Selection, stochastic

$E[n_i] = N_2 \cdot P_i$ → احتمالاً این تعداد از موجوداتی که انتخاب کرده‌ایم، n_i هستند

if $N_2 = 2$:

$E[n_1] = 2 \times 0.5 = 1$

$E[n_2] = 2 \times 0.3 = 0.6$ ⇝ [0,1] یا یکی یا صفر تا

$E[n_3] = 2 \times 0.2 = 0.4$ ⇝ [0,1] یا یکی یا صفر تا

PAPCO

Scanned by CamScanner

Subject
Date

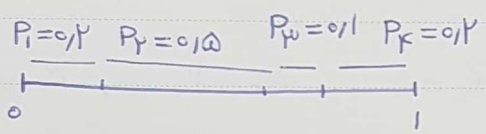
if $N_2 = 10$:

$E[n_1] = 10 \times 0.5 = 5$
$E[n_2] = 10 \times 0.3 = 3$
$E[n_3] = 10 \times 0.2 = 2$

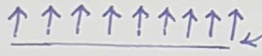
→ با rooled wheel این اتفاق نمی‌افتد چون تعداد تکرار (10) کم است و تصادفی یکنواخت رخ نمی‌دهد.

از نظر ریاضی فرمول $E[n_i] = N_2 \cdot P_i$ در هر دوی SUS و Roded wheel باید برقرار باشد
اما فرق SUS این است که در N_2 های کوچک، یکنواخت بودن را force می‌کند

یک خط‌کش از ۰ تا ۱ داریم (چون مجموع P_i های موجودات تمام ۱ است.) اینکه P_i ها را روی (توزیع شده) خط‌کش چه‌طور بچینیم، باید تصادفی باشد نه مرتب شده. (حتی می‌توان یک P_i بزرگ را تکه تکه کرد)
خط‌کش را به N_2 قسمت (در اینجا ۱۰ قسمت) تقسیم می‌کنیم. در روش SUS، از هر قسمت، یک نقطه را به طور تصادفی انتخاب می‌کنیم.



حالا یک خط‌کش دیگر به سایز $1 - \frac{1}{N_2}$ (در اینجا $\frac{1}{10}$) بری داریم هر $\frac{1}{N_2}$ تا، روش علامت می‌زنیم.



یک شروع تصادفی یکنواخت بین ۰ تا $\frac{1}{N_2}$ ← (مثلاً بین ۰ و ۱، ۰/۰۵)
از این نقطه‌ی شروع خط‌کش دوم را ایجاد می‌کنیم.
روی خط‌کش اول قرار می‌دهیم و
علامت‌ها روی بازه‌ی احتمال هر موجودی که بیفتد، یعنی آن موجود انتخاب می‌شود.

- اسلاید به نام روش‌های انتخاب:

$$\text{بایاس} = \left| P_i - \frac{f_i}{\sum_{j=1}^{N_1} f_j} \right|$$

بایاس بیان می‌کند که روش انتخابی ما چقدر با روش بیان احتمال‌ها بر اساس شایستگی موجودات تفاوت دارد. (طبق ! در ص ۱۵)

Scanned by CamScanner

Subject
Date

فشار انتخاب

spread = selective pressure ⟸ $n_i = n_b$ در حالت خاص اگر $E[n_i]$ ← spread (انتشار)

از دست رفتن تنوع ← $\frac{N_1 - N_2}{N_1}$: موجوداتی که انتخاب نشده اند

درصد

امید ریاضی متوسط شایستگی جمعیت بعد از انتخاب

سه معیار مهم برای سنجش میزان خوب بودن روش انتخاب

فشار انتخاب

Selective intensity : سرعت همگرایی را نشان می دهد

Selection variance : تنوع را نشان می دهد

مثلاً ۱۰۰۰ تا

اول جمعیتی با توزیع شایستگی نرمال تولید می کنیم (متوسط = ۰ واریانس = ۱)

موجودات را انتخاب می کنیم با یک روش

متوسط شایستگی ها را حساب می کنیم

۳۰ بار این کار را انجام داده و متوسط متوسط ها را می گیریم ⟸ Selection intensity را به ما می دهد

(بین ۱-و۱ ← حدود ۶۶٪
بین ۲- و ۲ ← حدود ۹۹٪)

این یعنی توزیع گوسی

* چطور از روی توزیع یکنواخت، توزیع نرمال تولید کنیم؟

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

max : $x=0$ → $\frac{1}{\sqrt{2\pi}}$ → $\frac{1}{\sqrt{2\pi}}$

یک عدد تصادفی یکنواخت بین ۰ تا $\frac{1}{\sqrt{2\pi}}$ ایجاد می کنیم (α) سپس:

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \alpha \rightarrow e^{-x^2/2} = \sqrt{2\pi}\,\alpha$$

$$-x^2/2 = \ln \sqrt{2\pi}\,\alpha$$

$$x = \sqrt{-2\ln\sqrt{2\pi}\alpha} \leftarrow x^2 = -2\ln\sqrt{2\pi}\alpha$$

← شیر و خط می اندازیم که یکبار + و یکبار - بدهد ...

x به دست می آید

PAPCO

Scanned by CamScanner

sel variance ← ماتد sel intensity به جای متوسط شایستگی ها ، واریانس شایستگی؟
را می گیریم. واریانس = $\frac{1}{N}\sum (f_i - \bar{f})^2$

هرچه امید ریاضی واریانس بعد از انتخاب بیشتر باشد، یعنی تنوّع بیشتر بوده است.

اگر شایستگی یک موجود بسیار بیشتر از سایرین باشد، بعد از چند نسل، تمام موجودات از آن نوع خواهند شد.
همگرایی زود رس

مثلاً $P_b = 0.9$ و $\sum_{i \neq b} P_i = 0.1$

① فشار انتخاب بسیار زیاد است.

اما آن موجود موجودی نیست که هدف تکامل است: (موجود دلخواه)

④ $f_{rb} \gg f_b \gg f_i$
↳ Real Best

② طبق روش مبتنی بر شایستگی

جمعیت ساکن ← اگر عدد شایستگی موجودات مختلف بسیار شبیه به هم باشد، با گذشت نسل ها در روش مبتنی بر شایستگی، اتفاق جدیدی رخ نمی دهد.

در حالتی که از مشکلات روش های مبتنی بر شایستگی هستند.

PAPCO

Scanned by CamScanner

Subject
Date

«جلسه ۷»

* روش های انتخاب ؟

۱- همه انتخاب شوند μ و λ } مبتنی بر شایستگی ... ؟
۲- تصادفی یکنواخت

۳- انتخاب برشی ← sort موجودات از بهترین تا بدترین و N_2 تا موجود برتر را بری داریم

۴- متناسب با شایستگی ← $P_i = \frac{f_i}{\sum_{j=1}^{N_1} f_j}$ ← SUS و RW

$f_i = \frac{1}{c_i}$ (f بر اساس شایستگی باید باشد حتما) ← هزینه

متناسب با شایستگی هست

۵ و ۶ که در صفحه بعد

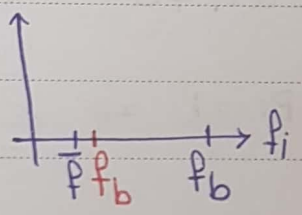
* در جمعیت ساکن، انتشار برابر با ۱ است.

متوسط تعداد copy های که از یک موجود انتخاب می شود (چه در انتخاب والدین چه در انتخاب بازماندگان) ← انتشار

* در روش انتخاب برشی، حتماً باید $N_1 > N_2$ باشد چون copy های موجود را نمی شود باز هم انتخاب کرد.

* جمعیت ساکن و همگرایی زود رس چه زمانی اتفاق می افتد؟ ← سه شمارهٔ صفحه قبل برای همگرایی زود رس بلی
متناسب با شایستگی رو یادت نره

① متناسب با شایستگی ② انتشار = ۱

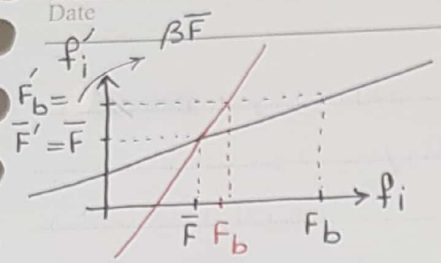


شایستگی متوسط با شایستگی best بسیار تفاوت دارد ← همگرایی زود رس
" " بسیار نزدیک به " " است ← جمعیت ساکن

PAPCO

Scanned by CamScanner

Subject
Date



$F'_i = aF_i + b$

اگر $F_i > F_j \Rightarrow F'_i > F'_j \Rightarrow$ مکان شایستگی در شایستگی واقعی و مجازی عوض نمی‌شود.

(قبلاً هم گفته بودیم مقدار شایستگی مهم نیست.)

یک شایستگی مجازی ایجاد می‌کنیم تا مشکل همگرایی زودرس یا جمعیت ساکن را حل کنیم.

روش اول حل:

$$\bar{F}'=\bar{F} \Rightarrow \frac{\sum f_i}{N_1} = \frac{\sum (af_i+b)}{N_1}$$

$$F'_b=\beta\bar{F} \Rightarrow af_b+b=\beta\bar{F}$$

روش حل دوم:

$$y-y_0=\frac{y_1-y_0}{x_1-x_0}(x-x_0) \Rightarrow F'-\bar{F}=\frac{\beta\bar{F}-\bar{F}}{F_b-\bar{F}}(F-\bar{F})$$

$$\Rightarrow F'_i=\bar{F}+\frac{(\beta-1)\bar{F}}{F_b-\bar{F}}(F_i-\bar{F}) \Rightarrow \boxed{a=\frac{(\beta-1)\bar{F}}{F_b-\bar{F}}} \text{ و } \boxed{b=\bar{F}\left(1-\frac{(\beta-1)\bar{F}}{F_b-\bar{F}}\right)}$$

$\beta \leftarrow$ خودمان تعیین می‌کنیم. که چقدر F_b می‌خواهیم از F متوسط فاصله داشته باشد.

با کنترل F_b

۵ - متناسب با شایستگی مقیاس شده‌ی خطی ✓ ← به صورت بالا

۵' - " " " " " " با فشار انتخاب ← $P_i = af_i + b$

$$\sum_{i=1}^{N_1} P_i = 1 \Rightarrow \sum_{i=1}^{N_1}(aF_i+b)=1 \Rightarrow a\frac{\sum_{i=1}^{N_1}F_i}{N_1}+\frac{bN_1}{N_1}=\frac{1}{N_1} \Rightarrow \boxed{a\bar{F}+b=\frac{1}{N}}$$

$$SP = N_2 P_b = N_2(aF_b+b) \xrightarrow{\div N_2} \boxed{af_b+b=\frac{SP}{N_2}}$$

ادامه: ص بعد

* فشار انتخاب بیشتر از ۲ ← با روش های خطی نمی‌شود.

* فشار انتخاب بین ۱ و ۲ ← " " " می‌شود.

Scanned by CamScanner

Subject
Date

$$b = \frac{1}{N_1} - a\bar{F} \Rightarrow P_i = a f_i + \frac{1}{N_1} - a\bar{F} \Rightarrow P_i = a(f_i - \bar{F}) + \frac{1}{N_1}$$

$$a(f_b - \bar{F}) = \frac{SP}{N_2} - \frac{1}{N_1} \Rightarrow a = \frac{SP/N_2 - 1/N_1}{f_b - \bar{F}}$$

$$\Rightarrow \boxed{P_i = \frac{SP/N_2 - 1/N_1}{f_b - \bar{F}}(f_i - \bar{F}) + \frac{1}{N_1}}$$

۶- متناسب با شایستگی رتبه بندی شده :

برحسب کوچک به بزرگ شایستگی ها را مرتب می کنیم و به ترتیب از ۱ تا N_1 به آن ها مقدار می دهیم. این مقادیر جدید را شایستگی موجودات می نامیم و در آن نه جمعیت ساکن رخ می دهد نه همگرایی زود رس.

$$SP = N_2(a f_b + b) = (a N_1 + b) N_2 \Rightarrow a N_1 + b = SP/N_2$$

ادامه در photo

۶' - متناسب با شایستگی رتبه بندی شده برحسب فشار انتخاب : ؟

در ۶ و ۶' لازم نیست شایستگی ها را به طور دقیق داشته باشیم. (نسبت به ۵ و ۵')
عیب ۶ و ۶' ← هزینه ی Sort کردن ممکنه زیاد باشه.

۷- tornument : Q (حتی عیب Sort کردن ۶ و ۶' را هم ندارد.)

به طور تصادفی یکنواخت Q تا از جمعیت بر می دارد. ← بعد بهترین شان را انتخاب می کند.
(Q معمولاً بین ۲ تا ۵ است ⇐ پیدا کردن max هزینه خیلی کمی دارد.)

N_2 بار این کار را انجام می دهد.

Scanned by CamScanner

Subject
Date

خلاصهٔ این جلسه ← photo

۲ وزن ← موجود انتخاب شده می‌تواند بازگردد به جمعیت یا بازنگردد (با یا بدون جایگذاری)

+ جزوه مهسان photo

«جلسه ۸»

$Q \uparrow$ ← تنوع کمتر سرعت همگرایی بیشتر

$Q=1$ ← الگوریتم تصادفی

$Q=2$ ← binary turnament ⇐ بیشترین تنوع فشار انتخاب (با جایگذاری) = ۲

بدون جایگذاری ← از هر موجود حداکثر ۱ ای می‌شه برداشت.

در حالت با جایگذاری: Q = فشار انتخاب ؟

$Q=N_1$ با جایگذاری ← برشی

〃 بدون 〃 ← یک جورهایی مثل روش متناسب با شایستگی

* این جا یه سری اسلاید داره

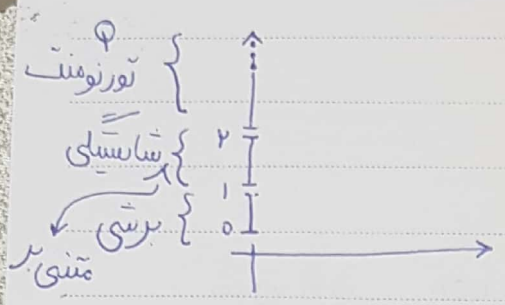
اسلاید: انتخاب - مبتنی بر رتبه - رتبه بندی خطی (s و μ) ← $N_1=N_2$ درنظر گرفته درحالی که نیاز نبوده.

Scanned by CamScanner

Subject:
Date

* منحنی selective variance و selection intensity اسلاید

صعودی و نزولی (تنوع)



نمودار داخل اسلاید

tornument - ♀ نسبت به روش های مبتنی بر شایستگی به ازای یک selection intensity یکسان، تنوع بیشتری ایجاد می کند

با سرعت همگرایی های یکسان

⇐ روش tournament - ♀ بهتر عمل می کند.

اگر برای روش خطی فشار انتخاب بالاتر از ۲ باشد، خط آن به قسمت منفی تجاوز می کند ⇐ ✗

اگر ؟ روش را می خواهید، رتبه‌بندی را انتخاب کنید

و اگر ؟ " " ♀ تورنمنت " " "

* مثلاً در $\mu + \lambda$ می شه ♀ تورنمنت را استفاده کرد که تنوع را بالا ببره و تعادل ایجاد کند!

- شرط های خاتمه:

① اگر شرط خاتمه را بگذاریم میزان شایستگی ای که می خواهیم به آن برسیم، ممکن است هیچ وقت الگوریتم به آن شایستگی نرسد و در loop بی نهایت بیفتد.

↳ نامحدود

② اصلاً ممکن است مقدار بهینه ی تابع هدف مان را ندانیم و ندانیم بهترین شایستگی چیست مثلاً در شطرنج، نقاشی،

↳ ممکن است بهترین مطلق نداشته باشد و همواره بتواند بهتر و بهتر شود.

PAPCO

Scanned by CamScanner

همواره در کنار ⇐ فقط این روش این مزیت را دارد.
یک روش دیگر باید بیاید. (OR شود)

• اگر شرط خاتمه = تعداد ارزیابی‌ها گذاریم، ← بهترین مزیت: مطمئنیم یک زمانی تمام می‌شود و حتی زمان خارج شدن از الگوریتم را می‌توان مشخص کرد.
↓
اول زمانی را که می‌توانیم برای الگوریتم منتظر بمانیم مشخص می‌کنیم بعد تعداد ارزیابی را در آوریم...

↓ عیب

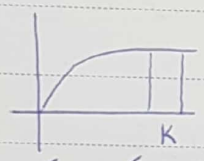
تعیین تعداد iteration های مناسب کار سختی است.

* اگر بتوان بهترین شایستگی را تعیین کرد، با روش بالا OR می‌کنیم.

• هرجا الگوریتم همگرا شود به یک جواب ⇐ جواب دیگر بهتر از آن نمی‌شود ⇐ ...

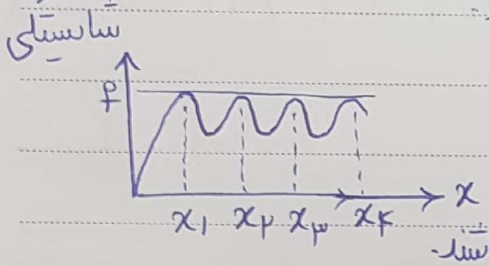
• هرجا دیگر تنوع از بین برود ⇐ تکامل دیگر رخ نخواهد داد. ⇐ فایده ندارد ادامه دهیم.

← در K تا iteration متوسط واریانس شایستگی را حساب می‌کنیم...



اگر از یک حد آستانه کمتر شده بود، می‌گوییم همگرا شده.

* تنوع < ژنوتیپ ← جمعیت، ویژگی‌های متفاوت داشته اند.
فنوتیپ ← مقدارهای شایستگی جمعیت متفاوت داشته اند.



⇐ شایستگی یکسان به معنای از بین رفتن تنوع نیست.
ممکن است با ویژگی‌های مختلف به شایستگی یکسان رسیده باشند.

* در الگوریتم‌های ژنتیک استاندارد، ژنوتیپ = فنوتیپ ← چون تمایل ژنتیکی؟

* در مورد تنوع، واریانس ژنوتیپ یا فنوتیپ اگه یکی باشد، محاسبه می‌کنیم اگر از حد آستانه کمتر بود، یعنی تنوع از بین رفته است.

برای هر ژن حساب کنیم و متوسط بگیریم و در K نسل باز متوسط می‌گیریم...

Scanned by CamScanner

Subject:
Date

در کارهای تحقیقاتی نه کاربردی ، فقط شرط خاتمه = تعداد ارزیابی

Scanned by CamScanner

«جلسه ۹»

شناسنامه‌ی یک الگوریتم تکاملی:

نام الگوریتم		تپه‌نوردی (برای توابع پیوسته)	ژنتیک (GA)
بازنمایی	μ = تعداد جمعیت	اعداد حقیقی $\mu=1$	باینری - گری کد
جمعیت اولیه		تصادفی یکنواخت	تصادفی یکنواخت $B \to x \to C(x) \to F(x) = 1/C(x)$
انتخاب والدین	λ = تعداد فرزندان	$\lambda=1$ همگی انتخاب می‌شوند -	$\lambda=\mu$ متناسب با شایستگی $F_i \to P_i \to RW$ و SUS
بازترکیبی	P_c = احتمال بازترکیبی	نداریم	$P_c>80$ چند نقطه‌ای - یکنواخت
جهش	P_m = احتمال جهش	$P_m=1$ $x_i^{new} = x_i^{old} + \lambda$ $\sigma N_i(0,1)$ اندازه گام	$\frac{1}{n} < P_m < \frac{1}{\mu}$ معکوس کردن
انتخاب بازماندگان	$\mu+\lambda$ μ و λ	1+1 برشی	μ و μ
شرط خاتمه		تعداد ارزیابی $T=10000$	
ویژگی‌های خاص		ایده تپه‌نوردی	تاکید روی بازترکیبی

* چرا σ اندازه قدم است؟ ← اگر 1000 تا قدم برداریم، به طور متوسط اندازه قدم‌هایمان σ است.

