

شبکه‌های عصبی مصنوعی

۱۳۹۵-۱۱-۰۷

بخش هشتم



مقدمه‌ای به یادگیری عمیق
DNN

دانشگاه شهید بهشتی

دانشکده‌ی علوم و مهندسی کامپیوتر

بهار ۱۳۹۵

احمد محمودی ازناوه

فهرست مطالب

- مقدمه‌ای بر یادگیری عمیق
- یادگیری خودآموقته
- یادگیری سلسله مراتبی فصلمند
- آشنایی با CNN
- مدل‌سازی شبکه‌های عمیق



دانشکده
پژوهشی

معرفی

<https://www.macs.hw.ac.uk/~dwcorne/index.htm>

- «یادگیری عمیق» باعث ایجاد پیشرفت‌های چشمگیری در زمینه بینایی ماشین و پردازش صوت شده است.

<http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/38131.pdf>

modeling technique	#params [10 ⁶]	WER		task	hours of training data	DNN-HMM	GMM-HMM with same data	GMM-HMM with more data
		Hub5'00-SWB	RT03S-FSH					
GMM, 40 mix DT 309h SI	29.4	23.6	27.4	Switchboard (test set 1)	309	18.5	27.4	18.6 (2000 hrs)
NN 1 hidden-layer × 4634 units + 2 × 5 neighboring frames	43.6 45.1	26.0 22.4	29.4 25.7	Switchboard (test set 2)	309	16.1	23.6	17.1 (2000 hrs)
DBN-DNN 7 hidden layers × 2048 units + updated state alignment + sparsification	45.1 45.1 15.2 nz	17.1 16.4 16.1	19.6 18.6 18.5	English Broadcast News	50	17.5	18.8	
Bing Voice Search (Sentence error rates)					24	30.4	36.2	
Google Voice Input					5,870	12.3		16.0 (>>5,870hrs)
Youtube					1,400	47.6	52.3	
GMM 72 mix DT 2000h SA	102.4	17.1	18.6					

go here: <http://yann.lecun.com/exdb/mnist/>

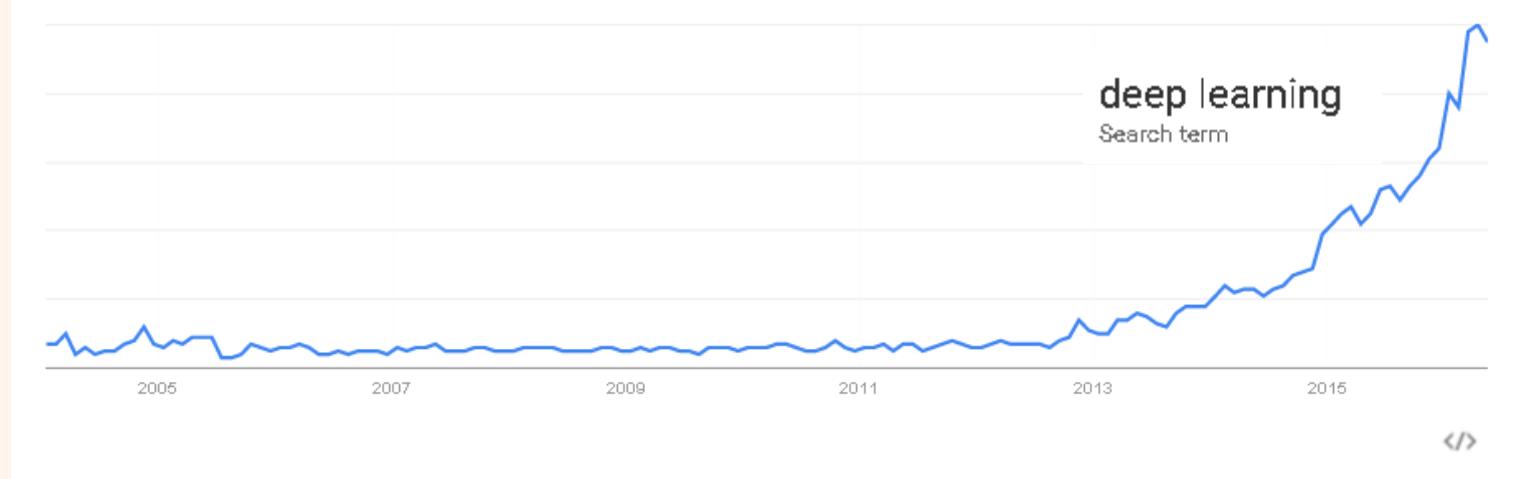
From here:

<http://people.idsia.ch/~juergen/cvpr2012.pdf>

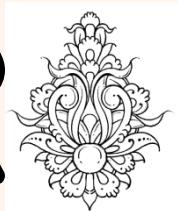
Dataset	Best result of others [%]	MCDNN [%]	Relative improv. [%]
MNIST	0.39	0.23	41
NIST SD 19	see Table 4	see Table 4	30-80
HWDB1.0 on.	7.61	5.61	26
HWDB1.0 off.	10.01	6.5	35
CIFAR10	18.50	11.21	39
traffic signs	1.69	0.54	72
NORB	5.00	2.70	46

یادگیری عمیق

Google Trends

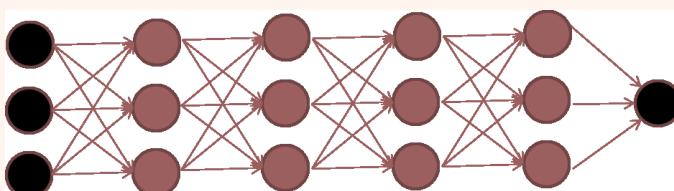


یادگیری عمیق اخیراً بسیار مورد توجه قرار گرفته است و در بسیاری از زمینه‌های توانسته است بهبود چشمگیری نسبت به سایر روش‌ها به دست آورده.

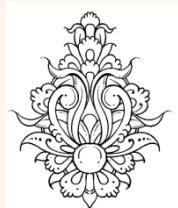


یادگیری عمیق چیست؟

- یک شبکه عصبی با چندین لایه بین ۰ و ۹ دی و خروجی
- این شبکه‌های از نمودهای پردازش اطلاعات در مخز انسان تقلید می‌کند، چندین لایه که وظیفه‌ی استخراج ویژگی و شناسایی را به صورت همزمان انجام می‌دهند.
- شبکه‌های چند لایه سال‌هاست که شناخته شده‌اند، اما تنها ویژگی شبکه‌های عمیق داشتن چندین لایهی مخفی نیست!

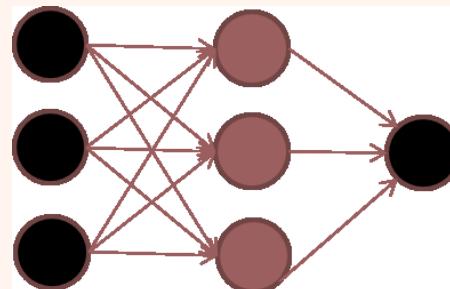


شبکه عصبی

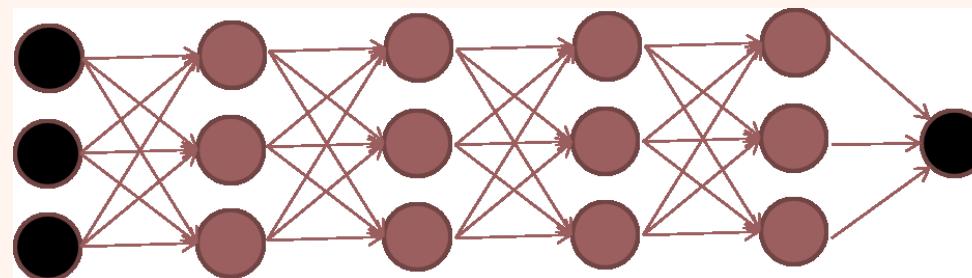


دانشکده
سینمایی
بهشتی

- شبکه‌ی عصبی با تعداد لایه‌های مخفی محدود به راهنمای آموزش می‌بیند:



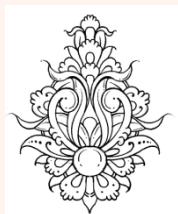
- اما در صورت افزایش تعداد لایه‌ها، الگوریتم‌های یادگیری کاری لازم را نفواهند داشت.



دانشکده
سینمایی

یادگیری عمیق

- در واقع، مهمترین تفاوت فوت و فن‌هایی است که برای آموزش این شبکه‌ها به کار گرفته شده است.
- یکی از این شیوه‌ها یادگیری بی‌نظرات *unsupervised feature learning* خصیصه‌هاست.
- کنترل پارامترهای آزاد، نیز یکی از شیوه‌های مورد استفاده است.

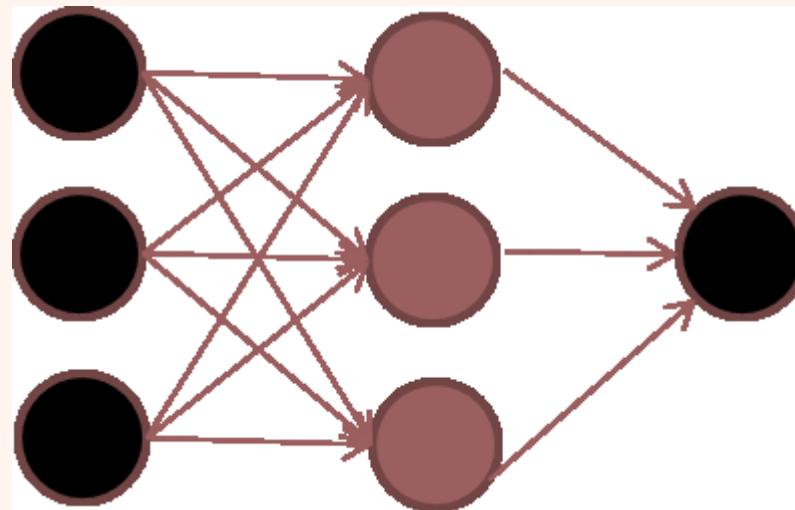


دانشکده
سینمای
بهریتی

آموزش در شبکه‌های پنهان لایه

مجموعی داده‌ها

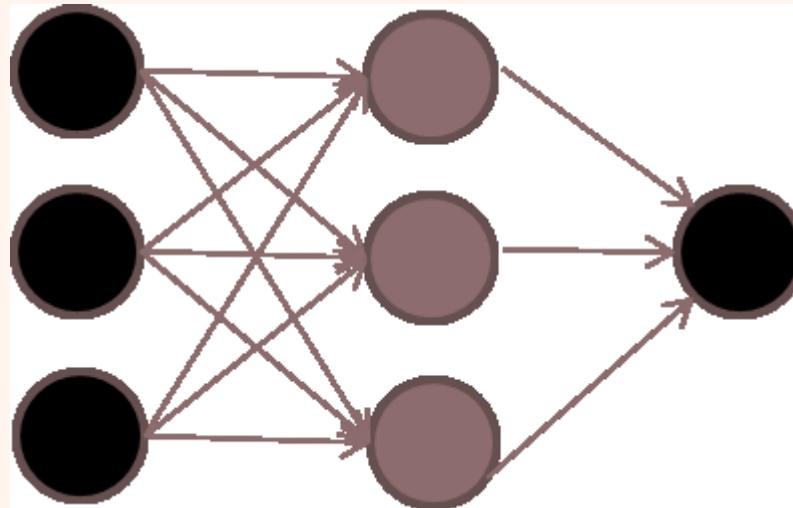
Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			