سیگما

$$\frac{\sum_{i=1}^{1000} |x_i^{new} - x_i^{old}|}{1000} \overset{?}{=} \sigma$$

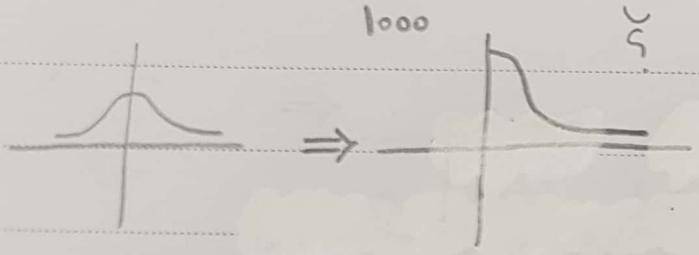
فرمول متوسط: $E[x] = \int x P(x)\, dx$

$E[|x|] = \dots$

$$= \sigma \frac{\sum |N(0,1)|}{1000} = \sigma \times 1 \overset{?}{=} \sigma$$

اندازه قدم یعنی قدر مطلق تفاضل

χ^2 متوالی



الگوریتم ژنتیک:

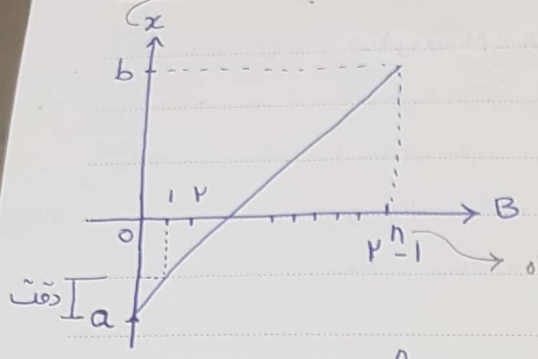
بالای صفحه

می‌خواهیم تبدیلی به دست آوریم که اعداد باینری را به حقیقی تبدیل کند.

Scanned by CamScanner

Subject: y هر فرودگاه در و x هر فرودگاه در محدوده‌ی c تا d

Date: محدوده‌ی a تا b می‌تواند باشد.



در مثال فرودگاه $n=10$

با دقت $\frac{b-a}{2^n-1}$ (مثلاً با دقت ۱۰۰ متری خواهیم مکان فرودگاه‌ها را در مثال فرودگاه محاسبه کنیم.)

$$0/1 = \frac{1000}{2^n-1} \rightarrow n = \log(10001)$$

$$\boxed{n=14}$$

فرمول کلی ص ۳۲

هر عدد حقیقی بین ۰ تا 2^n-1 متناظر می‌شود با یک عدد بین a و b

عدد حقیقی

$$x = a + (b-a)\cdot\frac{B}{2^n-1} \rightarrow B = \sum_{i=0}^{n-1} b_i \cdot 2^i$$

(اگه گری کد باشه فقط این فرمول B فرق داره.)

integer

$$y = c + (d-c)\cdot\frac{B}{2^n-1}$$

* عدد باینری (گری کد) ← تبدیل int ← تبدیل حقیقی ← تابع هزینه $c(x)$ ← معکوس تابع شایستگی $f(x)$

جایگزاری در

تعداد فرودگاه × ۲ × ۱۰ = سایز کروموزوم

تابع هدف در مثال فرودگاه:

$$\min \sum_{i=1}^{C} \min_{j=1}^{A} \left[(x_{c_i} - x_{a_j})^2 + (y_{c_i} - y_{c_j})^2 \right]$$

محاسبه تابع هزینه: جامونتم؟

فرمول تبدیل گری کد به integer:

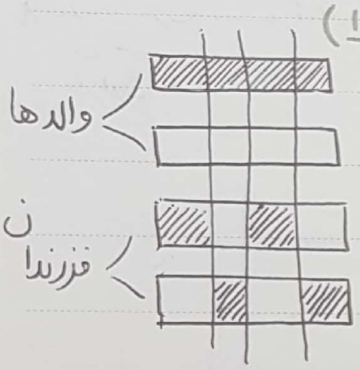
Scanned by CamScanner

انتخاب والدین: μ تا والد انتخاب می‌شه. مثلاً اگه $\mu=10$ بود، ۱۰ تا والد انتخاب می‌کنیم.
اگر یک موجود احتمال (شایستگی) نرمال شده‌اش ۰٫۳ بود، ۳ بار به عنوان والد در جمع والدها آورده می‌شود.
وقتی والدها انتخاب شدند، یک بار برشان می‌زنیم تا والدهای یکسان کنار هم نیفتند. (وگرنه بازترکیبی بی‌معنا می‌شود.)

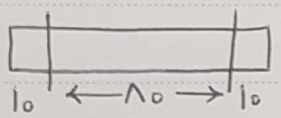
بازترکیبی: P_c (احتمال بچه‌دار شدن ۲ والد. ممکنه بچه‌دار نشوند) ← در الگوریتم ژنتیک $P_c > 0.8$
برای هر دو والد یک عدد بین ۰ و ۱ انتخاب می‌کنیم اگر کوچکتر از ۰٫۸ شد crossover انجام می‌دهیم
وگرنه crossover انجام نشود.

چند نوع بازترکیبی (روش) روی باینری و گری داریم: (۳ روش گفته شد سر کلاس)

۱- چند نقطه‌ای ← اول یه عدد تصادفی یکنواخت مثلاً ۳ (بین ۱ تا تعداد بیت‌ها (منهای ۱))
بعد ۳ عدد تصادفی بین ۰ تا ۶ (n = تعداد بیت‌ها)



به عنوان نقاط بازترکیبی



← عیب: احتمال انتخاب شدن ابتدا و انتها کمتر است.
بنابراین ابتدا و انتهای کروموزوم‌ها یکسان خواهد شد. ⇐ "بایاس شدن به وسط‌ها" ← راه‌حل ①: نقاط افزایش تعداد

59:23:92

↓ حالت افراطی

افزایش به تعداد بیت‌ها → روش «بازترکیبی یکنواخت» → $\boxed{\text{توضیح}}$ ؟
هر بیت شیر یا خط ؟

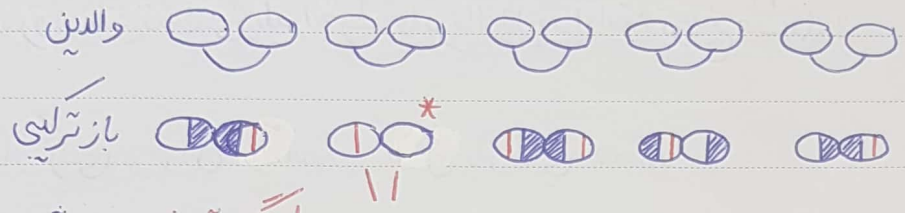
Scanned by CamScanner

Subject
Date

* اگر جهش رخ ندهد ممکن است هرگز به جواب نرسیم. مثلاً اگر تمام افراد جمعیت اولیه بیت ۱۰ ام شون صفر باشد اما در جواب اصلی این بیت باید ۱ باشد، هرگز جواب پیدا نخواهد شد.

* در الگوریتم ژنتیک، تأکید روی بازترکیبی است و احتمال جهش بسیار کم است: $\frac{1}{n} < P_m < \frac{1}{\mu}$

n ← تعداد بیت ها



$\lambda = \mu$?

اگر جهش داشتیم، معکوس می کنیم وگرنه ... ؟

در گری کد جهش جستجوی محلی انجام می دهد و بازترکیبی جستجوی عمومی ⇐ بهتر است.

* توانسته یک نفر از جمعیت اولیه نه بازترکیبی نه جهش روش انجام شه و همین طوری بره نسل بعد ولی انتخاب به طور مستقیم از جمعیت اولیه نداریم.

(خودم + Voice)

PAPCO

Scanned by CamScanner

Subject
Date

موجود جهش یافته — قدم جدید

* در تپه نوردی ← particle ← در الگوریتم تکاملی ← موجود
در PSO

* جهش در تپه نوردی ← یک قدم تصادفی برداشتن

* $N(0,1)$ ← نویز گوسی ، نرمال
↙ تولید یک عدد تصادفی نرمال

* تولید عدد تصادفی یکنواخت بین a و b:
$(b-a)\times x + a$

* در مسائل تحقیقاتی ، شرط خاتمه برابر با تعداد ارزیابی است.

* بازترکیبی و جهش وابسته به روش بازنمایی هستند.

* الگوریتم ژنتیک ۲ مشکل اساسی همگرایی زودرس و ساکن شدن را دارد.

* فرمول کلی به دست آوردن تعداد بیت مورد نیاز با داشتن دقت c :

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

$$\Rightarrow 2^n = \frac{b-a}{c} + 1 \Rightarrow \boxed{n = \frac{\ln\left(\frac{b-a}{c}+1\right)}{\ln 2}}$$

* اگر روش بازنمایی الگوریتم ژنتیک اعداد حقیقی باشد ، می شود الگوریتم DS
اما مزیت انتخاب روش باینری این است که در بازترکیبی می شه عدد را از وسط نصف کرد و...
ولی در عدد حقیقی نمی شه. بازترکیبی چند نقطه‌ای نمی‌توان داشت ؟

* برای هر تن باید $N(0,1)$ متفاوت باشه تا هر جهت رو بشه پیمایش کرد در تپه نوردی ← یعنی در هر iteration برای n بُعد کروموزوم هر بُعد یک $N(0,1)$ جدا دارد تا زاویه هم بتواند متفاوت باشد.

Scanned by CamScanner

Subject
Date

«جلسه ۱۱»

نام	ES	انتخاب بازماندگان	$\mu+\lambda$ ، تورنمنت λ و μ متناسب با شایستگی
بازنمایی	اعداد حقیقی		
جمعیت اولیه	$\mu=[1\ 10]$ تصادفی یکنواخت	شرط خاتمه ویژگی خاص	تعداد ارزیابی اندازه قدم خودتطبیقی
			SUS-RW
ارزیابی	$F(x)$		
انتخاب والدین	$\lambda=7\mu$ تصادفی یکنواخت		
جهش	$P_m=1$ گوسی $x_i'=x_i+\sigma N_i(0,1)$		
بازترکیبی	$P_c<0.4$ (اعتبار بازترکیبی) حسابی (کامل - تکی - ساده)		
استراتژی			

اثبات شده در اعداد حقیقی با $\lambda=7\mu$ سرعت max می‌شه.

اثبات شده این جوری سرعت max می‌شه.

self adaptive

فاکتور $\frac{1}{5}$ موفقیت فقط در ES(1+1) قابل استفاده است.

پس باید استراتژی برای پیدا کردن اندازه قدم پیدا کنیم ← روش خود تطبیقی (خودش اندازه قدم خودش را بتواند تعیین کند)

اول جهش می‌دهیم چون 7μ انتخاب شده ممکنه والدها یکی باشند باهم بعد بازترکیبی می‌کنیم تا والدهای متفاوت انتخاب شده باشند و ویژگی های متفاوت بتواند بازترکیبی شود.

بازترکیبی اگر هم انجام شه روی ۱ یا ۲ جفت فقط انجام می‌شه.

حسابی کامل ← روی همه ژن ها تکی ← ؟

ساده ← از یک ژن به بعد

← سه راهم ابتدا به طور تصادفی یکنواخت انتخاب می‌کنیم.

در خود - تطبیقی

Scanned by CamScanner

Subject
Date

چرا؟!

* در جهش اول x را جهش بدیم یا σ را؟

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

اگه τ نزدیک صفر ← تغییرات σ کوچک می‌شود.
معمولاً نزدیک به ۱/۲ و ۱/۵ است.
هرچه کوچکتر باشد، یادگیری تدریجی‌تر انجام می‌شود.

$$\boxed{x_i' = x_i + \sigma N_i(0,1)}$$

* x_i ها به طور خطی دارند جهش پیدا می‌کنند
* اگر σ را به طور خطی جهش دهیم مشکل دارد چون باید به طور تدریجی یاد گرفته شود یعنی به مقدار جهش بدیم ببینیم خوب بود یا نه بعد عوض کنیم متناسب با خروجی ...

~~تغییراتش شدید~~

یعنی $x_i' = x_i + \sigma' N_i(0,1)$ یا $x_i' = x_i + \sigma N_i(0,1)$ ؟

جهش x با σای که بوده و جایا دیگه وجود نداره!

اول جهش σ بعد جهش x ✓ ← می‌شه از روی σ جدید فهمید تأثیر x رو

* σ را مستقیم نمی‌شه فهمید خوبه یا بد. از روی خوب یا بد شدن x می‌شه فهمید
* احتمالش هست که به خاطر noise گوسی و نه σ جواب x خوب شده باشه. ولی احتمالش کمه.

x_1	...	x_n	σ

→

x_1'	...	x_n'	σ'

کروموزوم یه بیت اضافه می‌شه

* می‌توان اثبات کرد $\sigma N(0,1) = N(0,\sigma^2)$

اعراف معیار σ^2

عدد گوسی با

نرمال

توصیه می‌شه معمولاً از $\mu+\lambda$ استفاده شه در انتخاب بازماندگان
استفاده از μ و λ ← به خاطر اینکه نویز گوسی باعث خوب شدن x شده نه σ پس فراموششان کنیم قبلی‌ها رو.

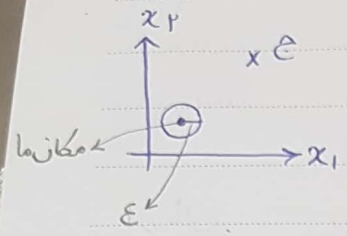
(کلاه رفته باشه سرمون)

(شاید هم خود x خوب بوده که x' خوب شده نه σ) ← $\mu+\lambda$ عه؟

Scanned by CamScanner

Subject
Date

* چه کار کنیم سرعت یادگیری بالاتر بره؟



بهبود → سطح دایره همون نمونه ولی شکلش شبه بیضی ~~به سمت هدف کشیده شده~~ → داریم ($\tilde{x}$ تا n) n بُعدی →

x_1	...	x_n	σ_1	...	σ_n	

بهبود →

x_1	...	x_n	σ_1	...	σ_n	α_1	...		

$\alpha_{\frac{n(n-1)}{2}}$

به سمت جواب بیضی‌ها بچرخه بهتره!
(زاویه هم می‌خواهیم)

عیب: فضای جستجو رو بزرگ می‌کنه ← اگر بازدهی و سرعت کلی را بدتر نکند، از آن استفاده می‌کنیم.

سرعت یادگیری
چرخاندن
یا پیدا کردن σ ها

یادگیری به پارامتر راحته. یادگیری ۲ تا پارامتر ← یعنی ترکیبات مختلف اون دو تا رو آزمایش کن! (توانی ی ره بالا!) : n^m ← تعداد پارامتر

روش CMA ← از روی داده‌ها کواریانس می‌گیره و جهت رو پیدا می‌کنه و با تکامل جهت رو پیدا نمی‌کنه.
خیلی بهتره از اینکه با الگوریتم تکاملی بخواهیم جهت را پیدا کنیم

معیار سرعت: تعداد ارزیابی‌ها

* فرزندان متناظر:

	x_i'		σ_1
	x_i		σ_1
	y_i		σ_2
	y_i'		σ_2

فرزند متناظر والد ۱
والد ۱
والد ۲
فرزند متناظر والد ۲

PAPCO

Scanned by CamScanner

Subject
Date

نام	دلخواه
بازنمایی	جایگشت
جمعیت اولیه	تصادفی یکنواخت
	$\mu = 6$
ارزیابی	وابسته به مسئله
انتخاب والدین	باینری تورنمنت
	$\lambda = 2\mu$
جهش	$P_m = 0.5$
	انواع جهش
بازترکیبی	$P_c = 0.5$

* معمولاً انتخاب بازماندگان رو متناسب با شایستگی می گیرند و انتخاب والدین را متنوع.

* randperm(n) → یه جایگشت قاطی پاتی رندم از اعداد 1 تا n
matlab

عدد تصادفی بین ۰ و ۱ ← ×n ← جزء صحیح ← عدد بین ۰ و n می شه. آرایه $\boxed{1 \cdots n}$ رو شیفت می دیم (اون عدد رو از توش در آورده ایم) بعد همین کار رو برای n-1 می کنیم ...

* جهش نباید جایگشت را خراب کند. (فرزند بازهم باید جایگشت باشد) ← مثلاً جابه‌جایی ۲ خانه‌ی موجود باهم.

* جستجو محلی و عمومی در جایگشت: عمومی: ترتیب عوض شه (یک رشته)
محلی: همسایگی عوض شه.
؟ منظورش تعداد کمه؟
مثلاً یک زیر رشته رو معکوس کنیم.

(۲ عدد تصادفی یکنواخت تولید ...) ↓ محلی

~~می شه چند تا جهش مختلف تولید کرد و برای هر کدوم احتمال گذاشت و با SUS یا RW انتخاب کرد~~ کدومشون انجام شه.

Scanned by CamScanner

Subject
Date

* بازترکیبی هم نباید جایگشت رو خراب کنه ← مثلاً : حلقه‌ها رو پیدا کنیم و جابه‌جا کنیم.

فرزند ۱	۲	۸	۴	۵	۶	۱	۷	۳
والد ۱	۲	۸	۴	۱	۶	۳	۷	۵
والد ۲	۷	۶	۲	۵	۴	۱	۸	۳
فرزند ۲	۷	۶	۲	۱	۴	۳	۸	۵

حلقه اول متناظر با هر فرزندی باشه ،
یعنی اون فرزند متناظره با اون والد!

جستجو محلی
↑

روش دوم : همسایگی های هر عدد رو در دو والد نگاه می کنیم و اگر یک همسایگی در هر دو والد تکرار شده بود ،
* می زنیم : بعد به طور تصادفی یکی از عدد ها را انتخاب می کنیم. (5)
بعد یکی از همسایه‌های آن را انتخاب می کنیم با این اولویت : ۱- * البته اول
۲- همسایه های کمتر (حذف 5 از کل جدول!)
۳- تصادفی

۱	۴*	۶	۸	
۲	۸	۵*	۶	
۳	۶	۷*	۸	
۴	۱*	۸	۵	
۵	۲*	۴	۷	
۶	۱	۳	۲	۷
۷	۳*	۶	۵	
۸	۱	۲	۳	۴

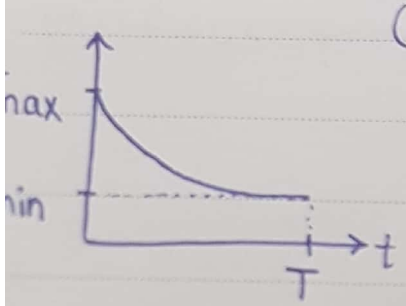
۵	۲	۶	۷	...

برای فرزند دوم ، عدد تصادفی اول رو متفاوت می دیم پس یه فرزند متفاوت تولید می شه.

«جلسه ۱۲»

مقدار پارامترها را چطور مشخص کنیم؟
(اندازه قدم ، λ ، μ ، ...)

① تابعی از زمان و مشخص (deterministic)
② (مشابه با $1/5$ موفقیت)
③ خود تطبیقی



$$\sigma_t = \sigma_{max} + \frac{\sigma_{min} - \sigma_{max}}{T} \times t$$

نسل

PAPCO

Scanned by CamScanner

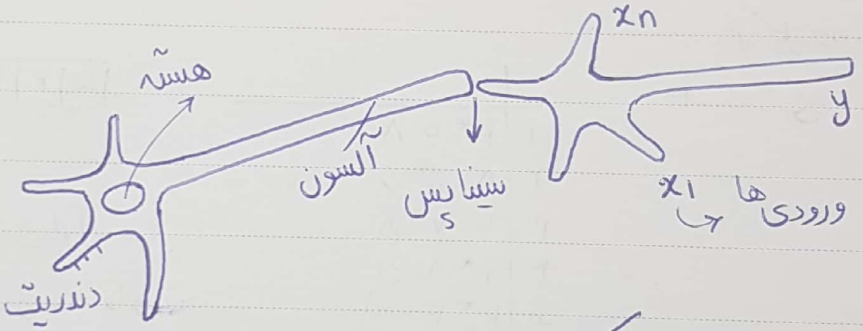
Subject
Date

روش ② : باید اول یک feedback بگیریم . مثل تنوع مان کم است.
پیدا کردن feedback مناسب همیشه ساده نیست. اما اگر پیدایش کنیم این روش بهتر است.
↓
بهتر
↓
اسمش : تطبیقی : Adaptive ← یعنی براساس feedback خودش را تطابق دهد با محیط

* ممکنه پارامترها براساس type های مختلف باشند اما برای اینکه ⊥ کروموزوم در بیاید، باید همه تبدیل شوند به ⊥ نوع مثلاً integer

فصل ۲ : شبکه های عصبی

الهام گرفته شده از شبکه مغز انسان ← کوچکترین المان داخل مغز انسان را باید معرفی کرد : نورون (سلول)
↓
دارای DNA ، هسته ...



* سیناپس واقعاً اتصال فیزیکی نیست شیمیایی است ← بسته های شیمیایی : نورو ترنسمیتر مانند ـــ و ـــ ؟

البته سیناپس الکتریکی

* دوپامین ← تأثیر روی بیماران دوقطبی و پارکینسون

روان شناسان و عصب شناسان خیلی موفق نیستند ← چون فقط یک نگاه کلی می توانند داشته باشند نه جزئی
مثلاً فقط می فهمند دوپامین در مغز سطحش آمده پایین پس ...
(در ابعاد نورون نمی توانند آزمایش کنند)

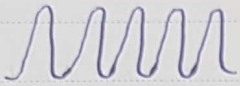
PAPCO

Scanned by CamScanner

Subject
Date

* آزمایش ها روی یک ماهی با اندازه نورون بزرگ انجام شده ولی کاربردی نیستند
یک سری مدل ساده هم به دست آمده و شبکه عصبی هم براساس همان ها نوشته شده.

* خروجی های اکسون به صورت فرکانس و دنباله ای از پالس هاست:



البته رفتارش در زمان یادگیری و عملکرد متفاوته
← اینو کشیده:

* این سیگنال ها در دندریت به ولتاژ تبدیل می شود: ← فرکانس coding و decoding و ولتاژ x و coding و ... فرکانس $f(x)$ $f^{-1}(f(x))$

با توجه به اینکه سیناپس تحریکی یا مهاری باشد، ولتاژ هم تحریکی یا مهاری خواهد بود.
(+ و - در نظر می گیریم) ← میزان تحریکی یا ... بستگی به قوی یا ضعیف بودن سیناپس دارد. ← نشان دادن با عدد w

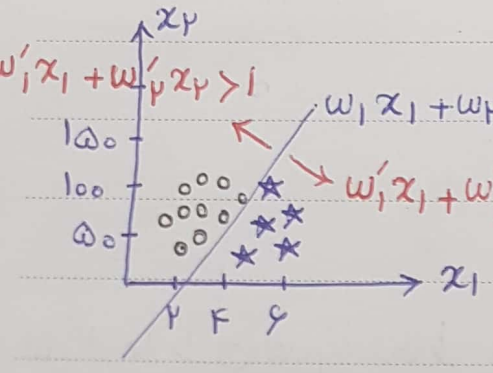
* در نهایت این + و - ها اگر ولتاژ از یک حد آستانه بیشتر باشد، آن تحریک می شود وگرنه آنقدر برایش نمی افتد.

$$\sum_{i=1}^{n} w_i x_i > b \quad y=1$$
$$\quad\quad\quad\quad\quad\quad \text{〃} \quad < b \quad y=0$$

تعداد بسته ها، قوی و ضعیف بودن را تعیین می کند.
یک سری دروازه دارد و بسته به آنکه باز باشند یا بسته، ... (در فرکانس تعداد مهم نه اندازه min و max فرکانس)

ما جنس x و y را یکی فرض کرده ایم

$x_2 \rightarrow$ وزن پرتقال ها و سیب ها
$x_1 \rightarrow$ 〃 〃 〃 قطر



$\rightarrow w_1 \cdot 7 + w_2 \cdot 200 = b \xrightarrow{\div b} \underbrace{\frac{w_1}{b}}_{w'_1} \cdot 7 + \underbrace{\frac{w_2}{b}}_{w'_2} \cdot 200 = 1$

$w_1 \cdot 4 + w_2 \cdot 50 = b \xrightarrow{\div b} \dots$

$$\Rightarrow \begin{cases} 7w'_1 + 200w'_2 = 1 \\ 4w'_1 + 50w'_2 = 1 \end{cases}$$

$$w'_1 = \frac{\begin{vmatrix} 1 & 200 \\ 1 & 50 \end{vmatrix}}{\begin{vmatrix} 7 & 200 \\ 4 & 50 \end{vmatrix}} = \frac{1}{3}$$

$$w'_2 = \frac{\begin{vmatrix} 7 & 1 \\ 4 & 1 \end{vmatrix}}{\begin{vmatrix} 7 & 200 \\ 4 & 50 \end{vmatrix}} = -1/150$$

+ o : سیب ها
- ★ : پرتقال ها

PAPCO

Scanned by CamScanner

Date

$$\frac{1}{3}x_1 - \frac{1}{150}x_2 = 1$$

اگر به جای ۲ بعدی n بعدی بود، باید ویژگی های دیگرشان را هم وارد کنیم (n تا ویژگی حداقل)
رنگ، میزان سدیم، میزان اسیدی و بازی
...

* در n بُعد چطور محاسبه کنیم؟ ← روش تحلیلی

* باید بیشتر از n تا داشته باشیم که بدانیم کدام سیب است و کدام پرتقال.
(یا حداقل n تا) سیب و پرتقال
↓
n' تا

نمایش با 1 — نمایش با -1

$$\begin{array}{c} x_0^1 \leftarrow \\ x_0^2 \leftarrow \\ \\ \\ \end{array}\underbrace{\begin{bmatrix} 1 & x_1^1 & \cdots & x_n^1 \\ 1 & & & \\ \vdots & & & \\ 1 & x_1^{n'} & & x_n^{n'} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ -1 \\ \vdots \\ -1 \end{bmatrix}}_{y} \underbrace{\begin{bmatrix} b \\ w_1 \\ \vdots \\ w_n \end{bmatrix}}_{w}$$

y^i — w_0

برای b اضافه شد

$L_1 + L_2$

$$\underset{n'\times(n+1)}{X} \times \underset{(n+1)\times 1}{W} = \underset{n'\times 1}{Y}$$

→ این را نمی‌شود حساب کرد؟ پس به جاش:

$$\min \underbrace{\tfrac{1}{2}(XW-y)^T(XW-y)}$$

مجموع مربعات خطا

این را می‌خواهیم min کنیم.

$$E = \sum_{i=1}^{n'}\Big(\sum_{j=1}^{n} w_j x_j^i - y^i\Big)^2$$

$$\left[\frac{\partial E}{\partial w_j}\right] = 0$$

یا مشتق نسبت به W (برداری)

$$dE/dW =$$

$$\Downarrow$$

$$X^T(XW-y) = 0$$

$$X^TXW = X^Ty$$

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

(مشتق نسبت به W)
مشتق ماتریسی جدول داره باید حفظ کرد
↓ در کتاب های جبرخطی چند صفحه آخرش

اگه قاطی کردی برحسب ابعاد ماتریس ها بگو.

این کارها را می‌کنیم تا خطا را min کنیم.

در نتیجه‌ی این صفحه: به داده نیاز داریم! ← بعد باید آموزش بدهیم به شبکه عصبی

Scanned by CamScanner

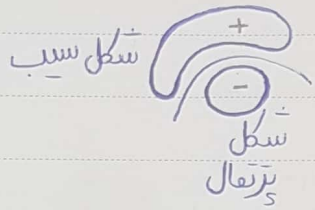
Subject
Date

* Data های پراهمیت ← Data های لب مرزی !

بچه تازه به دنیا آمده، سیناپس هایش data نداشته اند و یاد نگرفته اند هنوز که مثلاً یاد بگیرند
(یک برنامه ی بایاس دارد البته D: که با ژنتیک با خودش آورده.)
۹۰٪ با ژنتیک می آید و ۱۰٪ یادگیری است!
مثلاً شنیدن صدا، دیدن، تکان دادن دست، نفس کشیدن، غذا خوردن، ...

انسان از خیلی قسمت های مغزش استفاده نمی کند!

گاهی خط جداکننده یک خط نیست! مثلاً:



؟ شبکه عصبی با ۳ یا ۴ لایه می توان هر تابعی را با آن پیاده سازی کرد. ← pattem های بهتر برای یادگیری

اگر به جای b بگذاریم $-w_0$،

$$\sum_{i=1}^{n} w_i x_i > -w_0$$

$$\Rightarrow \begin{cases} \overbrace{\sum_{i=0}^{n} w_i x_i}^{Z} > 0 & y=1 \\ \quad '' \quad < 0 & y=0 \end{cases}$$

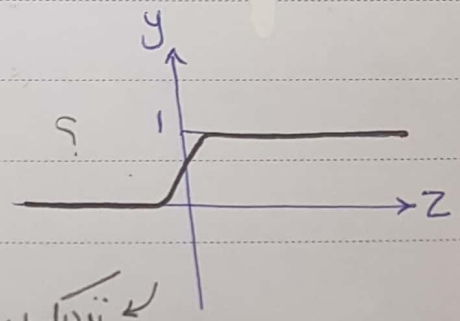
(x_0 همواره برابر با ۱)

$$y = S\left(\sum_{i=0}^{n} w_i x_i\right)$$

سیگموئی $\rightarrow f(x) = \dfrac{1}{1+e^{-x}}$?

$$S(Z) = \frac{1}{1+e^{-Z}}$$

$\begin{pmatrix} w_0 = \text{بایاس} \\ x_0 = 1 \end{pmatrix}$



؟

مزیت: مشتق پذیره

نزدیک به تابع پله

نمی شد اینجا از تابع پله استفاده کرد (؟)

PAPCO

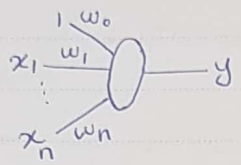
Scanned by CamScanner

Subject
Date

«جلسه ۱۳»

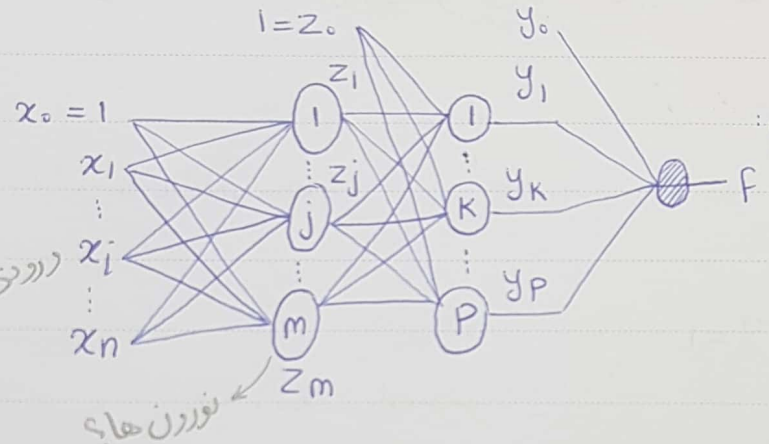
فقط ۱

* در perceptron به سیگموید نیاز نیست.؟
و دو کلاس + و - (+۱ و -۱)



اگر کل داده‌هامان L_2+ L_1 تا باشند و L_1 ای از کلاس +۱ و L_2 تا از کلاس -۱ باشند، داریم:

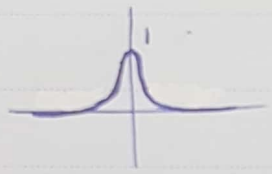
$$w_0 \times 1 + w_1 x_1^1 + \dots + w_n x_n^1 = 1$$
$$\vdots$$
$$w_0 \times 1 + w_1 x_1^{L_1} + \dots + w_n x_n^{L_1} = 1$$
$$w_0 \times 1 + w_1 x_1^{L_1+1} + \dots + w_n x_n^{L_1+1} = -1$$
$$\vdots$$
$$w_0 \times 1 + w_1 \times x_1^{L_1+L_2} + \dots + w_n \times x_n^{L_1+L_2} = -1$$



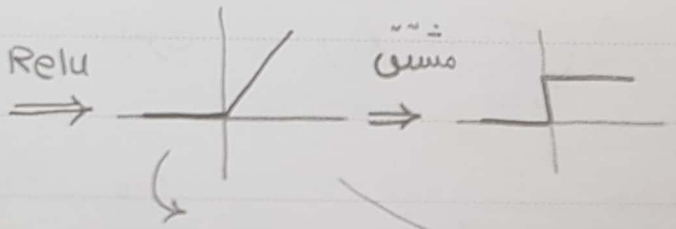
اگر چند perceptron داشته باشیم:
← شبکهٔ چند لایه؟

activation function ← S
← صف قبل

مشتق تابع سیگموید



در شبکه مثلاً ۱۰ لایه
سیگموید به درد نمی‌خوره چون نمی‌تونه یاد بگیره
و وزنش تغییر نمی‌کنه.



برای صفر:
ضرب شبه توش،
کلش صفر می‌شه و
یه سری داده از بین می‌ره.

سرعت learning
هم بالا می‌رود.

چون یه سری عدد بین ۰ و ۱ درهم ضرب می‌شن و
یه عدد نزدیک صفر به دست میاد.

Scanned by CamScanner

Subject:
Date

* activation function باید غیرخطی باشد وگرنه می‌شود perceptron (تبدیل به یک) ! چون می‌توان ادغام کرد با لایهٔ بعدی اش و در واقع یک لایه کم می‌شود!

لایهٔ آخر را می‌توان خطی گذاشت ولی لایه‌های قبلی را نه.

relu : Rectifire linear ...
(دیود)

(چون طرف منفی صفر هست)

Photo

w_{ij} , v_{jk} , u_{kl} : وزن‌ها

حالت کلی‌تر در نظر گرفته شده که چند تا خروجی داریم.

۲ لایه (یک ورودی دو تا مخفی یک خروجی)

supervised regression

۱ نظارتی ← label داریم. ← مثل MLP ← classification

۲ غیرنظارتی ← unsupervised ← clustering

۳ یادگیری تقویتی ← نیمه نظارتی semi-supervised

وقتی به هدف رسیدیم می‌گه به هدف رسیدی ولی در مسیر چیزی نمی‌گه.
(یعنی وسط کار نظارتی نمی‌کند ولی آخر کار نظارت می‌کند)
و پاداشی نمی‌دهد

در بعضی version ها چند step یکبار هم ممکنه پاداش بده

شبکه‌های عصبی می‌توانند با نظارت یا بدون نظارت باشند.

SOM : Self organized Mapping
یک روش خوشه بندی است.

PAPCO

Scanned by CamScanner

Date

regression : تابع پیوسته . هم دسته بندی ها را می دانیم و هم اینکه هر داده در کدام کلاس باید قرار بگیرد ؟

classification ← تابع گسسته . فقط دسته بندی ها را می دانیم.

حالت ۱ خروجی:

$$\text{Min Square Error} = MSE = \frac{1}{N_{data}} \sum_{o=1}^{N_{data}} \left(f^{o} - f^{*o} \right)^{2}$$

f از نوع اعداد حقیقی

حالت چند خروجی:

$$MSE = \frac{1}{N_{data}} \sum_{o=1}^{N_{data}} \sum_{\ell=1}^{q} \left(F_{\ell}^{o} - f_{\ell}^{*o} \right)^{2}$$

؟

	F_1	F_2	F_3		
label ها	1	0	0	C_1	کلاس اول
	0	1	0	C_2	〃 دوم
	0	0	1	C_3	〃 سوم

} classification → سیگموی / هایپربولیک

F_1	F_2	F_3	
1.2	2.3	4.1	regression
⋮	⋮	⋮	

→ تابع activation خطی

تابع هدف

cross on trophy در classification از MSE بهتر جواب می دهد و دقت بهتری دارد.

خروجی ها بین صفر و ۱ باید باشند چون مفهوم احتمال دارند ⇐ باید از سیگموی استفاده شود.

$$\frac{-1}{N_{Data}} \sum_{o=1}^{N_{Data}} \left(f^{*o} \log F^{o} + (1 - F^{*o}) \log(1 - F^{o}) \right)$$

دلیل: min محلی هاش کمه → چند بار run کنی جواب ها تقریباً یه جور در میاد ولی با MSE ...

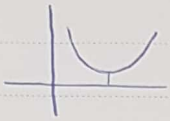
PAPCO

Scanned by CamScanner

Subject:
Date

$-p \ln p - (1-p) \ln (1-p)$

$\hookrightarrow p(\omega)$



$x_1^1 \quad \dots \quad x_n^1 \quad F^1$

$\vdots$

$x_1^{N_{data}} \quad \dots \quad x_n^{N_{data}} \quad F^{N_{data}}$

در حالت چند perceptron
نمی‌شه مشتق رو به طور بسته نوشت.

$\Downarrow$

در جهت عکس گرادیان حرکت کنیم $\Rightarrow$ بیشترین کاهش خطا $\Leftarrow$ با حداکثر سرعت برسیم به جواب

در جهت گرادیان، مقدار error افزایش پیدا می‌کند.

در آن جهت بیشترین افزایش تابع را داریم.

گرادیان:
از یک نقطه تصادفی X^t حرکت می‌کنیم...

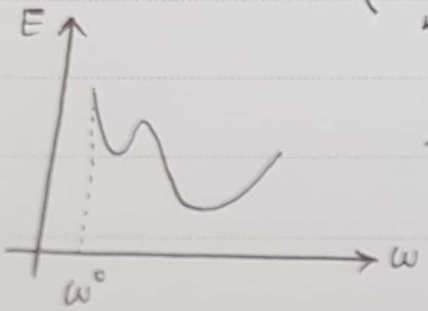
$$W^{t+1} = W^t - \eta \nabla E(W^t)$$ برداری $\rightarrow$ در خلاف G

$$w_{ij}^{new} = w_{ij}^{old} - \eta \frac{\partial E}{\partial w_{ij}} \Big|_{W^{old}}$$

اندازه قدم (آلفا)

مشکلات G
در min محلی گیر می‌کند.
پیدا کردن η سخته. (اگه بزرگ باشه)

با یه اندازه قدم متوسط شروع.
اگه خطا $\rightarrow$ بهتر $\rightarrow$ η رو یه کم بزرگتر کن.
$\rightarrow$ بدتر $\rightarrow$ برگرد سر جای اول و η رو کوچک کن.



Scanned by CamScanner

Subject:
Date

«جلسه ۱۴»

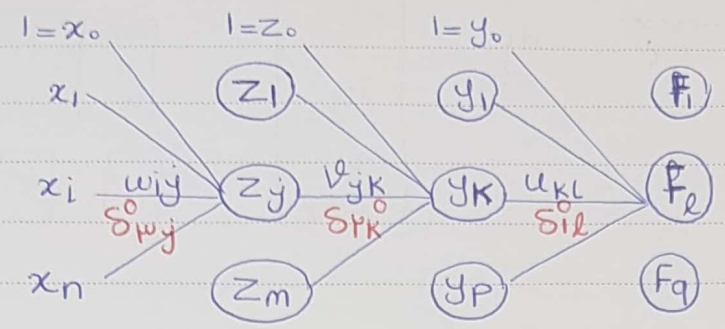
جزوه + Voice مهسان ؟

PAPCO

Scanned by CamScanner

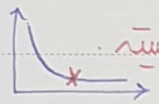
Subject:
Date

«جلسه ۱۵»



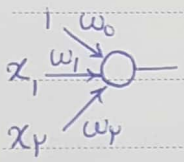
*دفعه قبل مشتق را برای u_{kl} به دست آوردیم

$$X=\begin{bmatrix}1 & x_1^1 & \cdots & x_n^1\\ \vdots & & & \\ 1 & x_1^{N_{data}} & \cdots & x_n^{N_{data}}\end{bmatrix} \qquad Z=\begin{bmatrix}1 & z_1^1 & \cdots & z_m^1\\ \vdots & & & \\ 1 & z_1^{N_{data}} & \cdots & z_m^{N_{data}}\end{bmatrix} \qquad W=\begin{array}{c|c} & j\\ \hline i & \left[w_{ij}\right]\end{array}$$



اینجا وای استیه

CNN را امروزی خواهیم



خطا $E=\frac{1}{2}\sum_{l=1}^{10}(y^l-y^{*l})^2=\frac{1}{2}(y-y^*)^T(y-y^*)$

$$y=\begin{bmatrix}y^1\\ \vdots\\ y^{10}\end{bmatrix} \qquad y^*=\begin{bmatrix}y^{*1}-\frac{20}{7}\\ \vdots\\ y^{*10}-20\end{bmatrix} \qquad w=\begin{bmatrix}w_0\\ w_1\\ w_2\end{bmatrix}$$

?