v

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			



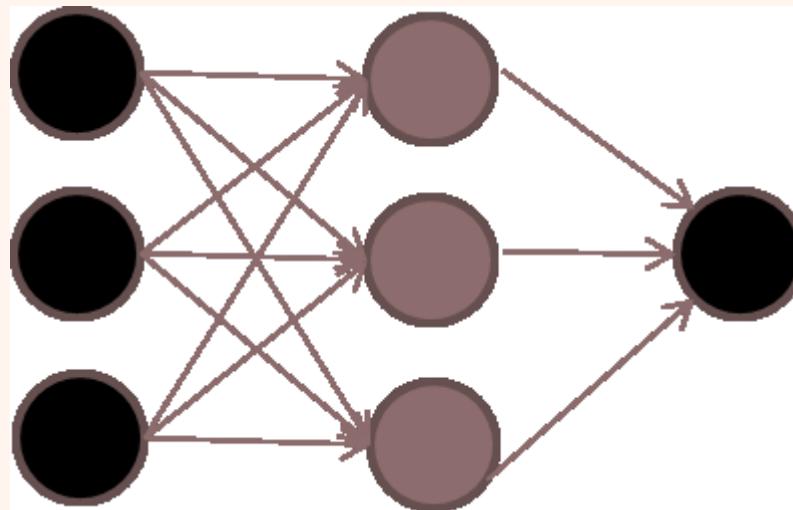
دانشکده
سینما
بهریتی



آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

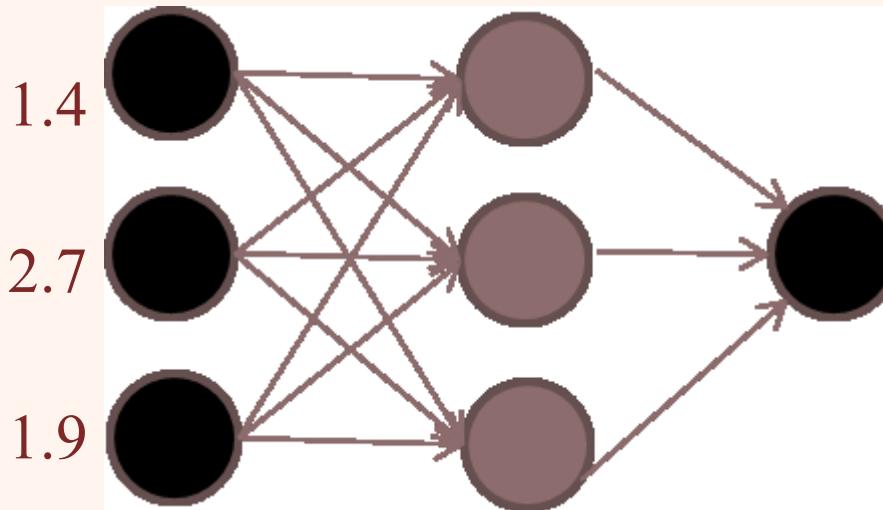
مقداردهی اولیه وزن‌ها به صورت تصادفی



آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

اعمال و اهدیا

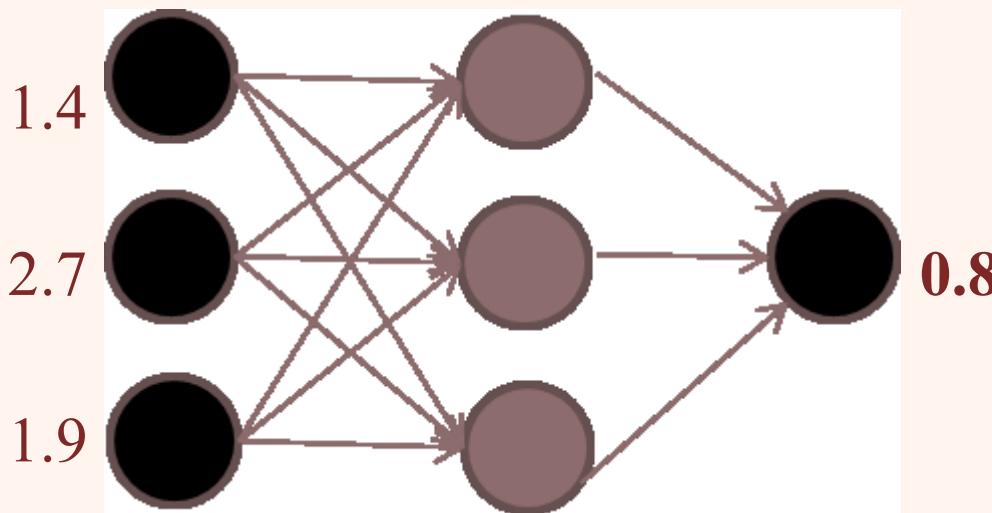


دانشکده
سینما
بهریتی

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

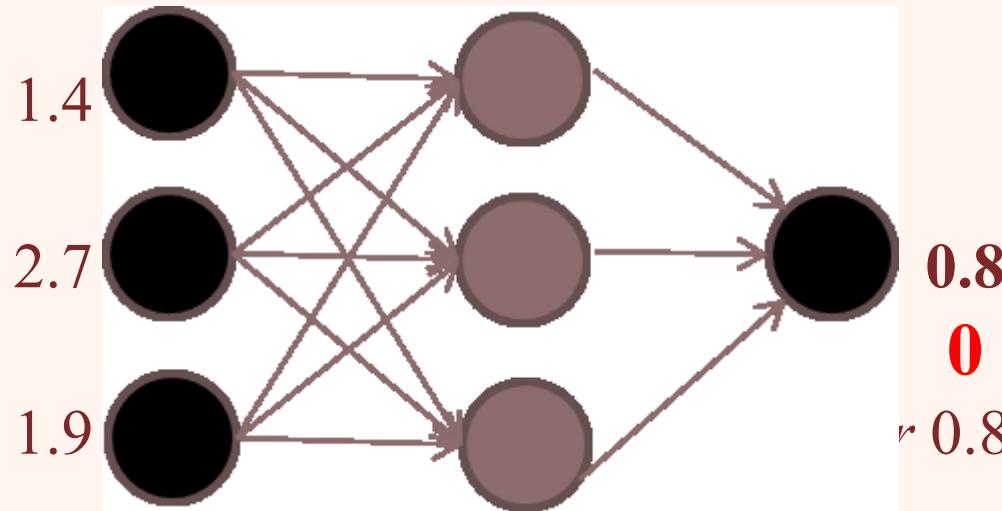
هماسه‌ی فروجی



آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

مقایسه با مقدار مطلوب

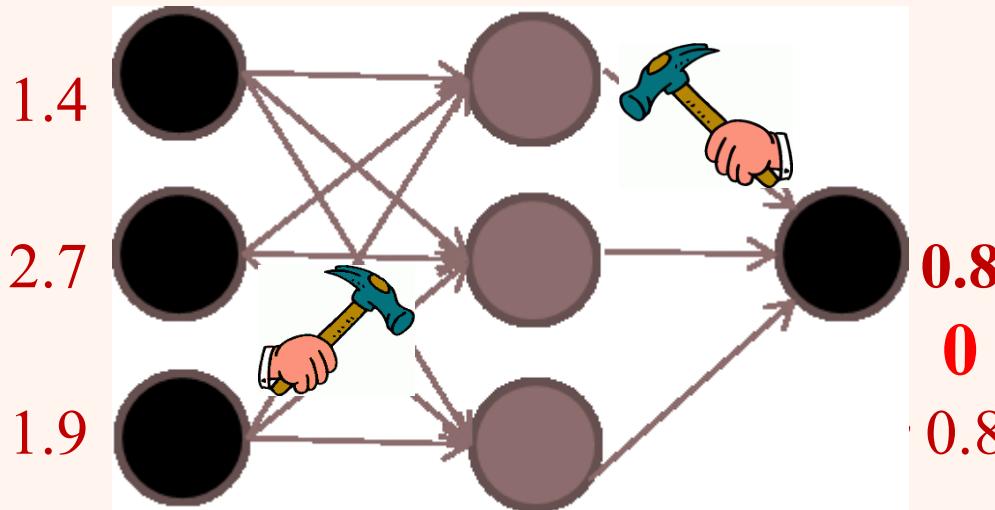


دانشکده
سینمایی
بهرمی

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

تنظیم وزن‌ها براساس فقط

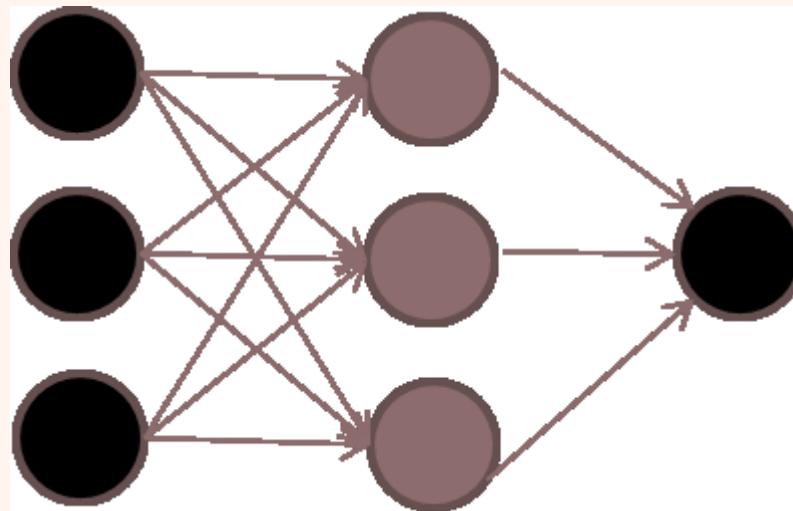


دانشکده
سینما و
بصیرتی

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

اعمال و اwendیها

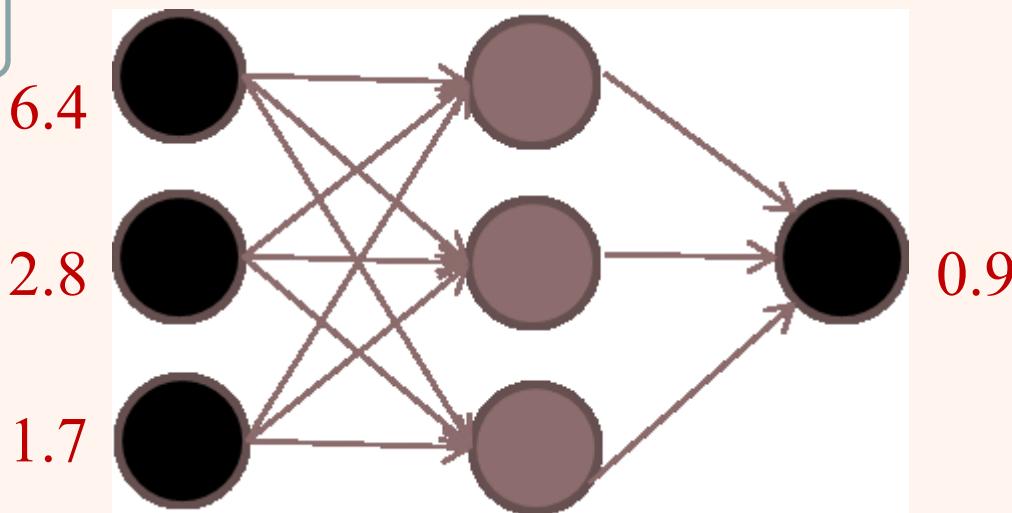


دانشکده
سینما
بهریتی

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

هماسه‌ی فروجی

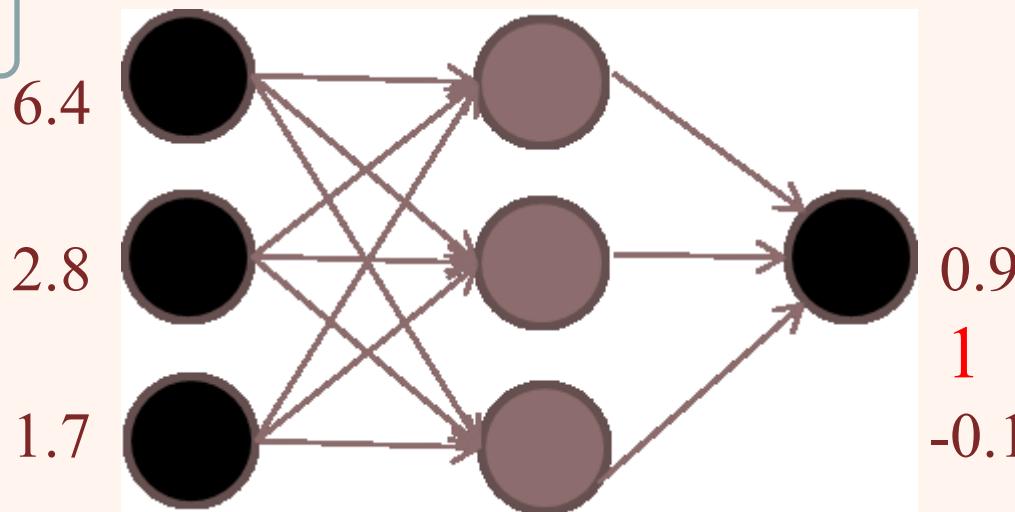


دانشکده
سینما
و تئاتر

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

مقایسه با مقدار مطلوب

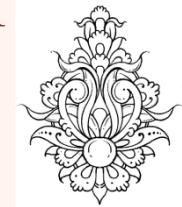
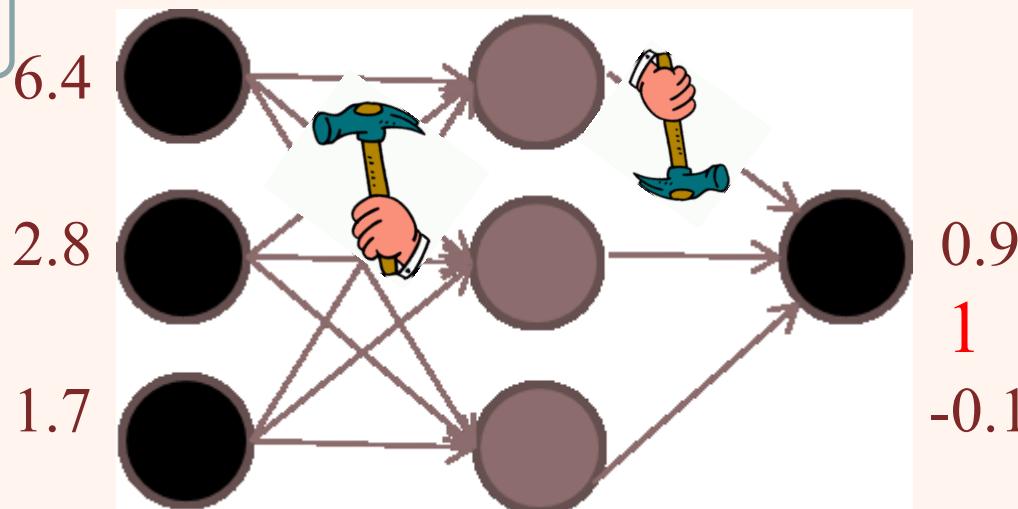


دانشکده
سینما
و تئاتر

آموزش در شبکه‌های چند لایه (ادامه...)

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

تنظیم وزن‌ها براساس فقط

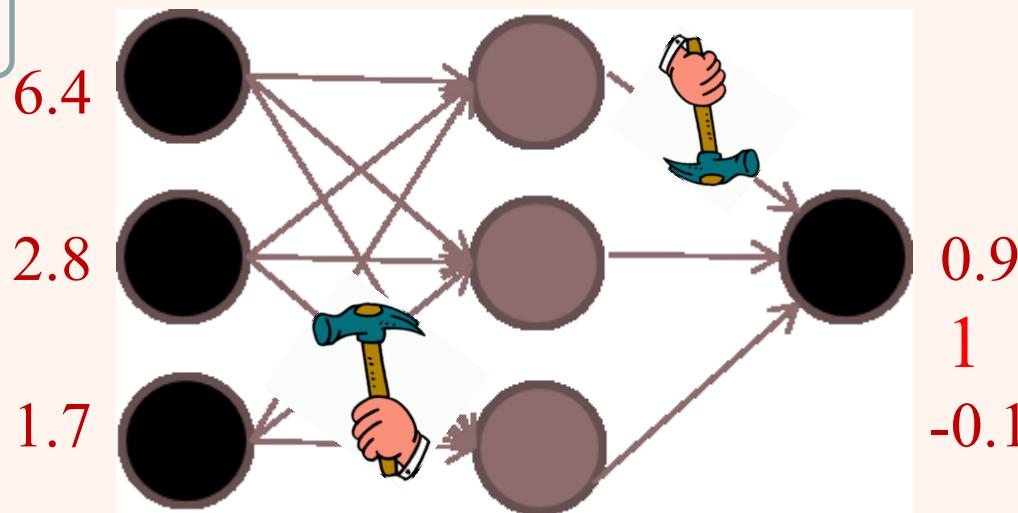


دانشکده
سینما و
بصیرتی

آموزش در شبکه‌های چند لایه

Fields	class		
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc ...			

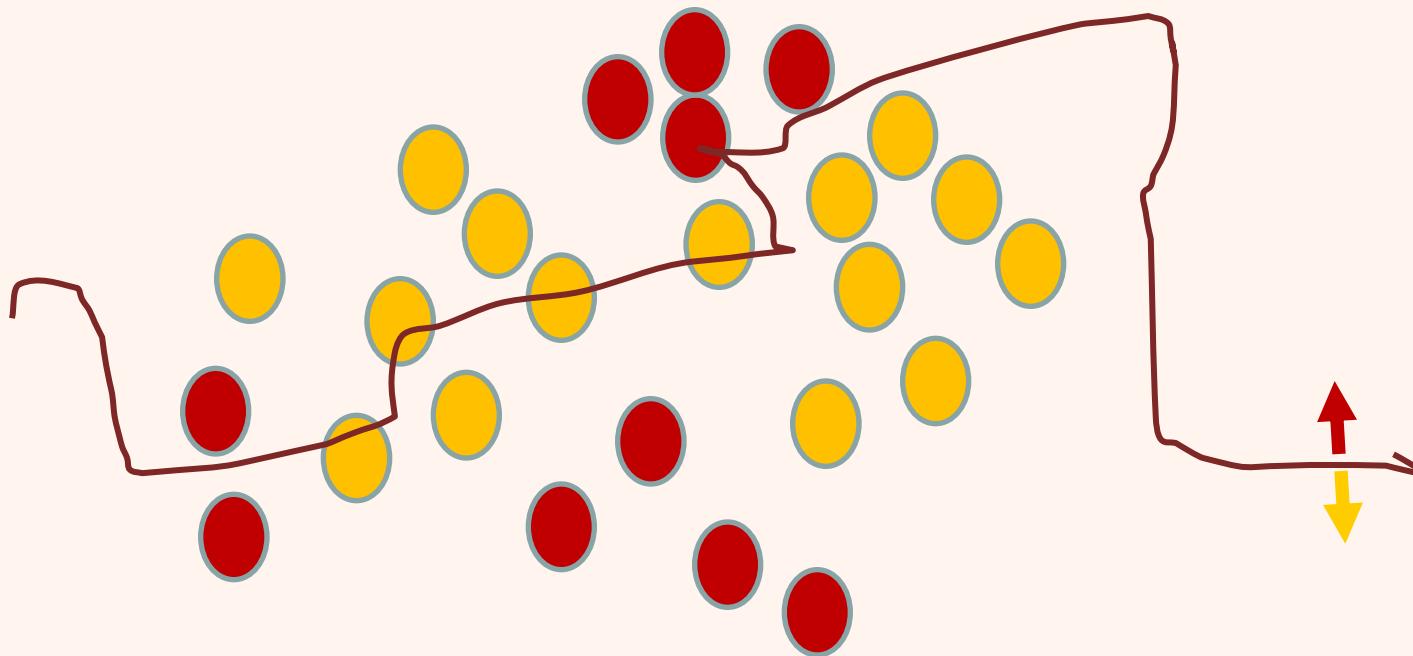
۹ به همین ترتیب



این روند هزاران بار و بلکه میلیون‌ها بار (بسطه به پیچیدگی مدل) تکرار می‌شود.
در این ترتیبی هر بار داده‌ای به صورت تصادفی انتخاب شده و به شبکه اعمال می‌شود.
وزن‌ها به گونه‌ای تنظیم می‌شوند که فقط گمینه شود.
تعداد زیاد پارامترهای آزاد می‌تواند منجر به گند شدن شبکه شود.