مشتق جهت دار f در جهت x : ؟

$$\underset{3\times1}{\frac{dE}{dw}}=\underset{\substack{10\times3\\ \to 3\times10}}{\left(\frac{dy}{dw}\right)^T}\underset{10\times1}{(y-y^*)} \quad \text{I} \qquad y^t=\begin{bmatrix}1 & x_1^1 & x_2^1\\ \vdots & & \\ 1 & x_1^{10} & x_2^{10}\end{bmatrix}w^t$$

$$w^{t+1}=w^t-\eta x^T(y-y^*) \qquad \left(\frac{dy}{dw}\right)^T=x^T \quad \text{II} \Longleftarrow \underset{10\times3}{x}\,\underset{3\times1}{w}=\underset{10\times1}{y}$$

Transpose

$$\text{I}, \text{II} \Rightarrow \frac{dE}{dw}=\underset{3\times10}{x^T}\underset{10\times1}{(y-y^*)}$$

$$\cancel{\frac{dE}{dw}=x^Tx^*}$$

PAPCO

Scanned by CamScanner

Subject:
Date

$$\frac{dE}{dw} = \begin{bmatrix} \partial E/\partial w_0 \\ \partial E/\partial w_1 \\ \partial E/\partial w_2 \end{bmatrix}$$

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0{,}01 \times X^T y^*$$

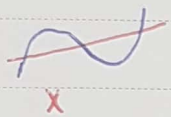
$$\begin{bmatrix} \ \end{bmatrix} = \begin{bmatrix} \ \end{bmatrix} - 0{,}01 \times X^T (\ )$$

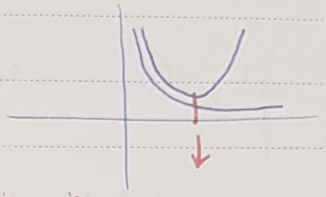
$y(w^t) - y^*$

over fit و under fit

به یک شکل خاص برای داده های دیده نشده برسد؟

آنقدر تعداد لایه‌هامان کم باشد که نتواند شکل شبیه شکل اصلی داده های دیده نشده را به دست آورد





باید همین جا متوقف شود

* فرق validation و test : از test به هیچ وجه نباید استفاده کرد و با نگاه به test، تغییرات بدهیم (نباید)

← برای تعیین meta parameter ها

* وقتی تعداد data ها زیاد ← validation : کم ← K-fold cross validation : ← برای پیدا کردن meta parameter ها

λ : یک مقدار انتخابی : چقدر اهمیت بدهیم به اینکه وزن‌ها کم شوند نسبت به اینکه خطا کم شود.

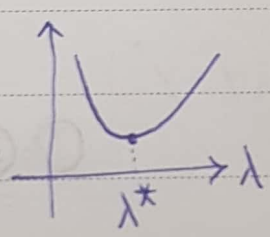
$$E = \frac{1}{2}(y-y^*)^T(y-y^*) + \lambda^* w^T w$$

1	2	3	4

1	2	3	4 ✓
1	4	3	2 ✓
1	2	4	3 ✓
2	3	4	1 ✓

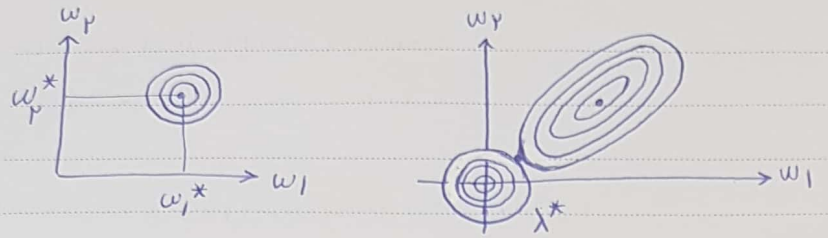
$\lambda = 0{,}1$ $\quad 0{,}7$

$\lambda = 0{,}2$ $\quad 0{,}8$



Scanned by CamScanner

Subject:
Date



- تعداد iteration ها در گرادیان ؟

$$E=\frac{1}{2}\sum_{o=1}^{N_{data}}\sum_{\ell=1}^{9}\left(F_\ell^o-F_\ell^{*o}\right)^2$$

نصف مجموع مربعات خطا
E تابع هدف ما است

قرمز در بخش بالای ۲ صفحه قبل

δ : اندازه قدم

$$w_{ij}^{t+1}=w_{ij}^{t}-\eta\sum_{o=1}^{N_{data}}x_i^o\,\delta_{3j}^o$$

$$v_{jk}^{t+1}=v_{jk}^{t}-\eta\sum_{o=1}^{N_{data}}z_j^o\,\delta_{2k}^o$$

$$u_{k\ell}^{t+1}=u_{k\ell}^{t}-\eta\sum_{o=1}^{N_{data}}y_k^o\,\delta_{1\ell}^o$$

آلفا

در مرحلهٔ آخر میزان خطا برابر است با:

$$\left(F_\ell^o-F_\ell^{*o}\right)$$

$$\Downarrow$$

دلتا $\delta_{1\ell}^o=S'\left(\sum_{k=0}^{P}u_{k\ell}\,y_k^o\right)\left(F_\ell^o-F_\ell^{*o}\right)$

$$\delta_{2k}^o=S'\left(\sum_{j=0}^{m}v_{jk}\,z_j^o\right)\sum_{k=1}^{\ell}u_{k\ell}\,\delta_{1\ell}^o$$

$$\delta_{3j}^o=S'\left(\sum_{i=0}^{n}w_{ij}\,x_i^o\right)\sum_{k=1}^{P}\underbrace{v_{jk}}_{\text{وزن}}\,\underbrace{\delta_{2k}^o}_{\text{خطا}}$$

مشتق سیگموید

خطوط قرمز در photo : اینکه خطاها از کجا می آیند در هر مرحله

PAPCO

Scanned by CamScanner

Date

MLP ← full-connected است

CNN ← " نیست

* به صورت ماتریسی این فرمول ها را به دست بیاورید!

$$\sum_{o=1}^{N_{data}} z_j^o \, \delta_{2k}^o = \frac{\partial E(t)}{\partial v_{jk}}$$ (در لحظه t) باید: $\Longrightarrow$ اثبات: $$\frac{\partial E}{\partial v_{jk}} = \sum_{o=1}^{N_{data}} \sum_{l=1}^{q} (F_l^o - F_l^{*o}) \frac{\partial F_l^o}{\partial v_{jk}}$$

$$F_l^o = S\Big(\overbrace{\sum_{k=0}^{p} u_{kl} y_k^o}^{A}\Big) \longrightarrow \frac{\partial F_l^o}{\partial v_{jk}} = \frac{\partial F_l^o}{\partial A} \times \frac{\partial A}{\partial v_{jk}} = S'\Big(\sum_{k=0}^{p} u_{kl} y_k^o\Big) u_{kl} \frac{\partial y_k^o}{\partial v_{jk}}$$

$$y_k^o = S\Big(\overbrace{\sum_{j=0}^{m} v_{jk} z_j^o}^{B}\Big) \longrightarrow \frac{\partial y_k^o}{\partial v_{jk}} = \frac{\partial y_k^o}{\partial B} \times \frac{\partial B}{\partial v_{jk}} = S'\Big(\sum_{j=0}^{m} v_{jk} z_j^o\Big) z_j^o$$

$$\Longrightarrow \frac{\partial E}{\partial v_{jk}} = \sum_{o=1}^{data} \sum_{l=1}^{q} \delta_{1l}^o \, u_{kl} \frac{\partial y_k^o}{\partial v_{jk}} = \sum_{o=1}^{N_{data}} \delta_{2k}^o z_j^o$$

طبق تعریف ص قبل

اسم این روش گرادیان، error-back-propagation هست (انتشار خطا به عقب) ← چون ...

هر شبکه عصبی که feedback نداشته باشد می توان از این روش در آن استفاده نمود.

Scanned by CamScanner

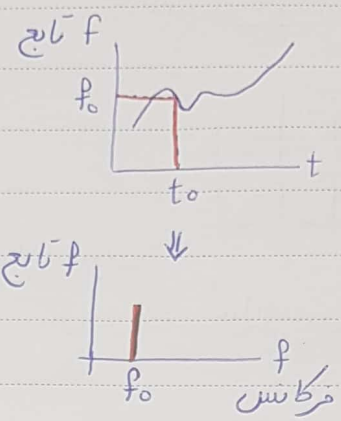
Subject:
Date

« جلسه ۱۶ »

CNN : convolutional Neural Network

از یک فضا مثل زمان ما را به فضای دیگر
مثلاً فرکانس می‌برد.
تبدیل فوریه: برای ساختن یه منحنی چه فرکانس
هایی نیاز است ...

فرکانس: فضای تکرار



چه تعداد تکرار از f_0 هست؟

convolution در تصویر:

* تصویر دیجیتالی در حالت گسسته: یک سری pixcel که هر کدام یک نقطه است (resolution ممکنه فرق کنه)
برای هر نقطه اگر سیاه سفید باشه ۱ عدد و اگر رنگی باشد ۳ عدد نیاز است.
هر کدام ۸ بیت — نشان دهنده شدت رنگ

⇐ ۱ ماتریس یا ۳ ماتریس

کانولوشن در تصویر: مثلاً یک فیلتر ۳x۳ داریم (از نقطه (۲و۲) شروع می‌کنیم تا نقطه (m-1و۲) و فیلتر را اعمال می‌کنیم
پیش می‌رویم و سطرهای بعد تا به نقطه (n-1و m-1) برسیم
پس یک ماتریس n-۲xm-۲ همین طور به دست می‌آید.

$$p'_{ij} = \sum_{i'=-1}^{+1} \sum_{j'=-1}^{+1} a_{i'j'} P_{(i-i')(j-j')}$$

a_{-1-1}	a_{-10}	a_{-11}
...	a_{00}	...
...	...	...

فیلتر 3x3

photo ?

PAPCO

Scanned by CamScanner

Subject:
Date

مشتق افقی

0	0	0
-1	0	1
0	0	0

لبه‌های افقی می‌بینیم

مشتق مورب

0	0	1
0	0	0
-1	0	0

مشتق عمودی

0	-1	0
0	0	0
0	1	0

فیلتر مشتق افقی^2 + فیلتر مشتق عمودی^2 ⇒ لبه‌های دایره‌ای را نشان می‌دهد.

فیلتر میانه ← خطی نیست: نویزهای نقطه نقطه را حذف می‌کند.

حذف نویزهای فرکانس بالا

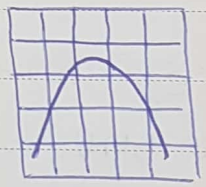
فیلتر پایین گذر

(تیزی‌ها را از بین می‌برد. جاهایی که تغییرات شدید، اون‌ها رو کم می‌کنه ⇐ مات می‌شه)

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

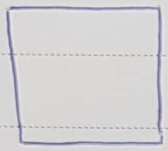
هر فیلتر یک pattern را می‌خواهد بدست بیاورد ← گاهی فیلترها را این گونه نشان می‌دهند:

مثلاً:



از اینجا تا ص بعد Voice

⇒ هر نورون ۹ تا ورودی می‌گیره و یه خروجی می‌ده.



n×m

(با فیلتر ۳×۳)

0
0
⋮
0

} ? تا m-۲ نورون که full-connected نیستند و هر کدام کار یک فیلتر را می‌کنند.

یعنی از تمام پیکسل‌ها ورودی نمی‌گیرد!

در لایهٔ اول ورودی نورون‌ها fix است ولی در لایه‌های بعدی نه.

مثلاً w۱ برای نورون اول برابر با w۲ برای نورون آخر (یعنی چی؟)

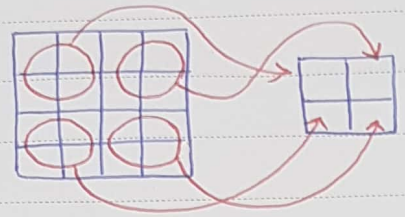
PAPCO

Scanned by CamScanner

Subject:
Date

Pooling ← معمولاً ۲x۲ ← یعنی به ازای هر ۴ پیکسل در تصویر اولیه فقط ۱ پیکسل قرار می‌دهند
(یا max یا متوسط آن ۴ تا را قرار می‌دهند)

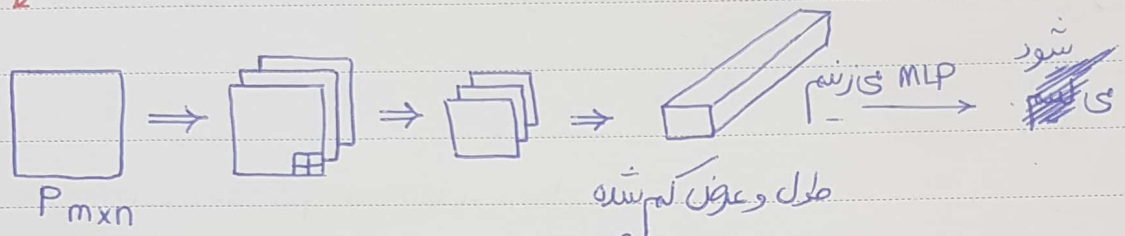
⟸ بالا بردن سرعت ≡ انجام محاسبات
پایین آوردن resolution



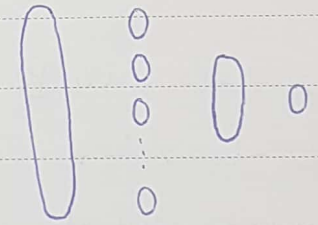
$y = w_{11} P_1 + w_{12} P_2 + \dots + w_{19} P_9$

لایهٔ اول

photo ✓



طول و عرض کم شده
ولی عمق زیاد شده.



Voice

Scanned by CamScanner

Subject:
Date

مهمترین روش شبکه عصبی بدون نظارت: خوشه بندی ← مثل kmeans

براساس شباهت ویژگی‌های ورودی، داده‌ها را خوشه خوشه می‌کنیم.

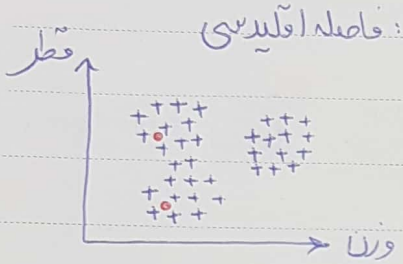
معیارهای شباهت: - فاصله اقلیدسی (هرچه فاصله ↓ شباهت ↑)

- ...

c-mean = k-mean ← FCM در فازی

- C-mean:

* فرض: داده‌ها ۲ بعدی و یه سری سیب/پرتقال بدون label / معیار شباهت: فاصله اقلیدسی



* اگه معیار شباهت درست انجام نشه، سیب و پرتقال‌ها از هم جدا نمی‌شوند.

① ۲ مرکز ← تولید به طور تصادفی یکنواخت
↳ انتخاب از بین داده‌ها به طور تصادفی یکنواخت به تعداد خوشه‌ها که می‌خواهیم

② بعد به داده‌ها براساس فاصله از نزدیک‌ترین مرکز label های ۱ و ۲ می‌دهیم

③ مرکزها را update می‌کنیم ← مختصات داده‌های مرکز ۱ باهم جمع ← تعداد ÷
" " ← " " " ۲ " " ← " "

④ تکرار مرحله ② و ③ ...

$$u_{ik} = \begin{cases} 1 & x^k \in i \\ 0 & x^k \notin i \end{cases} \qquad \|x^k - v_i\| < \|x^k - v_j\| \quad j \neq i$$

norm

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

مرکز جدید

مختصات مجموع داده‌های خوشه i ?

تعداد داده‌های خوشه i $(1+0+1+1+\dots)$

داده اول در خوشه i هست

داده دوم در خوشه i نیست.

PAPCO

Scanned by CamScanner

Subject:
Date

«جلسه ۱۷»

- ۲ نوع شبکه عصبی امروز گفته می‌شود ← مدل خوشه‌بندی : SOM : Self Organized Mapping
Redial Based Function : RBF
توابع پایه‌ی شعاعی

SOM : لایه اول :

z_1 تا z_m خوشه‌ها هستند

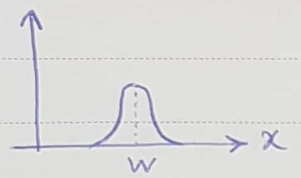
$$z_j^i = e^{-(x^i - w)^T (x^i - w)}$$

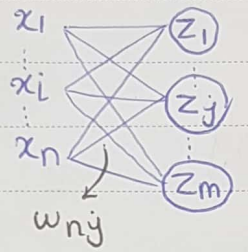
← نمایی : activation f

مرکز خوشه‌ی j ام : w_j ← w : مرکز خوشه‌ها

$$w_j = \begin{bmatrix} w_{1j} \\ \vdots \\ w_{ij} \\ \vdots \\ w_{nj} \end{bmatrix}$$

z_j^i : تعلق داده‌ی i به خوشه‌ی j ← هرچه x از w فاصله بگیرد، تعلق‌ش بیش‌تر می‌شود ؟





$$w_j^{t+1} = w_j^t + \cdots (x - w_j)$$

ضریب

اگر x نزدیک مرکز نباشد، نباید تأثیرگذار باشد. (مانند c-means)

$z(x)$: مقدار مناسب

$$\boxed{z_j(x) = e^{-(x-w_j)^T (x-w_j)}}$$

$$z_j^l = z_j(x^l)$$

متوسط‌گیری پنجره‌ای : $\bar{x} = (1-\beta)\bar{x} + \beta x_{new}$ $\quad \beta = \frac{1}{K}$ $\quad$ اندازه پنجره = K

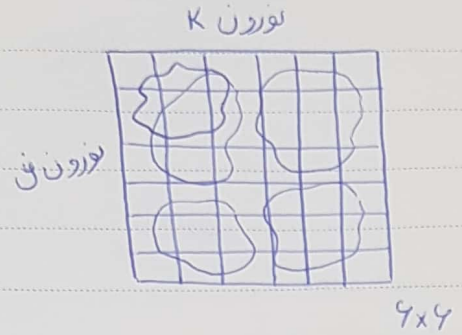
$$\Rightarrow \boxed{w_j^{t+1} = (1-\beta) w_j^t + \beta \underbrace{z(x)}_{\text{فاصله داده تا } w} (x - w_j)}$$

← این وزن همسایگی را لحاظ نمی‌کند. پس به تعداد w ها خوشه داریم.

PAPCO

Scanned by CamScanner

Subject:
Date



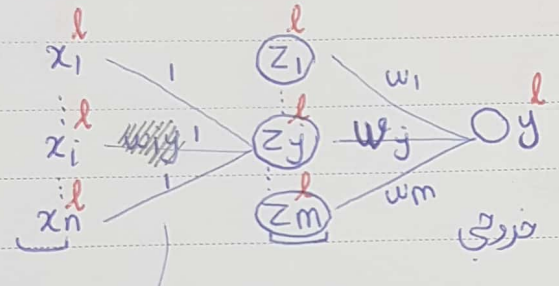
→ در این ۲ بعدی وزن لزوماً به تعداد w ها خوشه نداریم چون امکان دارد یک خوشه چند مرکز داشته باشد.

باید تعداد داده‌ها اینجا بیشتر باشد تا بتوان visualize کرد.

RBF ← برای رگرسیون و کلاسیفایی از آن استفاده می‌شود.

ترکیب خطی

← یک سری توابع شعاعی معروف را به عنوان پایه در نظر می‌گیریم و توابع جدید را از روی آن‌ها به دست می‌آوریم



وزن ها = 1

مرکز z ها

$$z_j = e^{-(x-v_j)^T(x-v_j)}$$

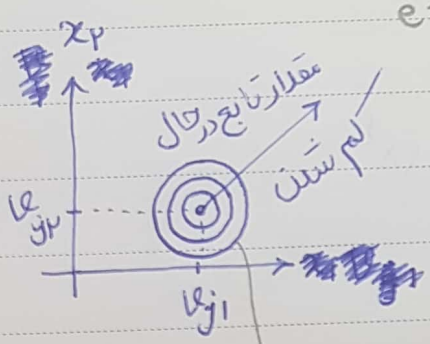
تابع پایه j ام:

① از هم مستقل باشند

② تمام توابع را بسته با آنها ساخت.