شبکه‌های پیچیده و مرز جداساز

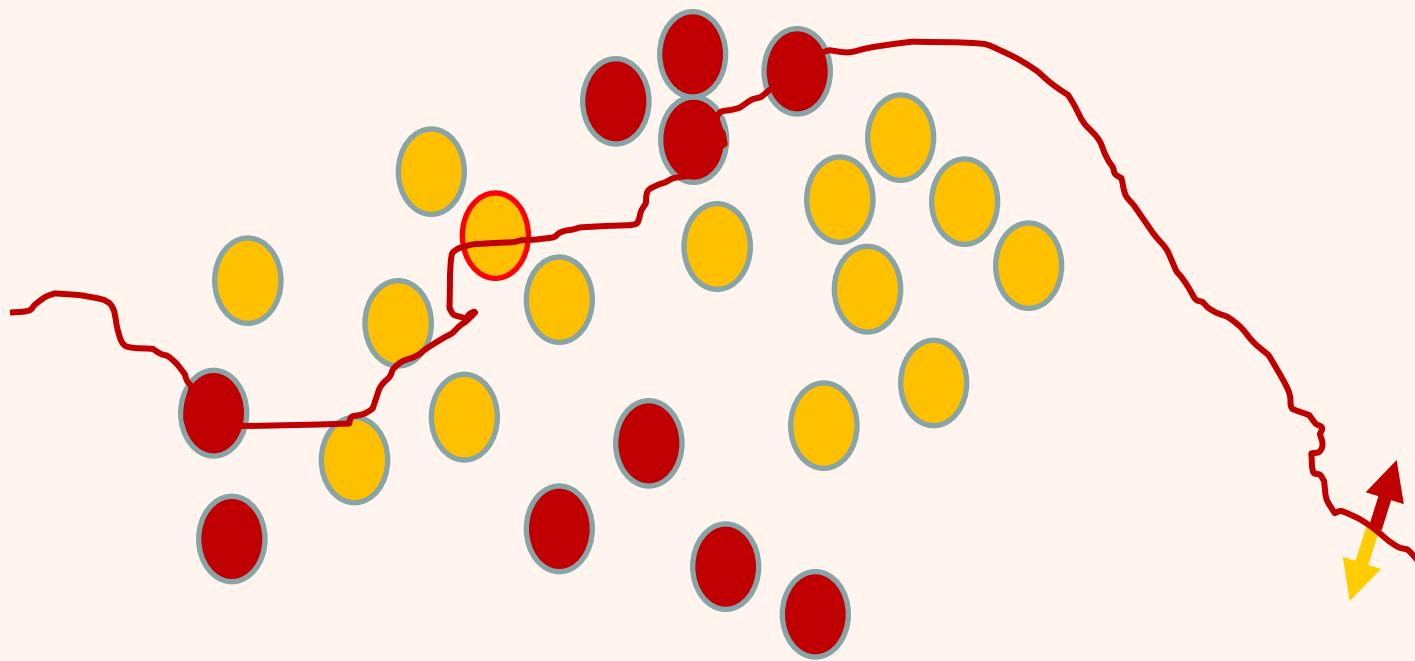
- در صورت پیچیدگی شبکه‌ی عصبی مرز جداساز هم قابلیت بیشتری برای جداسازی دارد:



دانشکده
سینمای
بهریتی

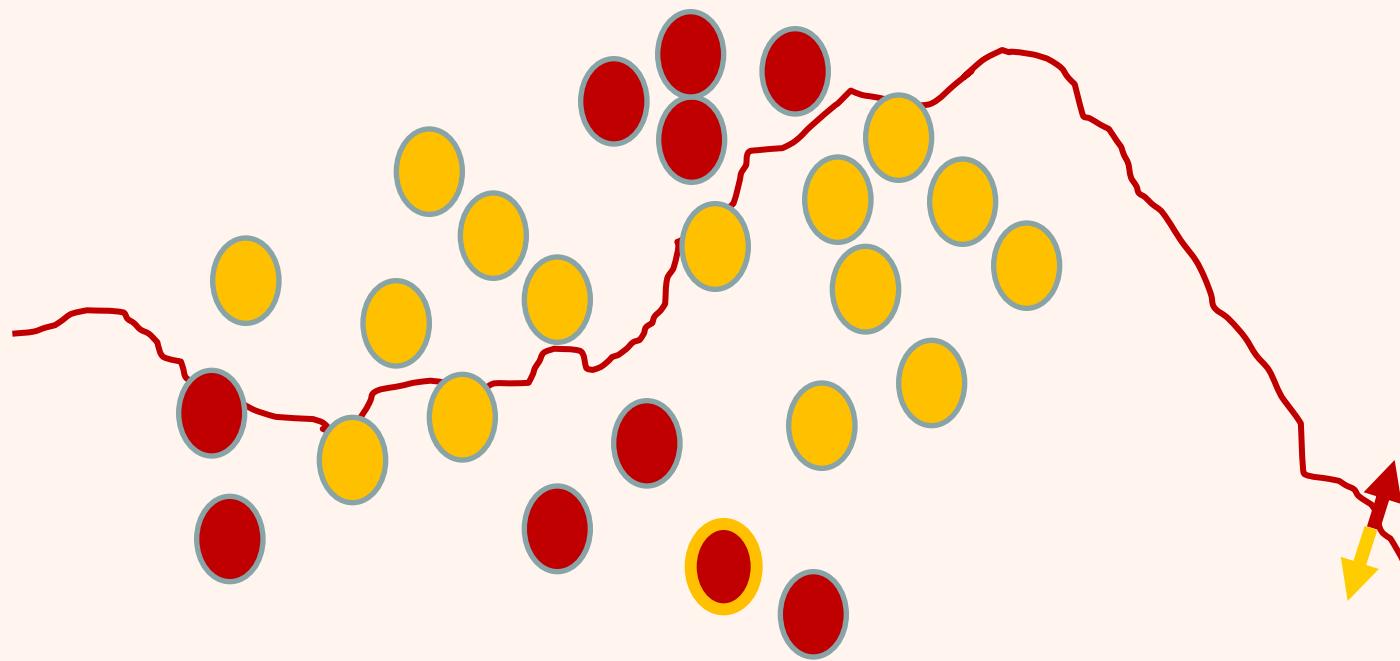
وزن‌های تصادفی اولیه

شبکه‌های پیچیده و مرز جداساز (ادامه...)



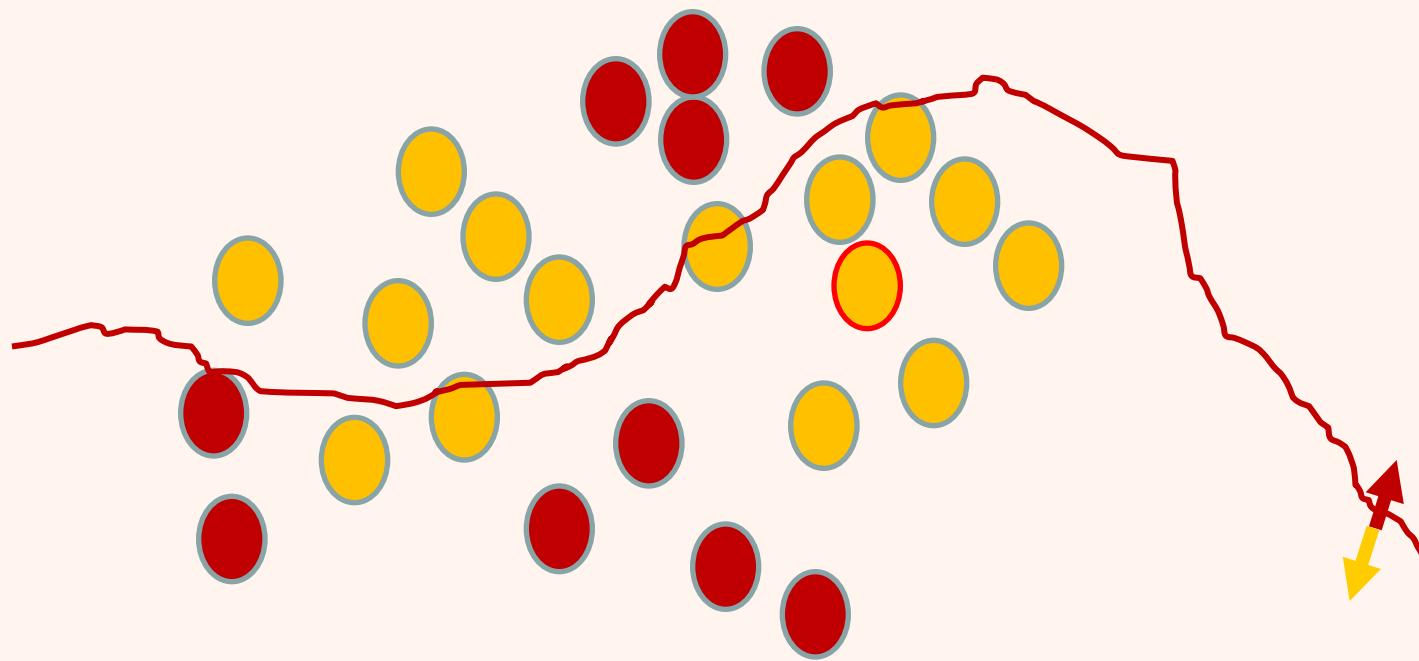
یک نمونه انتخاب و بر اساس آن وزن‌ها تنظیم می‌شوند

شبکه‌های پیچیده و مرز جداساز (ادامه...)



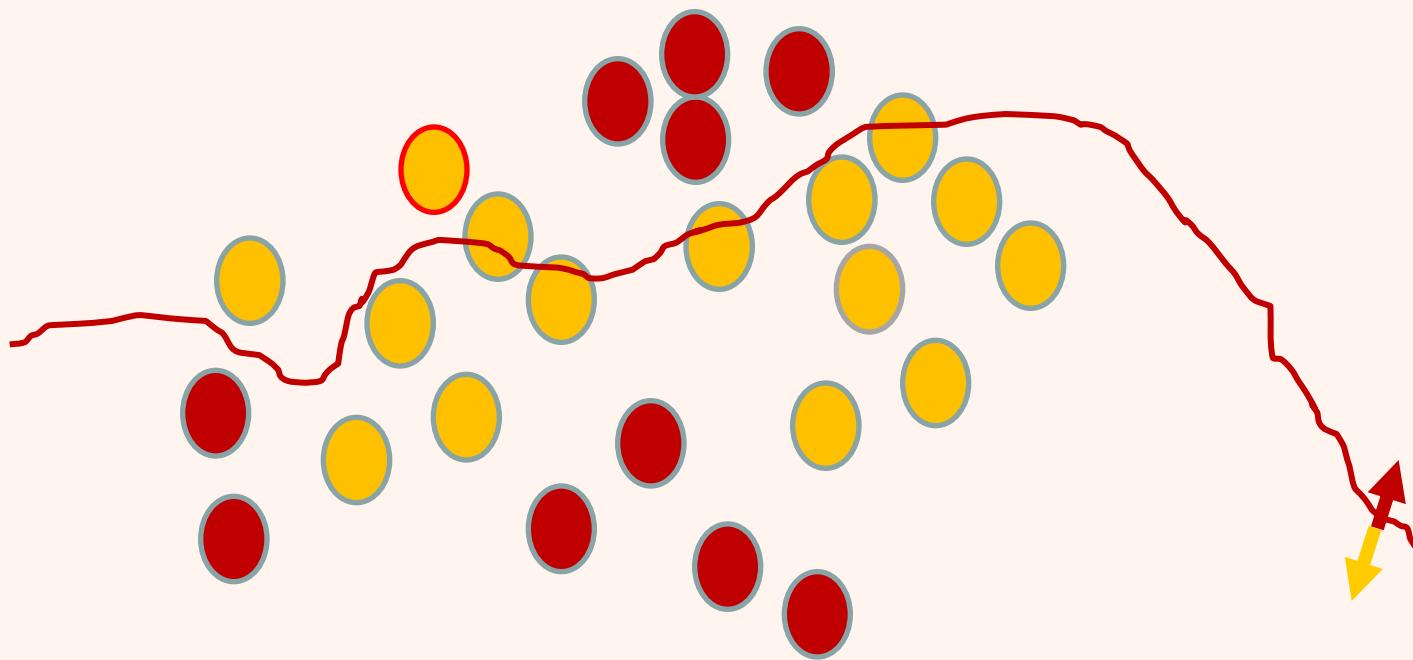
یک نمونه انتفاب و بر اساس آن وزن‌ها تنظیم می‌شوند

شبکه‌های پیچیده و مرز چداساز (ادامه...)



یک نمونه انتفاب و بر اساس آن وزن‌ها تنظیم می‌شوند

شبکه‌های پیچیده و مرز جداساز (ادامه...)

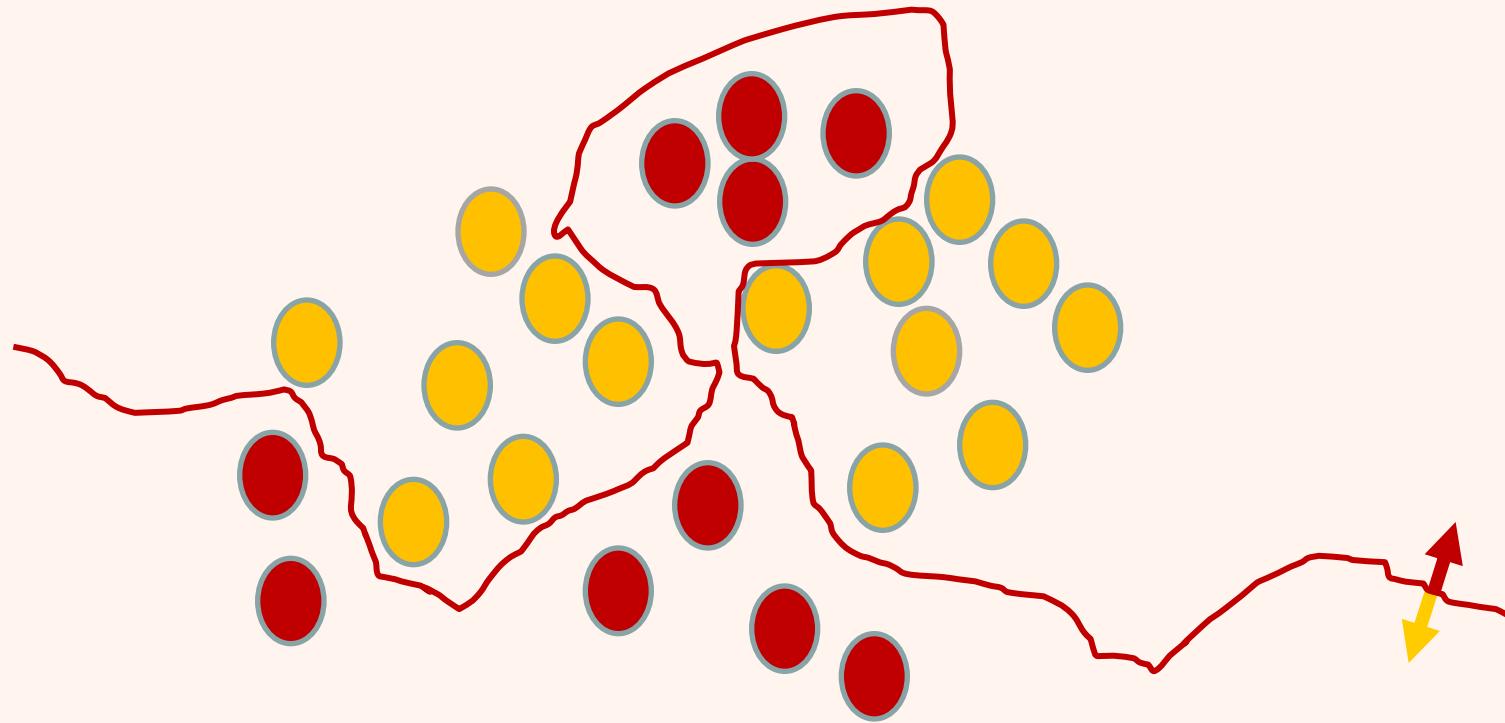


یک نمونه انتفاب و بر اساس آن وزن‌ها تنظیم می‌شوند



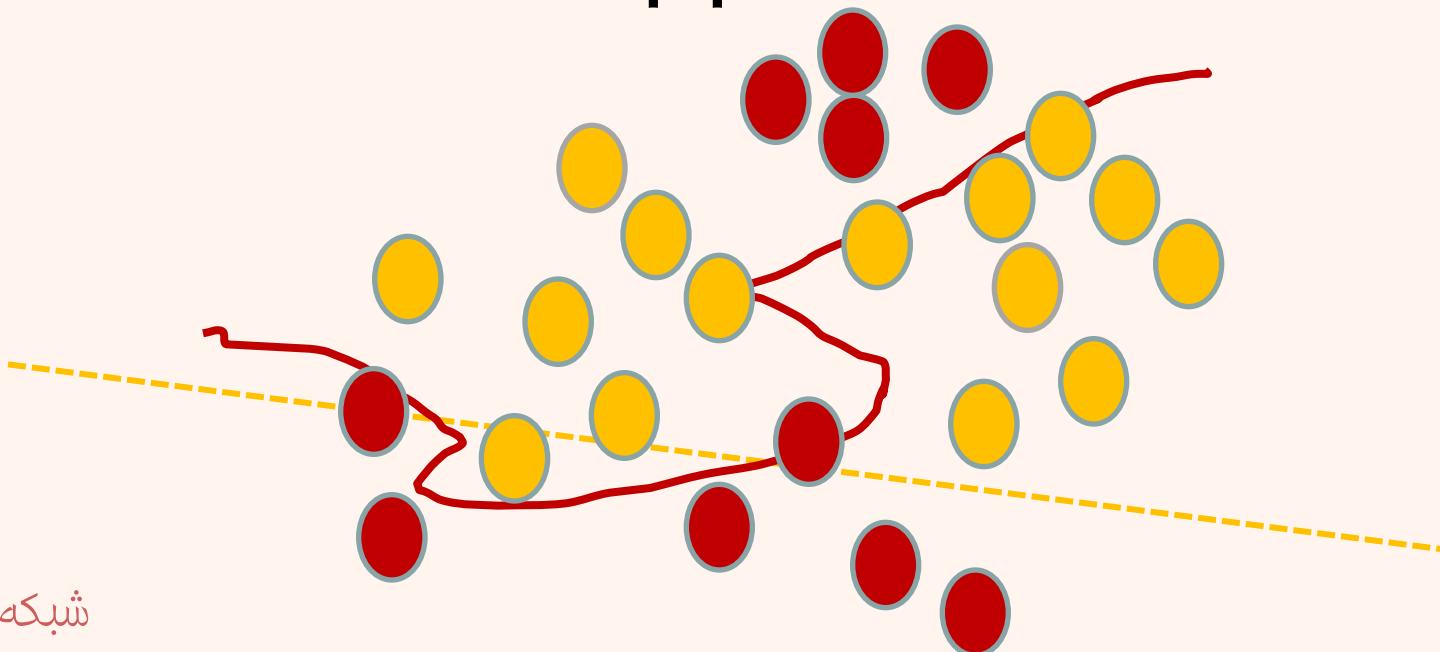
شبکه‌های پیچیده و مرز جداساز (ادامه...)