بیشترین مقدار در حالتی که $x = v_j$ باشد ⇐ $e^0 = 1$

هرچه از مرکز دور شویم مقدار کوچک می‌شود.



روی سطح هر دایره مقدار ثابت است.

اولین مقدار v_j مثل بقیه وزن‌های شبکه عصبی: تصادفی یکنواخت

$$E = \frac{1}{2}\sum (y^i - y^{*i})^2$$

خطا → min تا بعد از خطا مشتق می‌گیریم به دست بیاوریم.

PAPCO

Scanned by CamScanner

Subject:
Date

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

* l: سطر l ام داده های آموزشی (N_{data} تا ۱ از l)

$$y^l = \sum_{j=1}^{m} w_j z_j^l$$

* $\overbrace{y^{*1} \sim y^{*N_{data}}}^{y^*}$ در رگرسیون می شوند مقادیر رگرسیون و در کلاس بندی می شوند label ها

$$\frac{\partial y^l}{\partial w_j} = z_j^l$$

$$\boxed{w_j^{t+1} = w_j^t - \eta \sum_{l=1}^{N_{data}} (y^l - y^{*l}) z_j^l}$$

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_j \\ \vdots \\ w_m \end{bmatrix}$$

$$Z = \begin{bmatrix} z_1^1 & & \\ & & \\ & & z_m^{N_{data}} \end{bmatrix} \quad X = \begin{bmatrix} x_1^1 & & \\ & & \\ & & x_n^{N_{data}} \end{bmatrix} \quad y = Zw$$

$$\frac{\partial E}{\partial w} = (y - y^*) \frac{\partial y}{\partial w} \rightarrow \frac{\partial E}{\partial w} = Z^T (y - y^*) = 0 \Rightarrow Z^T y = Z^T y^* \Rightarrow Z^T Z w = Z^T y^*$$

$$w = \underbrace{(Z^T Z)^{-1}}_{} Z^T y^*$$

ممکنه معکوس نداشته باشه (پایه ها از هم مستقل نباشند)
(مقادیر ویژه اش صفر دارد)

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

$Z^T Z$: $\underbrace{\text{semi-possitive}}$ است

$$\underset{1\times n}{\alpha^T} \underset{n\times n}{A} \underset{n\times 1}{\alpha} \geq 0$$

A مقادیر ویژه ماتریس بزرگتر مساوی صفر هستند.

$$\alpha^T Z^T Z \alpha = (Z\alpha)^T (Z\alpha)$$

$$[\beta_1 \cdots \beta_n] \rightarrow \beta_1^2 + \cdots + \beta_n^2 \geq 0$$? چی گفت اینجا

PAPCO

Scanned by CamScanner

Subject:
Date

$Z^TZ+\lambda I \rightarrow$ حتماً مقادیر ویژه اش بزرگتر از صفر هستند. $\xrightarrow{\text{دلیل}}$ هر ماتریس Semi-positive را می شه به این صورت نوشت:

$(\lambda>0)$

$\alpha^T(Z^TZ+\lambda I)\alpha$

$A=\varphi^T\Lambda\varphi$

بردار ویژه؟

$\varphi^T(\Lambda+\lambda I)\varphi=\varphi^T\Lambda\varphi+\lambda\varphi^T\varphi \leftarrow \begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_m\end{bmatrix}$

مقادیر ویژه

$\frac{\partial E}{\partial w}=Z^T(Zw-y^*)+\lambda w=0$

$(Z^TZ+\lambda I)w=Z^Ty^* \Rightarrow w=(Z^TZ+\lambda I)^{-1}Z^Ty^*$

$A^{-1}=\varphi^T\Lambda^{-1}\varphi$

$\frac{\partial E}{\partial v_j}=\sum_l (y^l-y^{*l})\frac{\partial y^l}{\partial v_j} \rightarrow \frac{\partial y^l}{\partial v_j}=w_j\frac{\partial z_j^l}{\partial v_j}=w_j\underbrace{z_j^l}_{e^u=u'e^u}(x^l-v_j)\times 2$

z_j^l

$\Rightarrow \frac{\partial E}{\partial v_j}=2\sum_l (y^l-y^{*l})w_jz_j^l(x^l-v_j)$

$\Rightarrow v_j^{t+1}=v_j^t-\eta w_j\sum_{l=1}^{N_{data}}(y^l-y^{*l})z_j^l(x^l-v_j) \rightarrow$ قانون آپدیت مراکز

اگر با روش گرادیان مشتق می گرفتیم، تعداد min محلی ها زیاد می شد و به درد نمی خورد.
می شد w ها را با ES و v ها را با معکوس به دست آورد. ← ترکیبشون چیز خوبیه.

LM: $E=\frac{1}{2}(y-y^*)^T(y-y^*)$

$\Delta y=J\Delta w$

اگر پارامتر تنظیم کننده را نداشته باشد. $\rightarrow$ گوس - نیوتن

اگر پارامتر تنظیم کننده را داشته باشد. $\rightarrow$ LM

اگر خطی بود: $w=(Z^TZ+\lambda I)^{-1}Z^Ty^* \leftarrow$ در 1 iteration جواب به دست میاد.

" " نبود: با بسط تیلور تابع غ خطی را در 1 نقطه خطی می کنیم و بعد در نقطهٔ دیگر و...

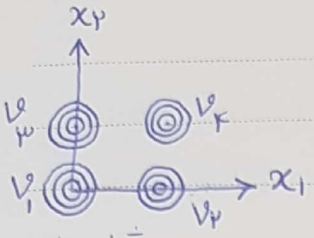
PAPCO

Scanned by CamScanner

Date

«جلسه ۱۸»

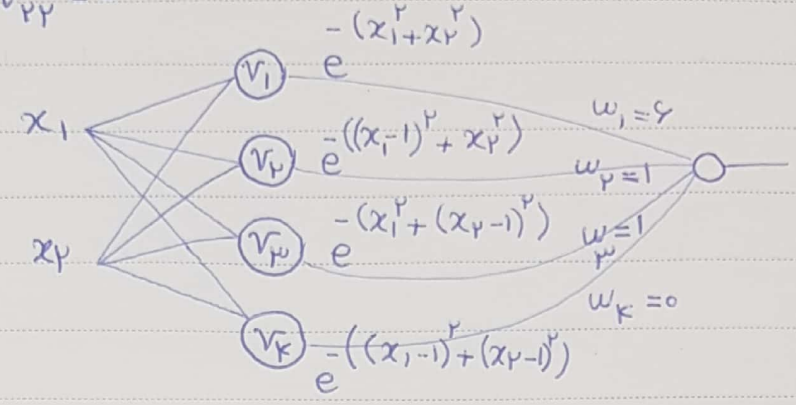
- مدل کردن مسئله XO :



شعاعی را این طوری فرض کنیم

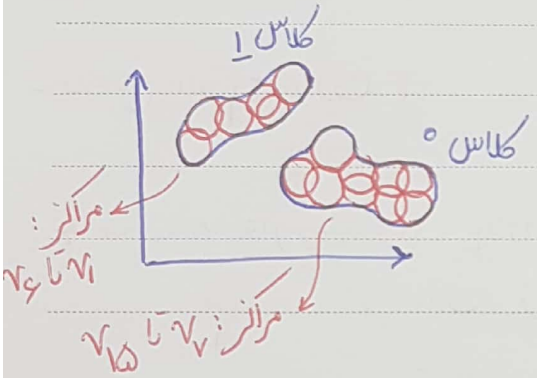
$v_{21} = 1$

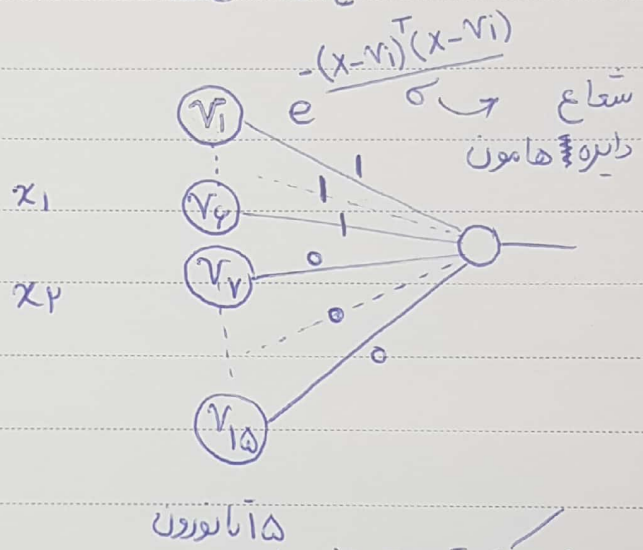
$v_{22} = 0$



v_1	0	0	0
v_4	1	1	0
v_2	1	0	1
v_3	0	1	1

$$e^{-(x-v_i)^T(x-v_i)} = e^{-x^Tx} = e^{-\|x\|^2}$$

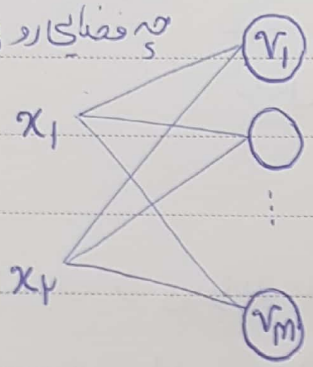




↑ از این سوال‌ها ممکنه تو امتحان بیاد!

$$e^{-\alpha(x-v_i)^T(x-v_i)}$$ ← معمولاً ثابت فرض می‌کنند.

مشخص می‌کنه هر نورون چه فضایی رو می‌پوشونه



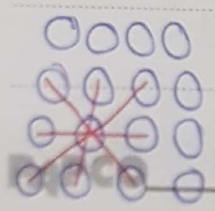
ساده‌ترین مدل SOM :

۰- مرکزهای v_i را به طور تصادفی یکنواخت مقداردهی می‌کنیم.

۱- برای هر داده آموزشی، نزدیک‌ترین نورون را پیدا می‌کنیم.

معمولاً: فاصله اقلیدسی

$e^{-(\dots)^T(\ )}$ هرجا بیشتر بود



۲- برای نزدیکترین نورون پیدا شده و نورون‌های همسایه، وزن‌ها را به صورت زیر اصلاح می‌کنیم:

$$v_i^{t+1} = (1-\beta)v_i^t + \beta\left(1-\frac{r}{d_{max}}\right)(x-v_i^t)$$

Scanned by CamScanner

$d_{max} = ?$

با SoM داده‌های n بُعدی را می‌توان در فضای ۲ بعدی نمایش داد.
یک سری نواحی مشخص می‌کند (مثل نقشه که یک صفحه ۲ بعدی است و کُره را نمایش می‌دهد)

(یه سری توضیح داد...)

- بخش سوم: فازی:

* سه دسته متغیر داریم
  - crisp (دقیق)
  - فازی (نادقیق)
  - تصادفی

* crisp: مقدارش مشخص باشه.
* تصادفی فرقش با نادقیق:

برایش نمی‌توان مقدار دقیق تعیین کرد ولی (بازه) می‌توان برایش تعریف کرد که ۱۰۰٪ در آن بازه بیفتد.

مثلاً سرعت زیاد رو می‌تواند نفرات مختلف به طور سلیقه‌ای بازه متفاوت برایش تعیین کنند ولی بالاخره برای هر نفر یک میزان ثابت دارد.
یا قد بلند یا صحبت آرامتر یا گرمای هوا

متغیرهای نادقیق معمولاً متغیرهای زبانی هستند.

Scanned by CamScanner

Subject:
Date

* یک درجه و تعلق برای متغیرهای نادقیق وجود دارد. مثلاً کسی که اعتقاد داره سرعت بالاتر از $70 \frac{km}{h}$ زیاد محسوب می‌شود، بین سرعت ۸۰ و ۹۰ تفاوت را متوجه می‌شود.

* متغیر تصادفی : مثلاً : با احتمال ۷۰ درصد باران می‌آید.

← در آن فراوانی مطرح است ولی در نادقیق فراوانی مهم نیست.

map کردن یه متغیر به فضای فراوانی و صحبت راجع به آن mapping و نه خود متغیر.

* ۴ نوع متغیر برحسب رفتارشون داریم (دنباله اعداد شون) در یک سری زمانی

- periodic
- آشوبی $x_n = A x_{n-1} (1 - x_{n-1})$ ← یک دنباله هیچ گاه تکرار نمی‌شود.
  بین ۳ و ۴
- ؟
- تصادفی

* منطق فازی تعمیم نظریه مجموعه هاست : (اگر فرض کنیم تعلق فقط ۰ و ۱ است، منطق فازی برابر با نظریه مجموعه‌های شود.)

یک شی به یک مجموعه

* در منطق فازی تعلق عددی بین ۰ و ۱ است. یعنی یک شی s با یک تعلق عضو یک مجموعه است و با (آن تعلق - ۱) جزو آن مجموعه نیست.

- مجموعه : تعدادی از اشیا که کنار هم باشند و لزوماً نه به دلیل یک ویژگی مشترک شان.

نمودار ون برای نمایش مجموعه ها

A (a b c d)

حرف بزرگ برای مجموعه

لاتین

- مجموعه مرجع : تمام اشیا در آن حوزه که داریم صحبت می‌کنیم. (تمام اشیا مورد نظر مسئله و نه تمام اشیاء موجود)

مجموعه

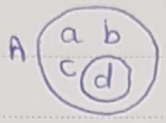
X ←

PAPCO!

Scanned by CamScanner

Subject:
Date

زیرمجموعه:

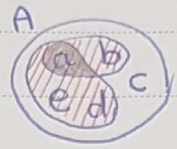


$A \subset B : a \in A \Rightarrow a \in B$

$A \subset B = \{x \mid x \in A \Rightarrow x \in B\}$

* به طور شهودی با نمودار ون می‌توان عملیات مجموعه‌ها را فهمید.

اجتماع، اشتراک، متمم:



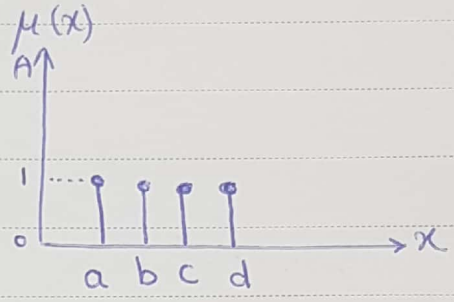
$A \cup B = \{x \mid x \in A \vee x \in B\}$

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

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

نمایش جدیدی برای مجموعه‌ها: ← قابلیت تعمیم را به ما می‌دهد.

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



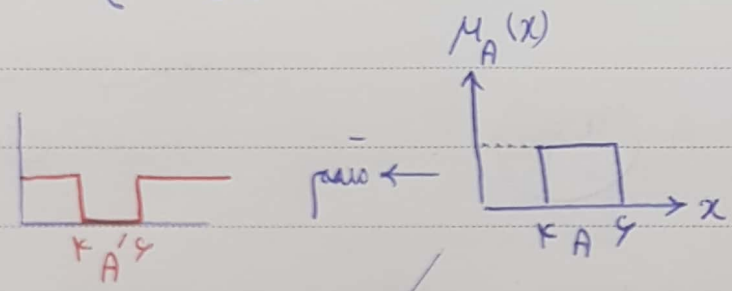
$\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max[\mu_A(x), \mu_B(x)]$

$\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) = \min[\mu_A(x), \mu_B(x)]$

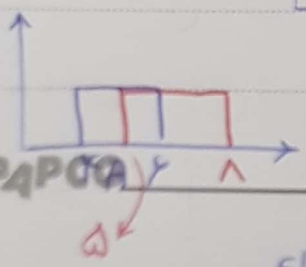
$\mu_{A'}(x) = 1 - \mu_A(x)$

؟ $\mu_{A \subset B}(x) = \begin{cases} 1 & \mu_A(x) \leq \mu_B(x) \\ 0 & o.w \end{cases}$

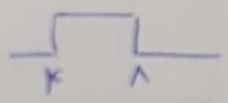
مجموعهٔ اعداد حقیقی $[4, 6]$:



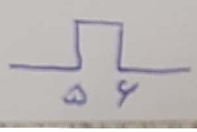
اجتماع و اشتراک:



اجتماع



اشتراک



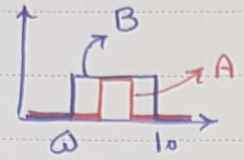
Scanned by CamScanner

Subject:
Date

«جلسه ۱۹»

$A = \{x \mid x \in A\}$

نمایش $A \subset B$:



نمایش مجموعه مرجع با μ

این جلسه، اسلاید داشت

22:40 به نام نمایش برداری یا ماتریسی ← یک نمایش دیگر هم هست
(استفاده در تابع تعلق)

Scanned by CamScanner

Subject:
Date

«جلسه ۲۰»

~~می‌خواهیم یک مجموعه داشته باشیم که نمرات درس هوش مصنوعی باشد و~~

نمرات درس هوش مصنوعی $X_1 = \{20, 19, 18, 17, 16, 15, 14, 12, 11, 10\}$

" " امنیت $X_2 = \{20, 19, 8, 17, 16, 15, 14, 12, 8, 7, 5\}$

به فرم بردار

$A_1 = \{1, 0, 0, 1, 0, 1, 0, 0, 0, 0\}$ → X_1 متناظر با $A_1 \subseteq X_1$

$A_2 = \{1, 0, 0, 1, 1, 0, 0, 1, 0, 0\}$ → X_2 " " $A_2 \subseteq X_2$

$\cancel{A_1 \cap A_2 = \{1, 0, 0, 1, 0, 0, 0, 0, 0, 0\}}$ ✗ → ** نمی‌توان اشتراک گرفت

چون مجموعهٔ مرجع‌شان یکی نیست!

از این سؤال می‌آید!

البته با مجموعهٔ ۲ بُعدی

→ اجتماع و اشتراک گرفتن: مزیت راحت می‌شود.

→ به ازای آن‌ها که آمده ۱ می‌گذارد و " " " " نیامده ۰ "

* می‌توانیم هم مرجع کنیم! ← A_1 و A_2 را ۲ بُعدی کنیم. ← بعد می‌توان اشتراک گرفت.

* اگر اسامی کلاس هوش را در یک مجموعه بنویسیم، ممکن است چند تا علی وجود داشته باشد ولی مجموعه تکرار نمی‌تواند داشته باشد ⇐ راه‌حل: نام خانوادگی و شمارهٔ تلفن و ... را هم اضافه کنیم.

⇐ یک مجموعه از اعضای n بُعدی به دست می‌آید.

* ابزاری برای افزایش ابعاد عناصر مجموعه: توسعهٔ استوانه‌ای

* در هم مرجع کردن باید ترتیب بُعدها رعایت شود. (مثلاً در هر دوی A_1 و A_2، بُعد اول شماره هوش مصنوعی و بُعد دوم شان امنیت باشد) $X_1 \times X_2 \neq X_2 \times X_1$

- حاصلضرب کارتزین

PAPCO

Scanned by CamScanner

Subject:
Date

عدد تعلق است

اگر گویا هستیم
از پرانتز استفاده نمیکنیم

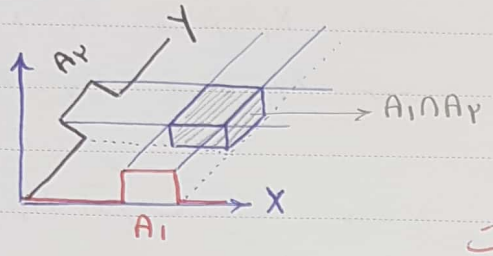
$X=\{\frac{1}{a},\frac{1}{b},\frac{1}{c}\}$ $\quad$ $Y=\{\frac{1}{1},\frac{1}{2},\frac{1}{3}\}$

$A_1=\{\frac{1}{a},\frac{1}{b}\}$ $\quad$ $A_2=\{\frac{1}{2},\frac{1}{3}\}$

- توسعهٔ استوانه ای:

توسعهٔ استوانه ای A_1 روی مرجع Y $=\{\frac{1}{(a,1)},\frac{1}{(a,2)},\frac{1}{(a,3)},\frac{1}{(b,1)},\frac{1}{(b,2)},\frac{1}{(b,3)}\}$

تمام مرجع Y را به A_1 اضافه می‌کنیم.



جدا کننده است
جمع نیست

توسعه استفاده ای A_2 روی X $=\{\frac{1}{a,2}+\frac{1}{b,2}+\frac{1}{c,2}+\frac{1}{a,3}+\frac{1}{b,3}+\frac{1}{c,3}\}$

$\underbrace{A_1\times A_2}_{\text{حاصلضرب کارتزین}}=A_1\cap A_2=\{\frac{1}{a,2}+\frac{1}{a,3}+\frac{1}{b,2}+\frac{1}{b,3}\}$

اشتراک توسعهٔ استوانه ای ۲ مجموعهٔ غیر هم مرجع

$U\times V=\{(u_1,v_1),(u_1,v_2),\dots,(u_n,v_m)\}$

$A_1\times A_2=(A_1\times Y)\cap(X\times A_2)$

- رابطه: زیر مجموعه‌ای از حاصلضرب کارتزین یک مجموعه است

شهودی: هر اسم چه شماره تلفنی دارد، هر اسم چه نمره‌ای در درس هوش محاسباتی دارد، ...

$\Downarrow$

معلومه چرا باید بشه زیر مجموعهٔ حاصلضرب کارتزین، چون حاصلضرب کارتزین تمام حالات مختلف این رابطه ها را در خود دارد

PAPCO

Scanned by CamScanner

Subject:
Date

* رابطه حداقل ۲ بُعدی است. ← معادل نمایش برداری : نمایش ماتریسی (مقادیر: صفر و یک)

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

رابطهٔ کامل $\neq$ رابطهٔ مرجع

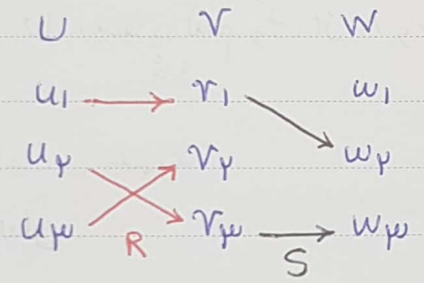
رابطهٔ حاصلضرب کارتزین ۲ مجموعهٔ مرجع $= X \times Y$

رابطهٔ کامل

$R \subseteq A \times B \subseteq X \times Y$

یک رابطه بین A و B

*عملگر ترکیب (∘) :



به ازای تمام اعضای مجموعه وسط،
اگر حداقل یک مسیر وجود داشته باشد از مجموعه اول به مجموعهٔ سوم،
پس یک ارتباط بین مجموعهٔ اول و سوم وجود خواهد داشت.

* کاهش بُعد توسط ابزار تصویر کردن انجام می شود.

$R \circ S = U(R \cap S)$

تصویر کردن است در واقع نه اجتماع

U مجموعه مرجع $= X$
V " $= Y$
W " $= Z$

$\Rightarrow$

R مجموعه مرجع $= X \times Y$
S " $= Y \times Z$

برای اشتراک باید هم مرجع شوند S و R

PAPCO

Scanned by CamScanner

Subject:
Date

اسلاید:

D : Delay

P : Power توان

E : E ... تکنولوژی

T : TTL تکنولوژی

C : CMOS "

* راه میان بُر: max گرفتن بین min های ...

* راه میان بُر به صورت جبری:

$$\mu_{RoS}(u_1, w_2) = \underbrace{\left(\mu_R(u_1, \underset{0}{v_1}) \wedge \mu_S(\underset{0}{v_1}, w_2)\right)}_{0} \vee \underbrace{\left(\mu_R(u_1, \underset{0}{v_2}) \wedge \mu_S(\underset{1}{v_2}, w_2)\right)}_{0} \vee \underbrace{\left(\mu_R(u_1, \underset{1}{v_3}) \wedge \mu_S(\underset{0}{v_3}, w_2)\right)}_{0}$$

تابع تعلق رابطه بین u_1 و w_2

$$0 \vee 0 \vee 0 = 0$$

نکته: کجا از $\cap$ و $\cup$ استفاده کنیم و کجا از $\wedge$ و $\vee$؟

~~AAB =~~

هنگام سر و کار داشتن با اعضای مجموعه
" " " " " " تعلق

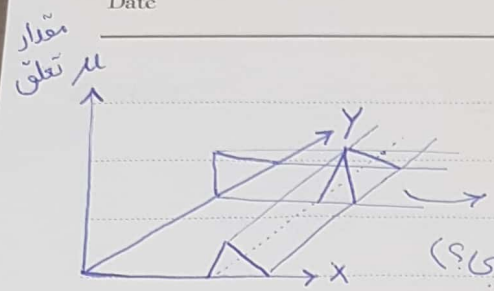
ادامه در ویس

Scanned by CamScanner

Subject:
Date

«جلسه ۲۱»

ضرب کارتزین فازی:



هم $A \times B = (A \times Y) \cap (X \times B)$

(۴ وجهی؟)

$\mu = 1 \leftarrow$ حداکثر تعلق

$\mu = 0 \leftarrow$ حداقل تعلق (تعلقی ندارد)

* "دوری" را چگونه به تعلق نسبت دهیم؟ نقطه مرکز تمام کشورها را به دست می‌آوریم. ۲ به ۲ فاصله مرکز کشورها باهم را حساب می‌کنیم. max فاصله را پیدا می‌کنیم. تمام فاصله‌های به دست آمده را تقسیم بر آن مقدار max می‌کنیم (همواره عددی بین ۰ و ۱ می‌شه) μ

means ← S-norm اجتماع

T-norm اشتراک

max و min و ...

توجیه فیزیکی شون رو بعداً می‌گه ← U, V

∩ و ∪

$$\mu_{R \circ S}(u_1, w_2) = \left(\mu_R(u_1, v_1) \wedge \mu_S(v_1, w_2)\right) \vee \left(\mu_R(u_1, v_2) \wedge \mu_S(v_2, w_2)\right) \vee \left(\mu_R(u_1, v_3) \wedge \mu_S(v_3, w_2)\right)$$

$$\Rightarrow \mu_{R \circ S}(u, w) = \bigvee_{v \in V} \left(\mu_R(u, v) \wedge \mu_S(v, w)\right)$$

ترکیب max-min:

$$= \max_{v} \left(\min\left(\mu_R(u, v), \mu_S(v, w)\right)\right)$$

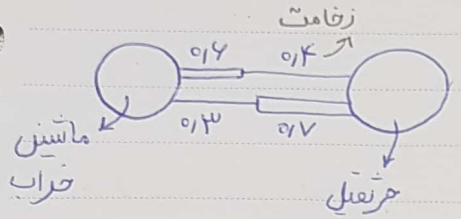
$$R \circ S = \overset{\max}{\cup}\left(\overset{\min}{R \cap S}\right)$$

PAPCO

Scanned by CamScanner

Subject:
Date

* دلیل منطقی که به‌جای ∪ و ∩ می‌گیم max و به ∩ و ∪ می‌گیم min در رابطهٔ صفحه قبل:



→ اگه این طناب‌ها کافی نباشند، اول ۰/۳ ای پاره می‌شه ولی ارتباط هنوزم وجود داره! (تا جایی که max بین min ها هنوز پاره نشده)

اجتماع: max
اشتراک: x
متمم: تعلق - 1

قوانین

آیا در این صورت، مجموعه‌ها را خراب می‌کند؟ (تحقیق کنید!)
آیا در این صورت بازهم یک کلاس normal داریم یا نه؟

که خاصیت شرکت‌پذیری، توزیع‌پذیری، دمورگان و جابه‌جایی را باهم داشته باشند.

* کلاس normal برابرست با اجتماع و اشتراک و متمم

اجتماع : max
اشتراک : min
متمم : 1 - تعلق

$\Rightarrow A' \cap A \neq \emptyset$

$A' \cup A \neq X$

$\forall x \ \mu_{A'}(x) \wedge \mu_A(x) \neq 0 \qquad \forall x \ \min(1-\mu_A(x), \mu_A(x)) \neq 0$

$\forall x \ \mu_{A'}(x) \vee \mu_A(x) \neq 1 \qquad \forall x \ \max(1-\mu_A(x), \mu_A(x)) \neq 1$

ولی اگر به‌جای max از جمع دراستیک و به‌جای min هم از ضرب دراستیک استفاده کنیم، این روابط بالا درست خواهند شد. ($\neq \leftarrow =$)

مثال داخل اسلاید: ۳ ماشین با مدل های V_1 و V_2 و V_3 داریم.
هرسه سرعت ۱۰۰ را می‌توانند بروند... V_1 سرعت ۲۰۰ را نمی‌تواند برود
V_2 را با وام می‌توان خرید. V_3 را با وام هم نمی‌شود خرید

Scanned by CamScanner

Subject:
Date

E : رابطهٔ کامل

O : رابطهٔ تهی

* اجتماع نقاط تعلق ← تصویرمان در آن نقطه

حذف بُعد Y : تصویر R روی X $= \bigvee_{y \in Y} \mu_R(x,y) = \mu_A(x)$

$p(x) = \int_{-\infty}^{+\infty} p(x,y)\,dy$ $\qquad p(x \cap y)$

$p(x,y) = p(x) \cdot p(y)$

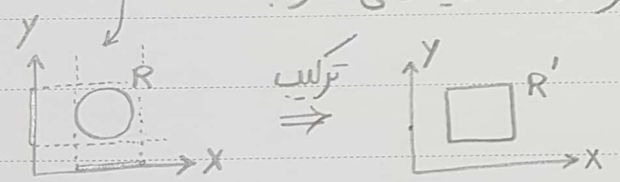
$\mu_R(x,y) = \mu_A(x) \wedge \mu_B(y)$

* رابطهٔ جداپذیر ← بتوان آن را به صورت حاصلضرب کارتزین تصاویر (اشتراک توسعه استوانه‌ای تصاویر) نمایش داد.

← اگر مؤلفه‌ها نسبت به هم وابستگی نداشته باشند، R می‌تواند منحصراً توسط تصاویرش بازسازی شود ← مثلاً در $x^2+y^2=R^2$ (دایره) (x,y) وابستگی دارند

⇓

یعنی ترکیب دوباره‌ی تصویرها، R اولیه را نمی‌سازد!



- ادامهٔ این جلسه:

؟ مشتق می‌خواهیم چی کار؟!

* مشکل (max و min): مشتق‌پذیر نیست ← اما استفاده از (max, ضرب) خیلی جاها مشتق‌پذیر است.

* ضرب بهترین تقریب برای min است.

* در هر نقطه که می‌خواهیم تصویر را حساب کنیم، یک صفحه موازی با صفحه‌ی MoY رسم کنیم. اگر صفحه، آن رابطه را در هیچ نقطه‌ای قطع نکند، تصویرش صفر می‌شود. ← هندسی

در غیر این صورت بین مقدار تعلق نقاط قطع شده اجتماع می‌گیریم.

PAPCO

Scanned by CamScanner

Subject:
Date

تا سر 50:32′ گوش دادم.

Scanned by CamScanner

Subject:
Date

«جلسه ۲۲»

* مهم ترین بخش های تا اینجای فصل: توسعه استوانه ای + تصویر کردن + رابطه جدایپذیر

Summary:

(مثال) $X=\left\{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\right\}$ $Y=\left\{\frac{1}{1}+\frac{1}{2}+\frac{1}{3}\right\}$

$A=\left\{\frac{0.8}{a}+\frac{0.3}{b}\right\}$ $B=\left\{\frac{0.9}{2}+\frac{0.1}{3}\right\}$

توسعه استوانه ای A روی Y $= A\times Y \rightarrow \mu_A(x,y)$

$y\in Y \rightarrow$ به ازای تمام y های Y

$$A\times Y=\left\{\frac{0.8}{a,1}+\frac{0.8}{a,2}+\frac{0.8}{a,3}+\frac{0.3}{b,1}+\frac{0.3}{b,2}+\frac{0.3}{b,3}\right\}$$

مقدار مرهای A و Y را می گذاریم. اشتراک (min)

توسعه استوانه ای B روی X $= X\times B \rightarrow \mu_B(x,y)$

$x\in X$

$$X\times B=\left\{\frac{0.9}{a,2}+\frac{0.9}{b,2}+\frac{0.9}{c,2}+\frac{0.1}{a,3}+\frac{0.1}{b,3}+\frac{0.1}{c,3}\right\}$$

$$(A\times Y)\cap(X\times B)=A\times B=\left\{\frac{0.8}{a,2}+\frac{0.1}{a,3}+\frac{0.3}{b,2}+\frac{0.1}{b,3}\right\}$$

$$A\cup B=\left\{\frac{0.8}{a,1}+\frac{0.9}{a,2}+\frac{0.8}{a,3}+\frac{0.3}{b,1}+\frac{0.9}{b,2}+\frac{0.3}{b,3}+\frac{0.9}{c,2}+\frac{0.1}{c,3}\right\}$$

max می گیریم

این دیگه راه میانبر نداره:

PAPCO

Scanned by CamScanner

Subject:
Date

$A' = X$ تصویر R روی $= \begin{cases} \bigcup_{y \in Y} R \\ \mu_{A'}(x) = \bigvee_{y \in Y} \mu_R(x,y) \end{cases}$

$A' = \left\{ \frac{0.8}{a} + \frac{0.3}{b} \right\}$

$B' = Y$ تصویر R روی $= \begin{cases} \bigcup_{y \in Y} R \\ \mu_{B'}(y) = \bigvee_{x \in X} \mu_R(x,y) \end{cases}$

$B' = \left\{ \frac{0}{1} + \frac{0.8}{2} + \frac{0.1}{3} \right\}$

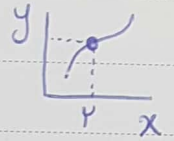
		1 (0)	2 (0.8)	3 (0.1)
0.8	a	✓ 0 / 0 0.8	✓ 0.8 / 0.8 0.8	✓ 0.1 / 0.1 0.8
0.3	b	✓ 0 / 0 0.3	✓ 0.3 / 0.8 0.3	✓ 0.1 / 0.1 0.3
0	c	✓ 0 / 0 0	✓ 0 / 0.8 0	✓ 0 / 0.1 0

$\Rightarrow$ $A' \times B'$ جداپذیر است

(ولی $A \times B$ جداپذیر نیست)

* نکته خیلی مهم:

وقتی یک رابطه بین x و y داریم مانند



و بخواهیم بدانیم y در $x=۲$ چند است،

① ورودی را روی محور خروجی توسعه استوانه‌ای می‌دهیم. $\mu_A(x,y)$ $y \in Y$

② اشتراک توسعه استوانه‌ای ① را با رابطه به دست می‌آوریم (نقطه روی نمودار) $\mu_A(x,y) \wedge \mu_R(x,y)$

③ تصویر اشتراک مرحله قبل را روی محور خروجی حساب می‌کنیم

$$\bigvee_{\substack{x \in X \\ y \in Y}} \mu_A(x,y) \wedge \mu_R(x,y)$$

PAPCO

$$B = \bigcup_{x \in X} (A \times Y) \cap R = \boxed{A \circ R}$$

Scanned by CamScanner

Subject: $R = A \times B$: یکی از ساده ترین حالت ها
Date

* اگر A و B را داشته باشیم، می توانیم R را مدل کنیم.

ورودی : A

خروجی : B

" R و A " " " " B را حساب کنیم $B = A^{\circ} R$

" R و B " " " " " A " $A = R^{\circ} B$

* استفاده از فازی در شبکه عصبی برای مدل کردن A : Voice

• فرض کنیم می دانیم اگر ورودی A باشد، خروجی باید B باشد.
فرض می کنیم به روشی رابطه R سیستم را هم به دست آورده ایم.

پس باید $B = A^{\circ} R$ برقرار باشد. ← ولی همیشه این طور نیست!
چون ممکنه R جدا پذیر نباشه! (اکثر رابطه های مسئله ها جداپذیر نیستند!)

قابل اثبات است. (قضیهٔ استدلال تقریبی)

* اگر A_1 ای داشته باشیم و با R ترکیب کنیم و به خروجی B_1 برسیم، هرچه A_1 را به A نزدیکتر کنیم، B_1 هم به B نزدیک خواهد شد.
وقتی A_1 با A دقیقاً برابر شود، بهترین تقریب برای B به دست خواهد آمد.

* اینکه بهترین R ممکن چه باشد یک مسئله دیگره!

PCA ← پیدا کردن بُعدی که بیشترین واریانس را دارد ← مناسب برای فشرده سازی و بازنمایی

fisher ← " " " " تمایز " " ← مناسب برای classification

PAPCO

Scanned by CamScanner

Subject:
Date

مثال) $A=\left\{\frac{0.1}{1}+\frac{1}{2}+\frac{0.3}{3}\right\}$ → یک عدد تقریباً "2"

$B=\left\{\frac{0.2}{2}+\frac{1}{3}+\frac{0.1}{4}\right\}$ → "3" " " "

خروجی Z را با فرض زیر حساب کنید:

$Z = 2x + 4y^2 = f(x,y)$

$x = A$

$y = B$

$A^* o R = ?$

$$A \times B = \left\{\frac{0.1}{1,2}+\frac{0.1}{1,3}+\frac{0.1}{1,4}+\frac{0.2}{2,2}+\frac{1}{2,3}+\frac{0.1}{2,4}+\frac{0.2}{3,2}+\frac{0.3}{3,3}+\frac{0.1}{3,4}\right\} = A^*$$

باید رابطه‌ای درست کنیم که سطرهای A^* با ستون‌های آن همخوانی داشته باشد

ستون‌های رابطه‌مان می‌شوند y

	18	38	66	20	40	68	22	42	70
1,2	1 0.1	0	0	0	0	0	0	0	0
1,3	0 0.1	1	0	0	0	0	0	0	0
1,4	0 0.1	0	1	0	0	0	0	0	0
2,2	0 0.2	0	0	1	0	0	0	0	0
2,3	0 1	0	0	0	1	0	0	0	0
2,4	0 0.1	0	0	0	0	1	0	0	0
3,2	0 0.2	0	0	0	0	0	1	0	0
3,3	0 0.3	0	0	0	0	0	0	1	0
3,4	0 0.1	0	0	0	0	0	0	0	1
	1 0.1	0.1	0.1	0.2	1	0.1	0.2	0.3	0.1

ممکن بود دو سطر دارای 1 مقدار شوند.

این عدد قرمزها که همه ستون اعداد A^* هستند را برای سطرهای نوشتیم و با عدد آن خانه min کرده و در نهایت بین اعداد ستون‌ها max می‌گیریم.

Scanned by CamScanner

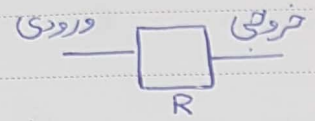
Subject:
Date

« جلسه ۲۳ »

* ورودی ماشین: گاز، ترمز، فرمان

خروجی = : میزان حرکت، میزان چرخش، میزان آلودگی ایجاد شده، میزان بنزین مصرف شده، ...

رابطه: فرمان را ۵ درجه به راست بچرخانیم، میزان فلان خروجی چقدر خواهد بود



تابع: ۲ ورودی یکسان فقط ۱ خروجی می‌توانند تولید کنند.

ورودی سیستم در مثال قبل: A^* ← $A^* = A \times B$ (۲ بُعدی کردن ورودی و هم مرجع کردن ورودی‌ها)

خروجی " " " " : B^*

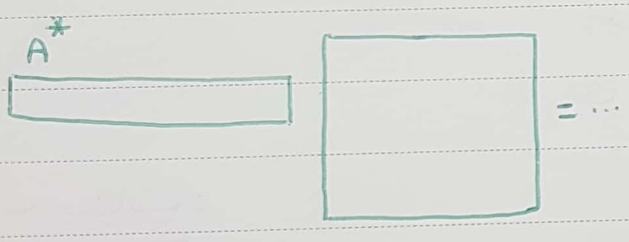
رابطه " " " " : $B^* = A^* \circ R$

↓

$$A^* = \left\{ \frac{0.1}{1,2} + \frac{0.1}{1,3} + \frac{0.1}{1,4} + \frac{0.2}{2,2} + \frac{1}{2,3} + \frac{0,1}{2,4} + \frac{0.2}{3,2} + \frac{0.3}{3,3} + \frac{0.1}{3,4} \right\}$$

ترکیبی که انجام می‌دهیم، ستون‌ها و سطرهایش باید یک جنس باشند. (؟)

ضرب ماتریسی معمولی انجام می‌دهیم فقط به جای ضرب min و به جای جمع max می‌گیریم.