و در نهایت:

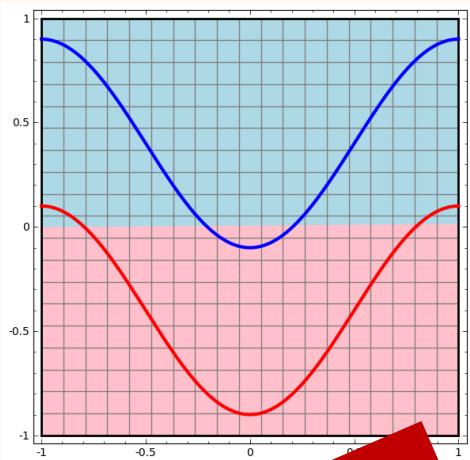


دانشگاه
بهشتی

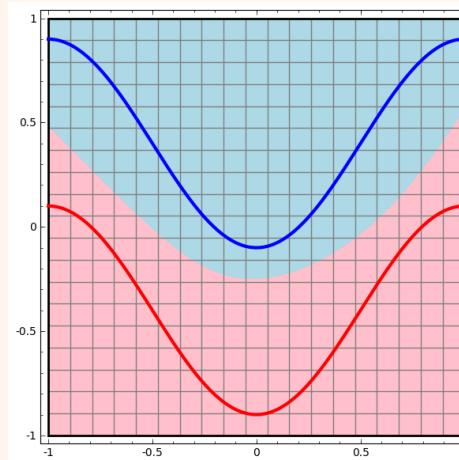
- وزن‌های آموزش دیده، **تفسیرپذیر** نیستند.
- یک شبکه‌ی **تک‌لایه** توانایی دسته‌بندی داده‌های جدایی‌پذیر فطی را دارد.
- می‌دانیم که یک شبکه‌ی عصبی با تنها **یک لایه** **مخفي** توانایی جداسازی پیچیده‌ترین اشکال را دارد.



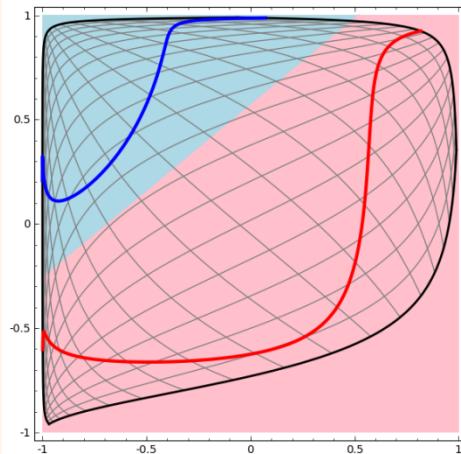
نقش لایه‌های مخفی



اپسیترون



شبکه چندلایه

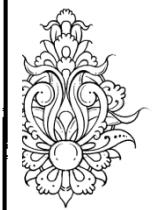
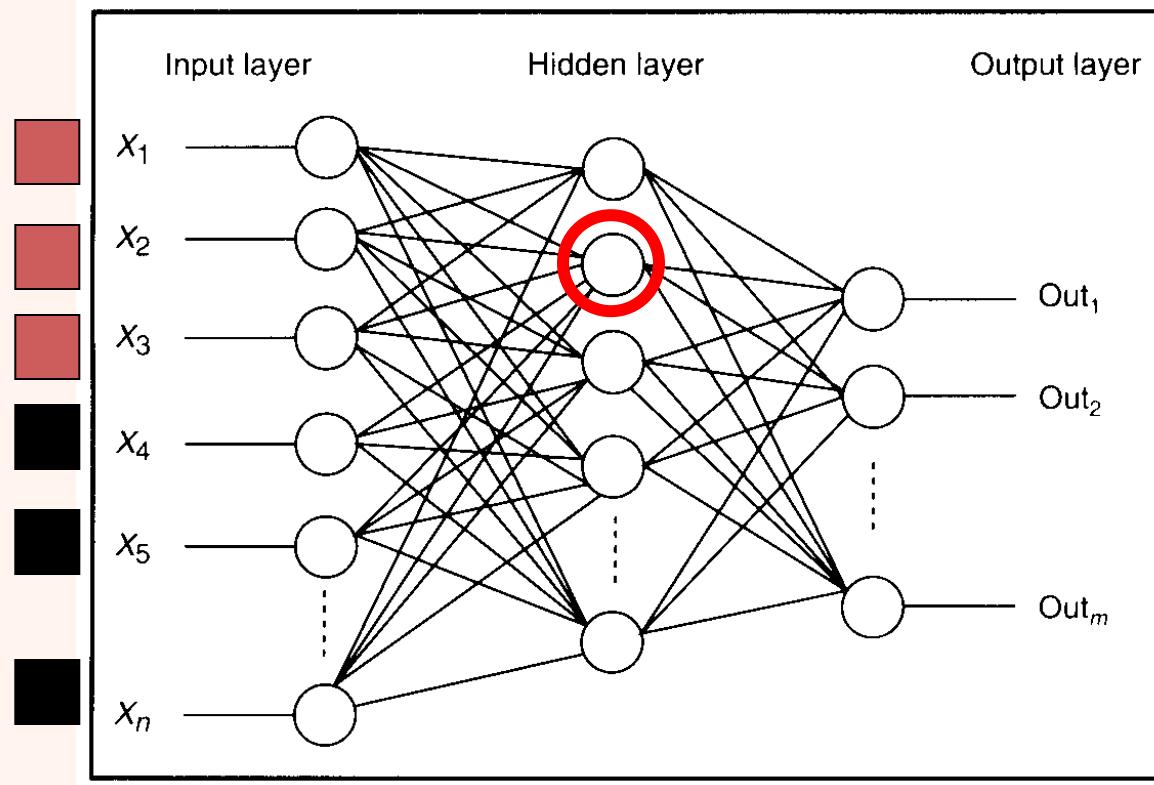
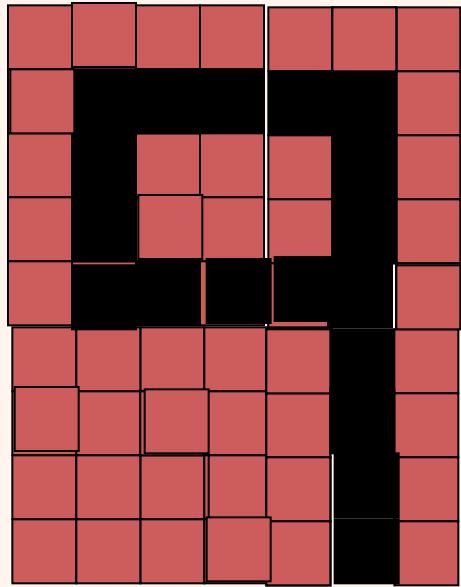