این روش را وقتی A و B پیوسته هستند نمی‌توان انجام داد چون تعداد سطرهای ماتریس ∞ خواهد شد

$$B^* = \left\{ \frac{0.1}{18} + \frac{0.2}{20} + \frac{0.2}{22} + \frac{0.1}{38} + \frac{1}{40} + \frac{0.2}{42} + \frac{0.1}{66} + \frac{0.1}{68} + \frac{0.1}{70} \right\}$$

↓

خروجی عدد تقریباً $\aleph_0$ است =)

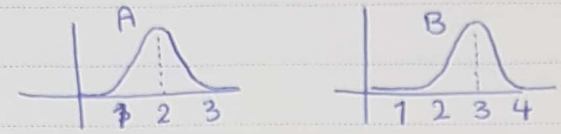
$$2(2) + 4(3)^2 = \aleph_0$$

PAPCO

Scanned by CamScanner

Subject:
Date

- در حالت پیوسته: ① اشانتیون تعیین می‌کنیم یعنی یک مقیاس در نظر می‌گیریم. مثلاً ۰٫۰۱
پس ۲۰۰ داده برای A و
۲۰۰ داده برای B داریم ...



در مورد مجموعه‌های بزرگ این روش هزینه‌بر است.

② مقادیر تعلق را گسسته کنیم. (برش λ)

$A_\lambda = \{x \mid \mu(x) \geq \lambda\}$ یک مجموعه غیرفازی

* می‌توان از روی برش‌های λ مجموعه فازی را بازسازی کرد!

$C = \bigcup C^\lambda$

قانون $\underbrace{C_\lambda}_{(A\star B)_\lambda} = (A_\lambda \star B_\lambda)$

$(A\star B)_\lambda$ → مورد نظر

$A_\lambda \star B_\lambda$ → عملیات

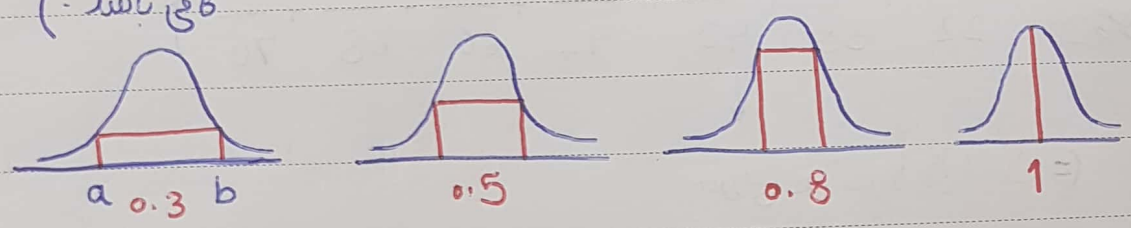
پس در مورد پیوسته اول روی A و B برش λ می‌زنیم و بعد عملیات را انجام می‌دهیم که البته عملیات ×، +، − و ÷ روی برش‌های λ عملیات متفاوت ولی ساده‌ای هستند.

- ترتیب عملیات:

۱- عملیات جبری روی برش‌های لاندا در حالت پیوسته به سادگی قابل انجام است.

۲- $C_\lambda = (A\star B)_\lambda = A_\lambda \star B_\lambda$

۳- از روی برش‌های لاندای C، مجموعه فازی C قابل بازسازی است. (اگر تعداد شون به اندازهٔ کافی باشد.)



$C_{0.3} = [a\ b]$ $\quad C^{0.3} = \left[\frac{0.3}{a, b}\right]$

C^λ: همان C_λ است که

PAPCO

Scanned by CamScanner

Subject:
Date

- منطق فازی: هر گزاره را به صورت تعلق داشتن یک عضو به یک مجموعه نگاه می‌کنیم و اگر $a \in A$ برقرار باشد، گزاره درست است و اگر $a \in A$ برقرار نباشد، گزاره غلط است.

مثلاً:

علی برادر حسن است.
$A =$ برادران حسن
$a =$ علی

$\Rightarrow a \in A \quad \mu_A(a)$

در: $p \wedge q$
منطق ↓ ↓ ↓ معادل
در: $A \cap B \Rightarrow \mu_A(x) \wedge \mu_B(y)$
فاز
مجموعه‌های فازی

پس معادل هر گزاره یک مجموعه پیدا می‌کنیم
پس منطق چیز جدیدی نیست.

استنتاج: $B = A \circ R$
← معادل rule های منطق رابطه به دست آوریم
(قانون)

قانون‌های منطق:

$p \Rightarrow q \equiv p' \vee q \Rightarrow A' \cup B$

$p \Leftrightarrow q \equiv (p' \vee q) \wedge (p \vee q')$

«جلسه ۲۴»

~~FCM Fuzzy K-means~~ ~~K-means~~

FCM: (یادآوری CM: ص آخر جلسه ۱۶)

تعداد خوشه‌ها

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

PAPCO

Scanned by CamScanner

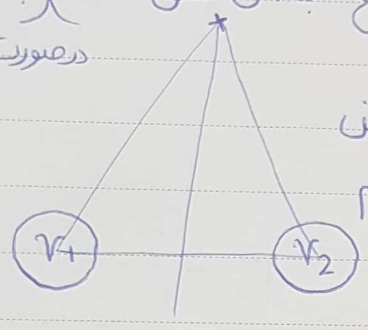
Subject:
Date

در این جا E مشتق پذیر چون تعلق فقط ۰ و ۱ ای نیست.

$$\frac{\partial E}{\partial u_{ik}} = m\, u_{ik}^{m-1} \|x^k - v_i\|^2 = 0$$

$u_{ik} = 0 \longrightarrow$ نباید این اتفاق بیفتد $\Rightarrow$ محدودیت: $\sum_{i=1}^{c} u_{ik} = 1$

مشکل این روش محدود کردن: فاصله‌ی مطلق، بسنی می‌شود. داده ای که out layer و noise است و به هیچ دسته ای تعلق ندارد، به هر دو دسته تعلق ۰٫۵ می دهد: درصورت فاصله برابر



و درضمن مراکز داده‌ها را هم عوض می کند!

$\Rightarrow$ ابتدا out layer ها را حذف کنیم و بعد محدودیت را اعمال کنیم

روش ها: ① واریانس ...

?

② دو به دو بین داده ها میانگین می گیریم ...

خواب بودم :)

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

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

PAPCO

Scanned by CamScanner

Subject:
Date

$$\frac{\partial E}{\partial u_{ik}} = m\, u_{ik}^{m-1} \| x^k - v_i \|^2 + \lambda_k = 0 \quad (۲)$$

$$\frac{\partial E}{\partial v_i} = \sum_{k=1}^{N} -2\, u_{ik}^{m} (x^k - v_i) = 0 \quad (۳)$$

$$\sum_{k=1}^{N} u_{ik}^{m} x^k = v_i \sum_{k=1}^{N} u_{ik}^{m} \Longrightarrow \boxed{v_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} x^k}{\sum_{k=1}^{N} u_{ik}^{m}}} \quad \text{II}$$

$$(۱) \Rightarrow \sum_{j=1}^{c} \left(\frac{-\lambda_k}{m}\right)^{\frac{1}{m-1}} \frac{1}{\| x^k - v_j \|^{\frac{2}{m-1}}} = 1 \Rightarrow \boxed{1 = \left(\frac{-\lambda_k}{m}\right)^{\frac{1}{m-1}} \sum_{j=1}^{c} \frac{1}{\| x^k - v_j \|^{\frac{2}{m-1}}}} \quad (۴)$$

$$(۲) \Rightarrow \boxed{u_{ik} = \left(\frac{-\lambda_k}{m}\right)^{\frac{1}{m-1}} \frac{1}{\| x^k - v_i \|^{\frac{2}{m-1}}}} \quad (۵)$$

$p_i = \alpha f_i$ $\qquad u_{ik} \propto S_{ik}$ $\qquad u_{ik} = k S_{ik} = \dfrac{S_{ik}}{\sum_{j=1}^{c} S_{jk}}$ $\qquad$ I

$p_i = \dfrac{f_i}{\sum f_j}$

$$k = \frac{1}{\sum_{j=1}^{c} S_{jk}}$$

$$\boxed{u_{ik} = \frac{\dfrac{1}{\| x_k - v_i \|^{\frac{2}{m-1}}}}{\sum_{j=1}^{c} \dfrac{1}{\| x_k - v_j \|^{\frac{2}{m-1}}}}}$$

معمولا: $1 < m < 2$ ⟶ نمایانگر چیزی از فازی بودن

هرچه نزدیکتر به ۱ : فازی تر

" " بزرگتر از ۱ : ۰ و ۱ ای تر

تعبیر این فرمول: (۶)

Scanned by CamScanner

Subject:
Date

جواب هایمان به هم وابسته اند ⇐ باید از یک نقطه شروع کنیم و مقدار اولیه بدهیم
⇐ مشکل min محلی + مقداردهی اولیه را دارد (مانند K means)
FCM

* مراکز FCM با Kmeans متفاوت اند و در FCM مراکز بهتر اند زیرا:
داده های مرزی بین ۲ دسته در FCM می توانند با مقداری تعلق، به هر دو دسته تعلق داشته باشند و به همین دلیل مراکز واقعی تر خواهند بود

- سیستم های خبره:

* قوانین منطق تفسیرهای مختلفی دارند.
* منطق می گوید اگر شرط برقرار نبود، هر اتفاقی ممکن است بیفتد.
فلسفه و

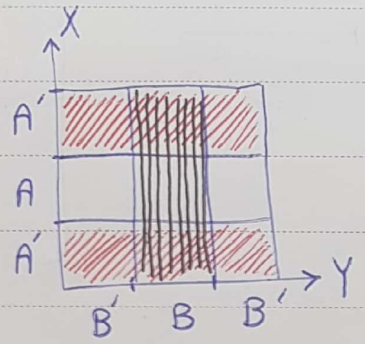
اما در واقعیت نه. مثلاً: اگر هوا سرد است، لباس گرم بپوش.
نمی شود اگر هوا سرد نبود، هر کار خواستیم بکنیم.

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

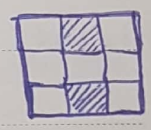
وقتی شرط برقرار نباشد، هر اتفاقی ممکنه (یعنی A') بیفته (یعنی Y)

$P \Rightarrow Q$

$R = A' \cup B$



→ با شکل داخل اسلاید یکی شد → اما در اینجا



این دو خانه ۲ بار در اجتماع آمده اند

if $P \cup P = P \Rightarrow \dots$? (Voice +)

(+ جزوه مهسان)

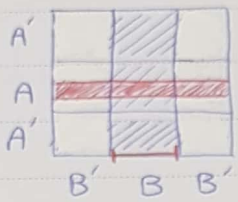
PAPCO

Scanned by CamScanner

Subject:
Date

«جلسه ۲۴»

$A_1 \subseteq A \Rightarrow B_1 = A_1 \circ R = B$



ادامه در اسلاید

- $P \Rightarrow Q \equiv A \times B \equiv A' \cup B \equiv (A \times B) \cup (A' \times Y)$

مجموعه $A \Rightarrow B$

در حالت کلی

از این فرمول

استفاده می‌کنیم

چون اگر شرط برقرار نباشد هیچکاری

نمی‌کند نه اینکه هرکاری دلش بخواهد

بکند!

$$\mu_{A \Rightarrow B}(x,y) = \min(\mu_A(x), \mu_B(y)) = \max(1-\mu_A(x), \mu_B(y)) = \max(\min(\mu_A(x), \mu_B(y)), 1-\mu_A(x))$$

بهترین حالت

$\Downarrow$

$$\mu_{A \Rightarrow B}(x,y) = \mu_A(x) \cdot \mu_B(y)$$

PAPCO

Scanned by CamScanner

Subject:
Date

- $A \Rightarrow B$ else C

$$\mu_R(x,y) = \text{Max}\left(\text{Min}\left(\mu_A(x), \mu_B(y)\right), \text{Min}\left(1-\mu_A(x), \mu_C(y)\right)\right)$$

؟ اگر چند قانون داشتیم، بستگی دارد هر قانون چقدر fire شود.
قانون ها باید کل فضا را پوشش دهند.

x_2 A_2 — y B — و خیلی کثیف — x_1 A_1

۱ قانون: اگر لباس ها زیاد باشند، آب باید خیلی داغ باشد.
اگر لباس ها نسبتاً زیاد باشند و کثیف،

A_2^* نسبتاً — A_1^*

ورودی $= A^*$

$$\boxed{B^* = A^* \circ R}$$

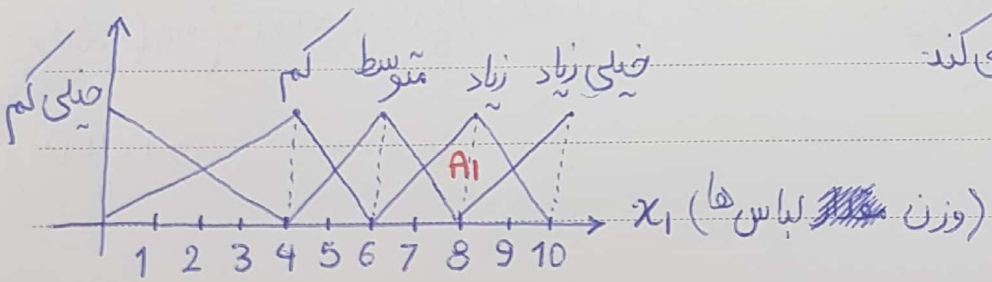
* در اینجا از روی قانون داده شده، رابطه‌ی سیستم (R) را به دست می‌آوریم
(روش های دیگر به دست آوردن R در مسائل دیگر: از روی تابع، با خوشه بندی ...)

$A_1 \times A_2 \Rightarrow B$

$R = A_1 \times A_2 \times B$

$$\mu_R(x_1, x_2, y) = \text{Min}\left(\mu_{A_1}(x_1), \mu_{A_2}(x_2), \mu_B(y)\right)$$

سیستم خبره این سه نمودار را تعریف می‌کند



؟ یک سنسور میزان غلظت آب را حساب می‌کند بعد ...

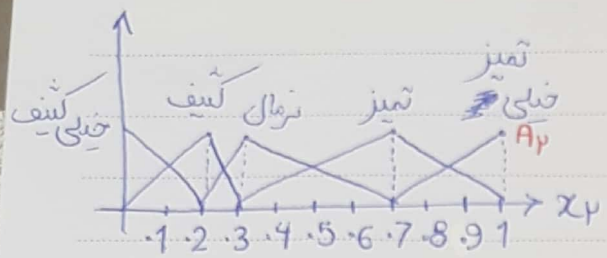
A2D → Analog To Digital

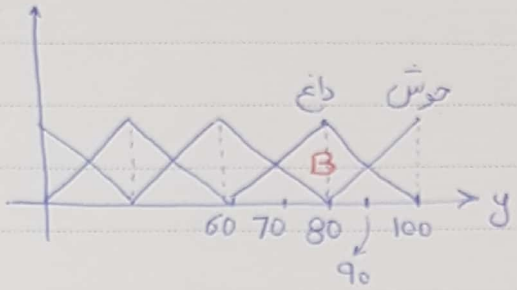
بعد عدد تولید شده توسط A2D رو تقسیم کنیم بر max مقداری که می‌تواند تولید کند ← یک عدد بین ۰ و ۱.

PAPCO

Scanned by CamScanner

Subject:
Date

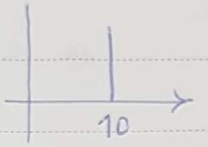




$$A_1^* = \left\{ \frac{0.2}{6} + \frac{0.8}{7} + \frac{0.1}{8} \right\}$$

$$\mu_A(x) = \begin{cases} 1 & x = 10 \\ 0 & x \neq 10 \end{cases}$$

لباس ها تقریباً 7 kg هستند.



ورودی غ فازی ← فازی ← رابطه ← خروجی فازی ← خروجی غیرفازی

ورودی می‌دهیم به

باید تبدیل شود — system — باید تبدیل شود

← آن جا که بیشترین مقدار تعلق را دارد می‌گیریم (این ساده‌ترین روشه)

که یک مقدار فازی است

A_1^* و A_2^* را سنسور به ما می‌دهد و ورودی هایمان هست و ما تعیین شان نکرده‌ایم!

$$A_2^* = \left\{ \frac{0.3}{0.7} + \frac{0.9}{0.8} \right\}$$

اینجا سنبا کثیف تقریبا روی 0.8 عه درحالی که روش سیستم خبره برای کثیف 0.2 در نظر گرفته شده!

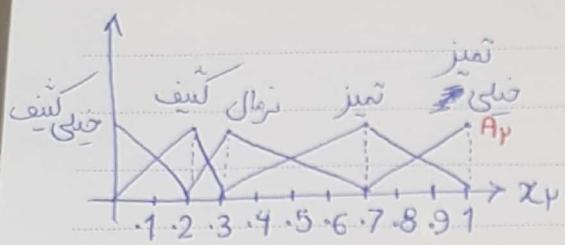
این ایراد سیستم خبره است که باید منطبق باشد با تعریف ورودی از میزان کثیفی!

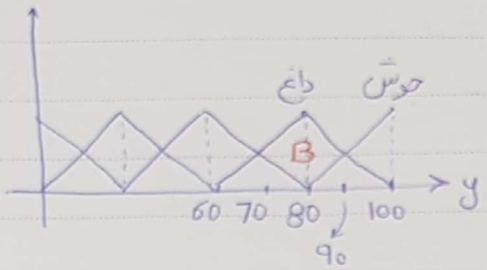
$$A_1^* \times A_2^* = \left\{ \frac{0.2}{6,0.7} + \frac{0.2}{6,0.8} + \frac{0.3}{7,0.7} + \frac{0.8}{7,0.8} + \frac{0.1}{8,0.7} + \frac{0.1}{8,0.8} \right\}$$

P4PCO

Scanned by CamScanner

Subject:
Date

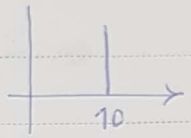




$$A_1^* = \left\{ \frac{0.2}{6} + \frac{0.8}{7} + \frac{0.1}{8} \right\}$$

$$\mu_A(x) = \begin{cases} 1 & x = 10 \\ 0 & x \neq 10 \end{cases}$$

لباس ها تقریباً 7 Kg هستند.



ورودی غ فازی ← فازی ← رابطه ← خروجی فازی ← خروجی غیر فازی

ورودی می‌دهیم به

باید تبدیل شود — system — باید تبدیل شود

← آن جا که بیشترین مقدار تعلق را دارد می‌گیریم (این ساده‌ترین روشه)

که یک مقدار فازی است

A_1^* و A_2^* را سنسور به ما می‌دهد و ورودی هایمان هست و ما تعیین شان نکرده‌ایم!

$$A_2^* = \left\{ \frac{0.3}{0.7} + \frac{0.9}{0.8} \right\}$$

اینجا سنسور کثیف تقریبا روی 0.8 عه درحالی که روش سیستم خبره برای کثیف 0.2 در نظر گرفته شده!

این ایراد سیستم خبره است که باید منطبق باشد با تعریف ورودی از میزان کثیفی!

$$A_1^* \times A_2^* = \left\{ \frac{0.2}{6,0.7} + \frac{0.2}{6,0.8} + \frac{0.3}{7,0.7} + \frac{0.8}{7,0.8} + \frac{0.1}{8,0.7} + \frac{0.1}{8,0.8} \right\}$$

PAPCO

Scanned by CamScanner

Subject:
Date

	70	80	90
6, 0.7	0	0	0
6, 0.8	0	0	0
7, 0.7	0	0	0
7, 0.8	0	0	0
8, 0.7	0	0	0
8, 0.8	1/3	1/3	1/3
	0.1	0.1	0.1

→ تعلق های داغ بودن

این قانون برای این ورودی نمی‌تواند جوابی بدهد.

اگر $A_1^* = \left\{ \frac{0.8}{8} \right\}$, $A_2^* = \left\{ \frac{1}{0.9} \right\}$

	70	80	90
8, 0.9	1/2	2/3	1/2

$$B = \left\{ \frac{0.5}{70} + \frac{0.75}{80} + \frac{0.5}{90} \right\}$$

«جلسه ۲۶»

یک قانون استاندارد فازی:

$$\text{if } \underbrace{x_1 = A_1 \ \& \ x_2 = A_2 \ \dots \ x_n = A_n}_{\text{مجموعه های فازی معادل هر بُعد را می دانیم.}} \text{ then } y = B$$

به طور معادل داریم:

$$A_1 \times A_2 \times \dots \times A_n \Rightarrow B$$

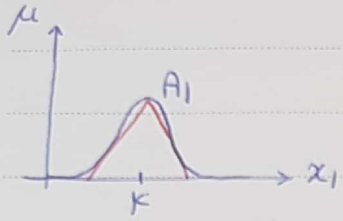
(در حالت یک بُعدی داشتیم: $A \Rightarrow B$)

در حالت ۲ بُعدی:

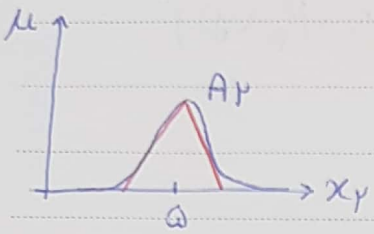
$$\underset{x_1}{A_1} \times \underset{x_2}{A_2} \Rightarrow B$$

PAPCO

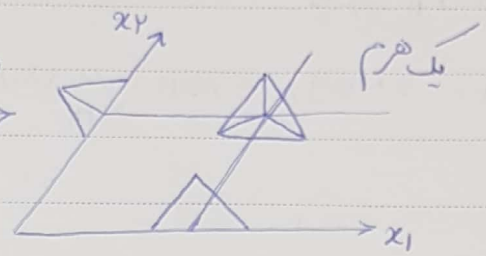
Scanned by CamScanner



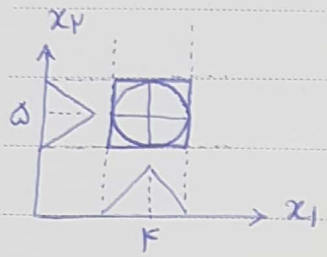
در ۲ حالت بررسی می‌کنیم یکی گوسی یکی مثلثی:



در حالت مثلثی



در حالت گوسی گوشه‌های هرم دیگر تیز نبود.

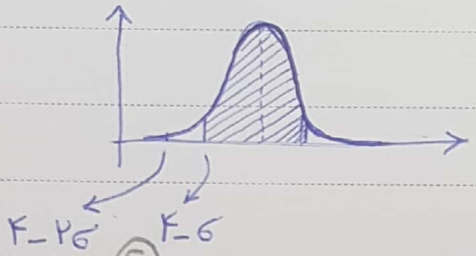


در حالت مثلثی این قانون برای $A_1 \times A_2$ فضایی که پوشش می‌دهد، یک مستطیل (یا مربع) است.

$$e^{-((x_1-4)^2+(x_2-5)^2)} = 0.63 \longrightarrow -\ln 0.63 = (x_1-4)^2+(x_2-5)^2 \Rightarrow$$

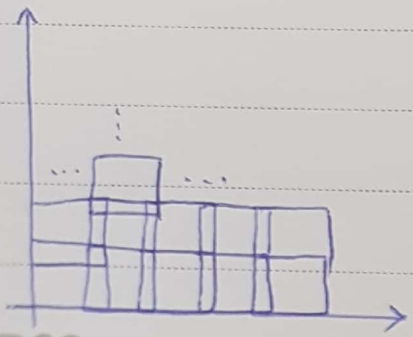
فرمول یک دایره

در حالت گوسی فضایی که این قانون پوشش می‌دهد، یک دایره است.

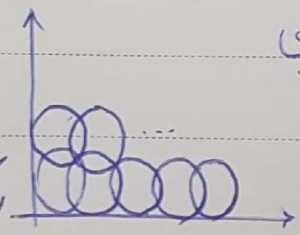


؟

این قانون حداکثر تا محدوده‌ی σ۱ در حالت گوسی جواب می‌دهد. ؟



یک سری مربع با ضلع برابر که overlap می‌کنند



* افراز فضا به یک سری ورودی‌ی خوشه که با هم overlap دارند و شعاع یکسان

استفاده از قانون برای هر قسمت:

Scanned by CamScanner

Subject:
Date

توجیه : $A_1 \times A_2 \times B$ ؟

① سیستم خبره بلد قوانین چه باشند
② اگر data داریم ، ورودی‌هایشان را خوشه بندی کنیم قوانین فازی را به دست آوریم و...

۲ راه وجود داره ،

$A_1 \times A_2 \times \dots \times A_n \Rightarrow B$ $\qquad$ $R = A_1 \times \dots \times A_n \times B$

$$\mu_R(x_1, x_2, \dots, x_n, y) = \text{Min}\left(\mu_{A_1}(x_1), \mu_{A_2}(x_2), \dots, \mu_{A_n}(x_n), \mu_B(y)\right)$$

$x_1 = 4$

$x_2 = 2$

$x_n = 8$

$$\mu_{A_1^*}(x_1) = \begin{cases} 1 & x_1 = 4 \\ 0 & x_1 \neq 4 \end{cases}$$

$$\mu_{A_2^*}(x_2) = \begin{cases} 1 & x_2 = 2 \\ 0 & x_2 \neq 2 \end{cases}$$

$$\mu_{A_n^*}(x_n) = \begin{cases} 1 & x_n = 8 \\ 0 & x_n \neq 8 \end{cases}$$

$\mu_{A_3^*}(x_3) = 1$

$\vdots$

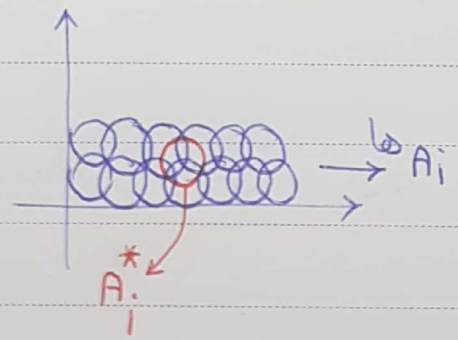
$\mu_{A_{n-1}^*}(x_{n-1}) = 1$

$$B^* = \underbrace{(A_1^* \times \dots \times A_n^*)}_{A^*} \circ R = A^* \circ R = \bigcup_{x \in X} (A^* \circ R) \longrightarrow$$

؟ اجتماع گیری نیازی نیست چون فقط در 1 نقطه اشتراک دارند.

A_i^* → مجموعه فازی ورودی

A_i → مجموعه فازی قانون



PAPCO

Scanned by CamScanner

Subject:
Date

$$\mu_B(y) = Max\left(Min\left(\mu_{A^*}(x), \mu_R(x,y)\right)\right) = Min\left(\mu_{A^*}(x^*), \mu_R(x^*,y)\right)$$

$$Min\left(\mu_{A^*}(x^*), \mu_R(x^*,y)\right) \longleftarrow \mu_{A^*}(x) = \begin{cases} 1 & x = x^* \\ 0 & x \neq x^* \end{cases}$$

$$0$$

$$\mu_{B^*}(y) = Min\left(\mu_{A_1^*}(x_1^*), \dots, \mu_{A_n^*}(x_n^*), \mu_R(x_1^*, \dots, x_n^*, y)\right) = \dots = \mu_{B^*}(y)$$

ورودی یک نقطه باشد و یک قانون داشته باشیم: $\mu_{B^*}(y) = Min\left(\mu_{A_1}(x_1^*), \dots, \mu_{A_n}(x_n^*), \mu_B(y)\right)$

* درصورتی می توان رابطهٔ کلی برای سیستم به دست آورد که بین قوانین اجتماع باشد. (اگر بین شان اشتراک باشد نمی شه گفت.)

$$Min\left(\mu_{A_1^*}(x_1^*), \dots, \mu_{A_n^*}(x_n^*), \mu_{A_1}(x_1^*), \dots, \mu_{A_n}(x_n^*), \mu_B(y)\right)$$

$$R = R_1 \cup R_2 \cup \dots \cup R_m$$

$$B^* = B_1^* \cup B_2^* \dots \cup B_m^* = \bigcup_{i=1}^{m} (A^* \circ R_i) = \bigcup_{i=1}^{m} \left(\bigcup_{x \in X} (A^* \cap R_i)\right) = \bigcup_{x \in X} \left(A^* \cap \overbrace{\left(\bigcup_{i=1}^{m} R_i\right)}^{R}\right)$$

$$= \bigcup_{x \in X} (A^* \cap R) = A^* \circ R$$

$$B^* = B_1^* \cap B_2^* \cap \dots \cap B_m^* = \bigcap_{i=1}^{m} (A^* \circ R_i)$$

$A^* \quad R_1 \rightarrow B_1^* = A^* \circ R_1$

$\vdots$

$A^* \quad R_i \rightarrow B_i^* = A^* \circ R_i$

$\vdots$

$A^* \quad R_m \rightarrow B_m^* = A^* \circ R_m$

PAPCO

Scanned by CamScanner

Subject:
Date

$$\mu_{B^*}(y) = \max_{i=1}^{m}\left(\min\left(\mu_{A_1}(x_1^*), \dots, \mu_{A_n}(x_n^*), \mu_{B_i}(y)\right)\right)$$

$$\mu_{B^*}(y) = \min\left(\mu_{A_1}(x_1^*), \dots, \mu_{A_n}(x_n^*), \max\left(\mu_{B_i}(y)\right)\right) \quad ?$$

از این حالت خاص در امتحان سوال میاد معمولاً!

« جلسه ۲۷ »

- قوانین XOR:

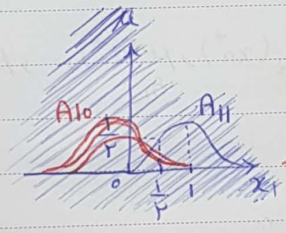
	A_{10}	A_{20}	B_0
۱-	اگر x_1 تقریباً صفر	و x_2 تقریباً صفر	باشد، آن‌گاه y تقریباً صفر است
۲-	" "	" " صفر	" " یک "
۳-	" "	" " یک	" " یک "
۴-	یک	یک	صفر
	A_{11}	A_{21}	B_1

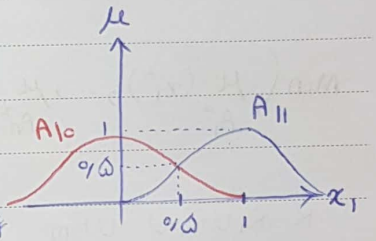
$A_{10} \times A_{20} \Rightarrow B_0 \quad : \quad R_1$

$A_{11} \times A_{20} \Rightarrow B_1 \quad : \quad R_2$

$A_{10} \times A_{21} \Rightarrow B_1 \quad : \quad R_3$

$A_{11} \times A_{21} \Rightarrow B_0 \quad : \quad R_4$





$$\boxed{\mu_{A_{11}}(x_1) = e^{-(4\ln 2)(x_1-1)^2}}$$

$$\boxed{\mu_{A_0}(x_1) = e^{-(4\ln 2)x_1^2}}$$

برای μ_{A21} و μ_{A20} هم همین طوری

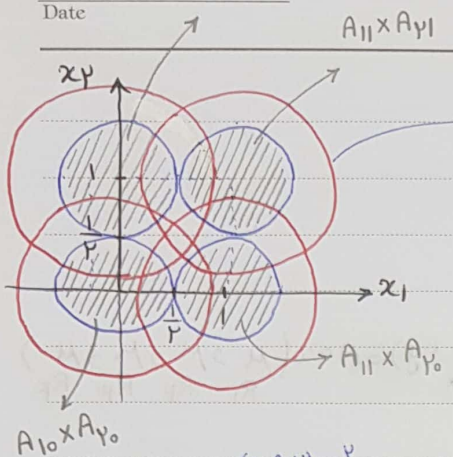
$$\mu_{A_{11}}(0.5) = 0.5 \rightarrow 0.5 = e^{-\gamma/4} \rightarrow \ln 2 = \frac{\gamma}{4} \Rightarrow \gamma = 4\ln 2$$

PAPCO

Scanned by CamScanner

Subject: $A_{10} \times A_{21}$

Date



$$0{,}1 = e^{-(4\ln 2)x_1^2} \qquad x_1^2 = \frac{\ln 10}{4\ln 2} \Rightarrow \boxed{x = 0{,}9}$$

$$+\ln 10 = (+4\ln 2)\, x_1^2$$

$$\mu_{B_0}(y) = e^{-(4\ln 2)y^2} \qquad \mu_{B_1}(y) = e^{-(4\ln 2)(y-1)^2}$$

$$\mu_{R_1}(x_1, x_2, y) = \mu_{A_{10}}(x_1)\,\mu_{A_{20}}(x_2)\,\mu_{B_0}(y) = e^{-(4\ln 2)(x_1^2 + x_2^2 + y^2)}$$

$$\mu_{R_2}(x_1, x_2, y) = \mu_{A_{11}}(x_1)\,\mu_{A_{20}}(x_2)\,\mu_{B_1}(y) = e^{-(4\ln 2)\left((x_1-1)^2 + x_2^2 + (y-1)^2\right)}$$

$$\mu_{R_3}(x_1, x_2, y) = \mu_{A_{10}}(x_1)\,\mu_{A_{21}}(x_2)\,\mu_{B_1}(y) = e^{-(4\ln 2)\left(x_1^2 + (x_2-1)^2 + (y-1)^2\right)}$$

$$\mu_{R_4}(x_1, x_2, y) = \mu_{A_{11}}(x_1)\,\mu_{A_{21}}(x_2)\,\mu_{B_0}(y) = e^{-(4\ln 2)\left((x_1-1)^2 + (x_2-1)^2 + y^2\right)}$$

* به ازای ورودی $x_1 = x_1^* = 0{,}8$ و $x_2 = x_2^* = 0{,}2$ خروجی چه می‌شود؟

$$\mu_{A_1^*}(x_1) = \begin{cases} 1 & x_1 = x_1^* = 0{,}8 \\ 0 & x_1 \neq x_1^* \end{cases}$$

$$\mu_{A_2^*}(x_2) = \begin{cases} 1 & x_2 = x_2^* = 0{,}2 \\ 0 & x_2 \neq x_2^* \end{cases}$$

$$B_1^* = A^* \circ R_1 = (A_1^* \times A_2^*) \circ R_1$$

$$\mu_{B_1^*}(y) = \max\left(\mu_{A_1^*}(x_1)\,\mu_{A_2^*}(x_2)\,\mu_{R_1}(x_1, x_2, y)\right)$$

$\mu_{A_1^*}(x_1^*) \to 1$

$\mu_{A_2^*}(x_2^*) \to 1$

$\mu_{R_1}(x_1^*, x_2^*, y)$ حالت اول

خروجی فازی وقتی قانون اول را داریم و ورودی A^* است. برای رابطه‌ی R_1

PAPCO

Scanned by CamScanner

Subject:
Date

حالت دوم: $\mu_{A_1^*}(x_1 \neq x_1^*)\ \mu_{A_2^*}(x_2 \neq x_2^*)\ \mu_R(\dots)$

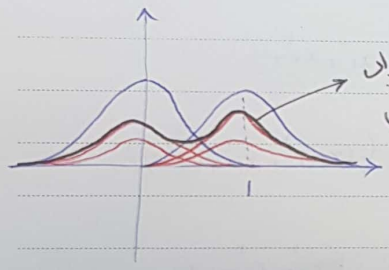
$\mu_{A_1^*} \to 0$ ، $\mu_{A_2^*} \to 0$

بین Max حالت اول و دوم $\Rightarrow \mu_{B_1^*}(y) = \mu_{R_1}(x_1^*, x_2^*, y)$

$\mu_{B_2^*}(y) = \mu_{R_2}(x_1^*, x_2^*, y)$

$\vdots$

$$\Rightarrow \mu_{B^*}(y) = \mathrm{Max}(\mu_{R_1}, \mu_{R_2}, \mu_{R_3}, \mu_{R_4})$$



توان نهایی $\xrightarrow{\text{غیر فازی}}$ روش: $y^* = \dfrac{0\times\mu_1 + 0\times\mu_2 + 1\times\mu_3 + 1\times\mu_4}{\mu_1+\mu_2+\mu_3+\mu_4}$

$$\mu_{R_1}(x_1^*, x_2^*, y) = e^{-(4\ln 2)(0.8)^2} \times e^{-(4\ln 2)(0.2)^2} \times \underbrace{e^{-(4\ln 2)y^2}}_{\mu_{B_0}(y)} = 0.15\,\mu_{B_0}(y) = \mu_{B_1^*}(y)$$

$$\mu_{R_2}(x_1^*, x_2^*, y) = 0.8\,\mu_{B_1}(y) = \mu_{B_2^*}(y)$$

$$\mu_{R_3}(x_1^*, x_2^*, y) = 0.02\,\mu_{B_1}(y) = \mu_{B_3^*}(y)$$

$$\mu_{R_4}(x_1^*, x_2^*, y) = 0.15\,\mu_{B_0}(y) = \mu_{B_4^*}(y)$$

- روش فازی سازی متوسط وزنی مراکز: خروجی = اجتماع یک سری مجموع فازی

PAPCO

Scanned by CamScanner

Subject:
Date

$$\frac{\sum_{i=1}^{m} \mu_i(c_i) \times c_i}{\sum_{i=1}^{m} \mu_i(c_i)} = \frac{0.15\times 0 + 0.8\times 1 + 0.02\times 1 + 0.15\times 0}{0.15 + 0.8 + 0.02 + 0.15} = 0.73$$

۱- اشتراک ← ضرب
۲- max دار
} شرط های فرض شده در این مسئله

$$F(x_1^*, x_2^*) = \frac{\sum_{i=1}^{NR} \mu_{c_{1i}}(x_1^*) \times \mu_{c_{2i}}(x_2^*)\, \bar{y}_i}{\sum_{i=1}^{NR} \mu_{c_{1i}}(x_1^*)\, \mu_{c_{2i}}(x_2^*)}$$

$\Leftarrow$ فرض کنیم اندیس ها این شکلی بودند:

$c_{11} = A_{10}$
$c_{21} = A_{20}$
$c_{12} = A_{11}$
$c_{22} = A_{22}$

- خلاصه‌ی فرض های این مسئله:

۰- فازی سازی ورودی به صورت عدد دقیق
۱- قوانین به این شکل تعریف شوند:
۲- رابطه " " " " " :

$$R_i \quad \underbrace{A_1^i \times A_2^i \times \cdots \times A_n^i}_{=A^i} \Rightarrow B^i$$

$$\begin{cases} R_i = A^i \times B^i \\ \mu_{R_i}(x_1, \ldots, x_n, y) = \mu_{A_1^i}(x_1) \cdots \mu_{A_n^i}(x_n)\, \mu_{B^i}(y) \end{cases}$$

۳- ترکیب به صورت max dot
۴- استفاده از روش غیر فازی سازی متوسط وزنی مراکز

* در این مثال از قبل رابطه را مستقیم می‌شد حساب کرد: $y = x_1\bar{x}_2 + \bar{x}_1 x_2$

PAPCO

Scanned by CamScanner

Subject:
Date

$$\left(\mu_{A_1}(x_1) \wedge \mu_{A'_2}(x_2)\right) \vee \left(\mu_{A'_1}(x_1) \wedge \mu_{A_2}(x_2)\right) \Rightarrow \mathrm{Max}\Big(\mathrm{Min}\big(\overset{0,8}{\mu_{A_1}(x_1)}, 1-\overset{0,2}{\mu_{A_2}(x_2)}\big),$$

$$\mathrm{Min}\big(1-\mu_{A_1}(x), \mu_{A_2}(x_2)\big)$$

$$= 0,8$$

(اگر هم به جای Min ضرب بگیریم : جواب = ۰٫۶۴)

PAPCO

Scanned by CamScanner