داده‌ی وجودی به لایه‌ی آفر



دانشکده
سینماسازی
بهشتی

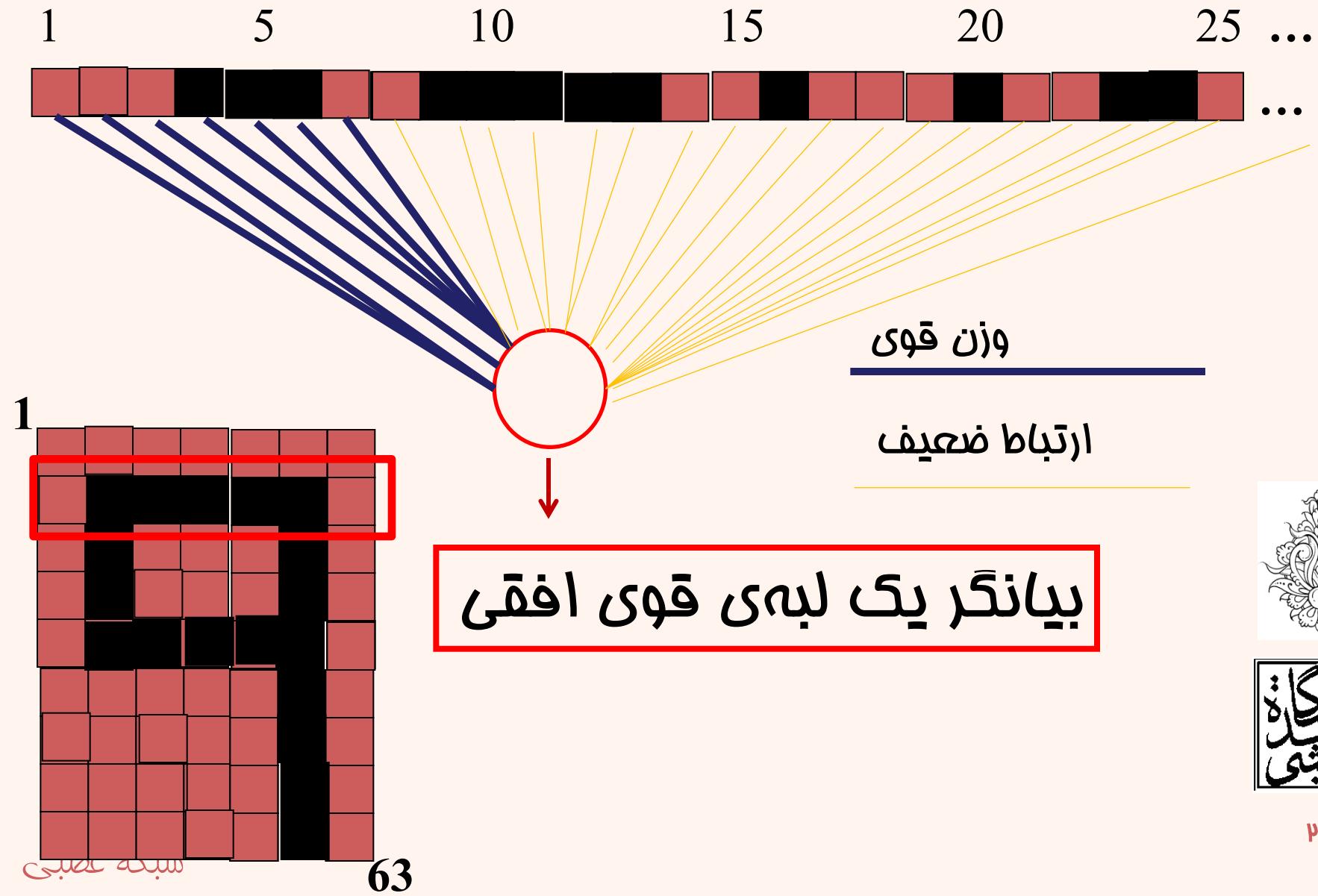
اسندهای ویژگی



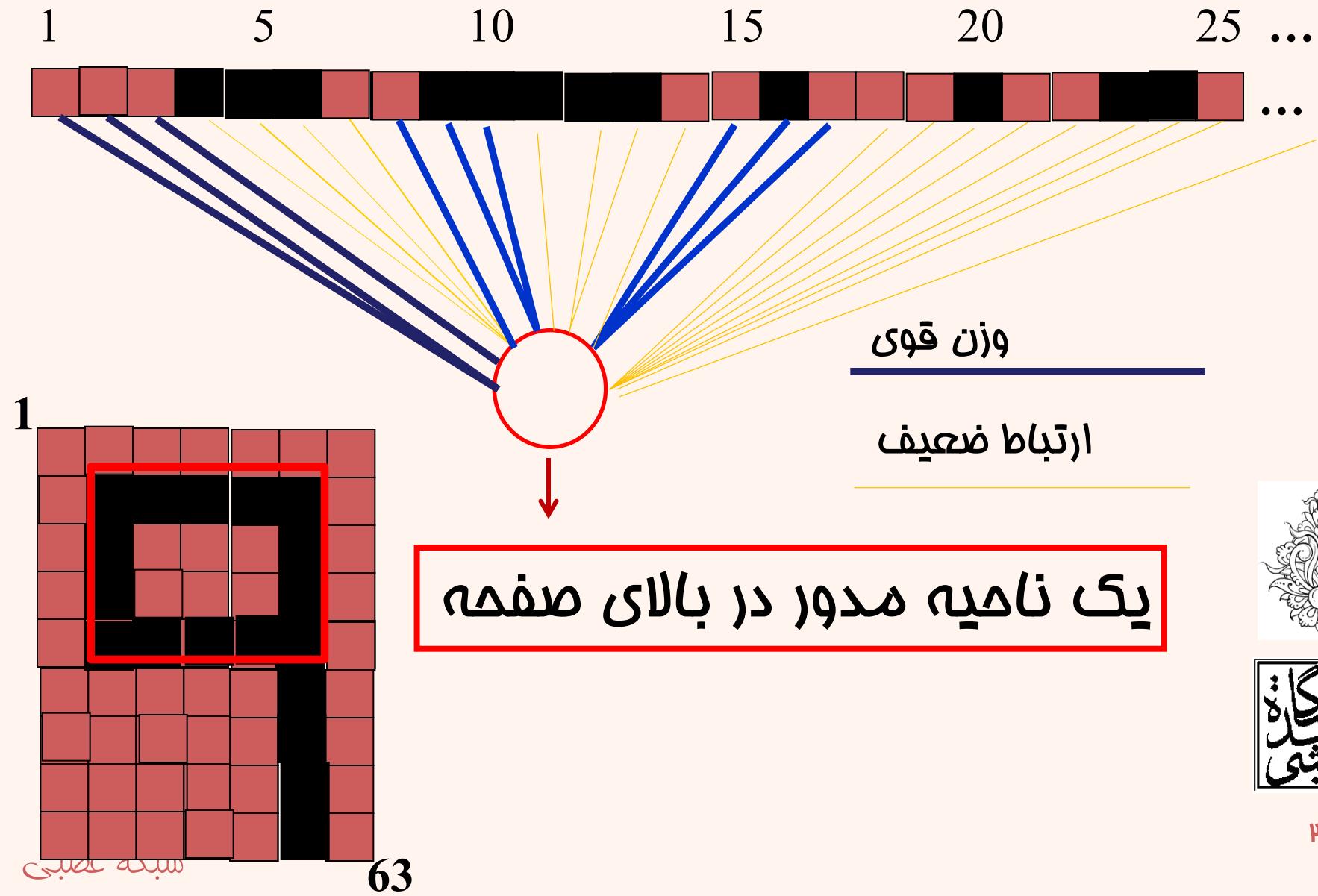
نقش نوونهای لایه مخفی پیست؟

دانشکده
بهشتی

اسنفراچ ویزگی (ادامه...)



اسندهای ویژگی (ادامه...)



دانشکده
سینمای
بهرستانی

اسندهای ویژگی (ادامه...)

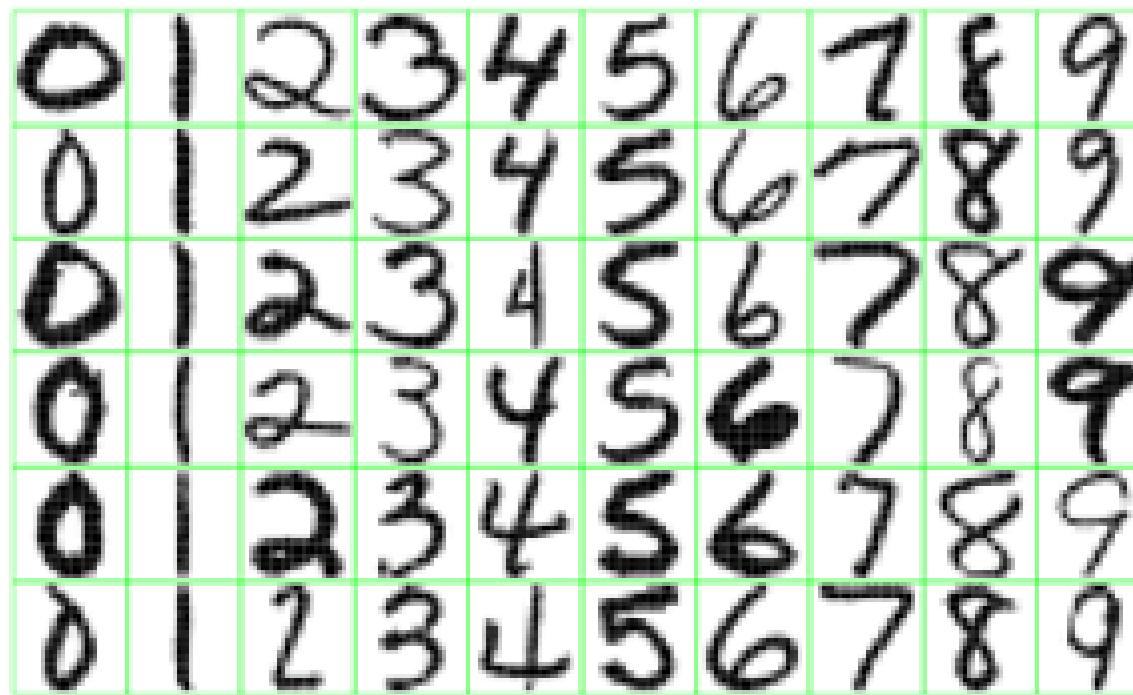


Figure 1.2: Examples of handwritten digits from U.S. postal envelopes.

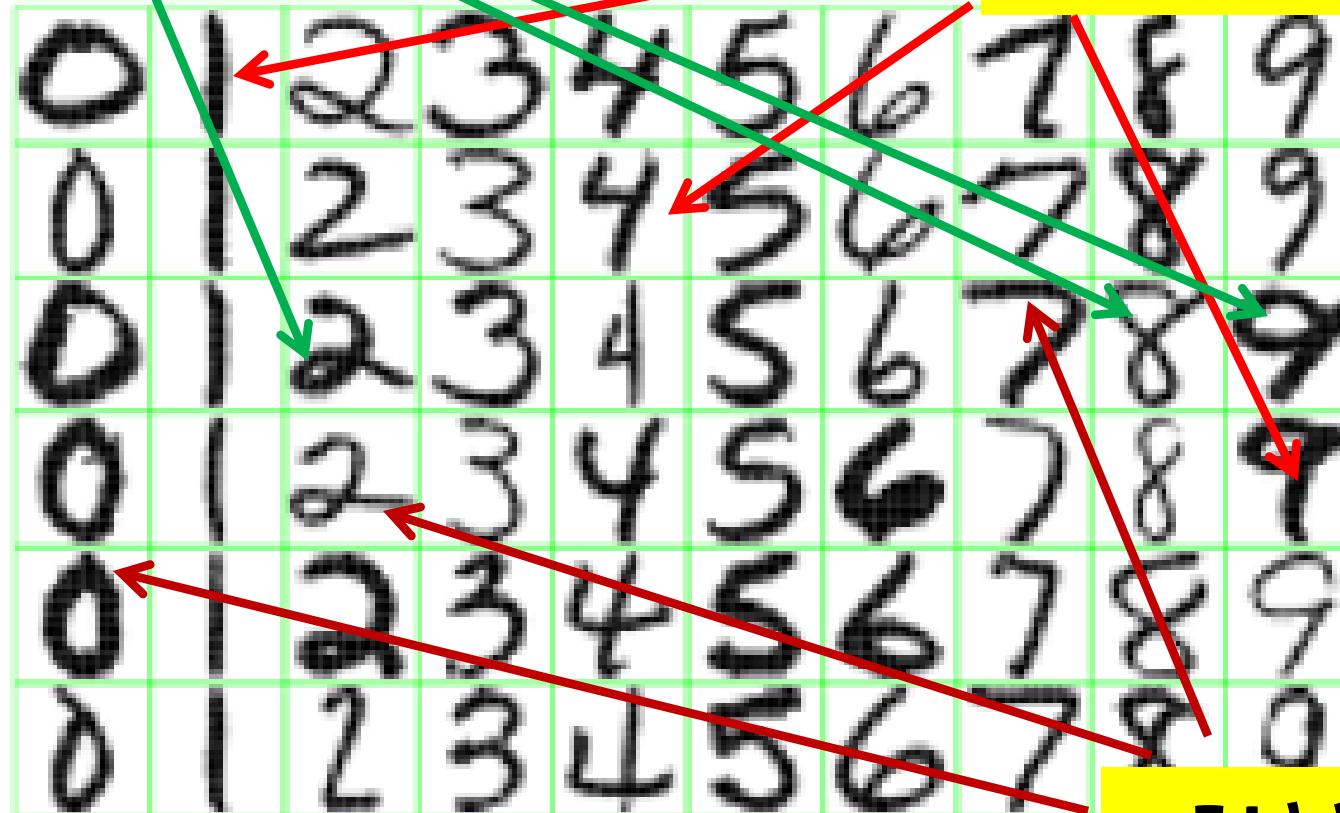
شبکه‌ی عصبی چه نوع خصیصه‌ای برای تشخیص کاراکتر یاد می‌کیرد؟



دایردهای کوچک

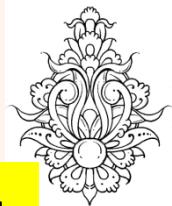
اسندهای اسخراج ویژگی (ادامه...)

خطوط عمودی



خطوط افقی

Figure 1.2: Examples of handwritten digits from U.S. postal envelopes.



دانشکده
سینمایی
بهشتی

خصوصیات اسندهای اسخراج شده نسبت به مکان مساس هستند!

استخراج ویژگی (ادامه...)

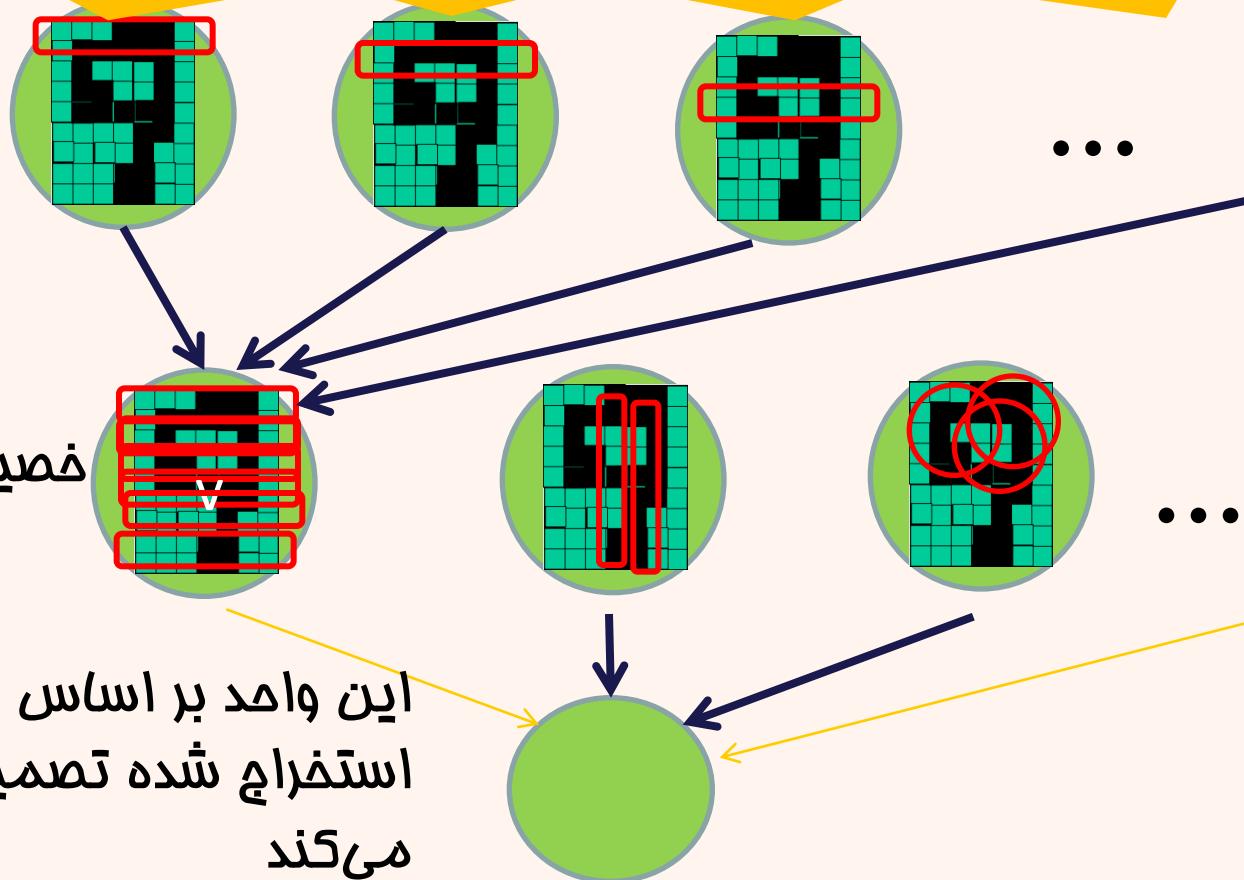
- لایه‌های مخفی می‌توانند خصیصه‌های سطح بالا

تشخیص فقط در
مکان‌های مختلف

خصیصه‌های سطح بالا

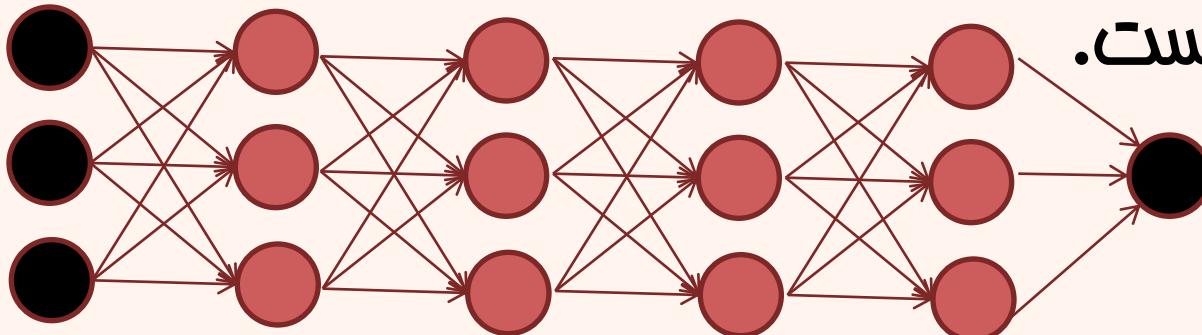
این واحد بر اساس ویژگی‌های
استخراج شده تصمیم‌گیری
می‌کند

شبکه عصبی

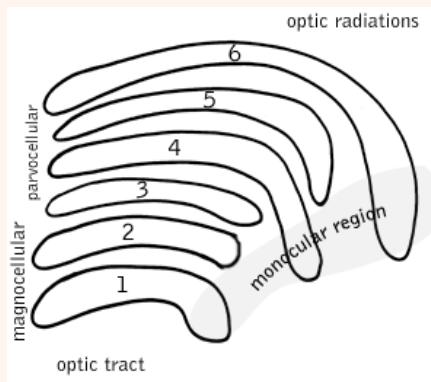


اسنفراچ ویزگی (ادامه...)

- در این حالت اسنفاده از چندین لایه متفلف توجیه پذیر است.



- درست مانند مخز

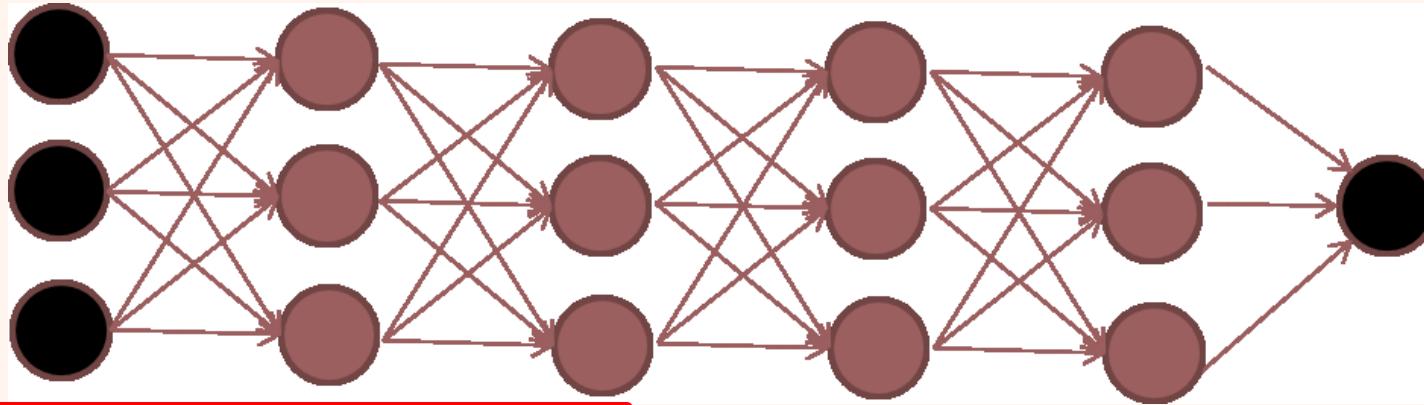


- لایه‌های مخفی باید بتوانند فضیصه‌های مناسب را اسنفراچ کنند.

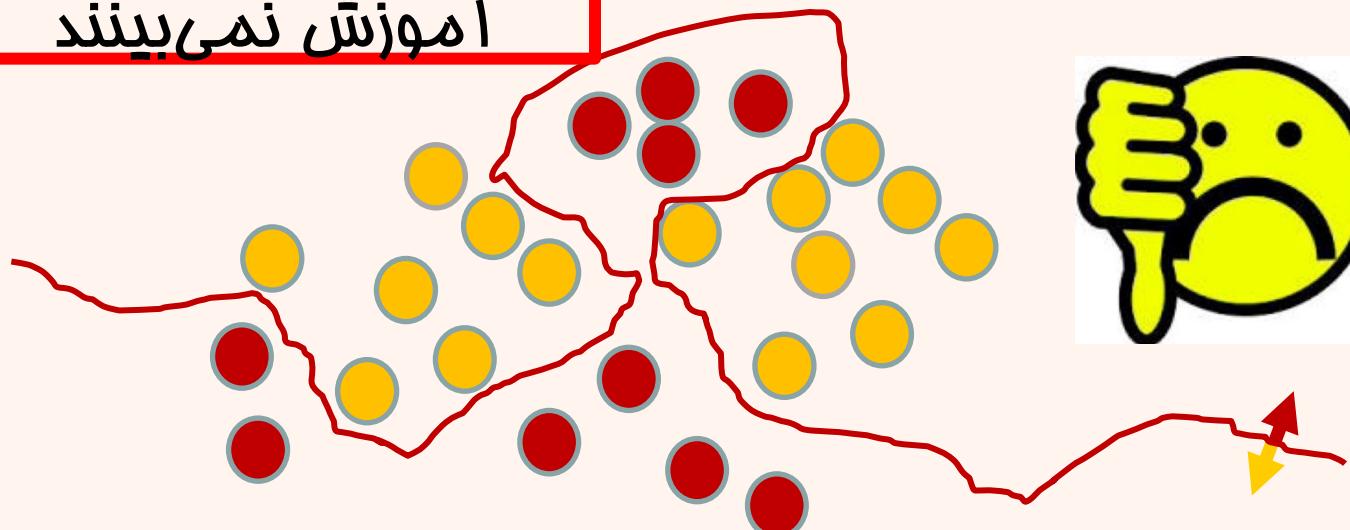


دانشکده
سینمایی

اسندهای خودکار (ادامه...)



لایه‌های نفست خوب
آموزش نمی‌بینند



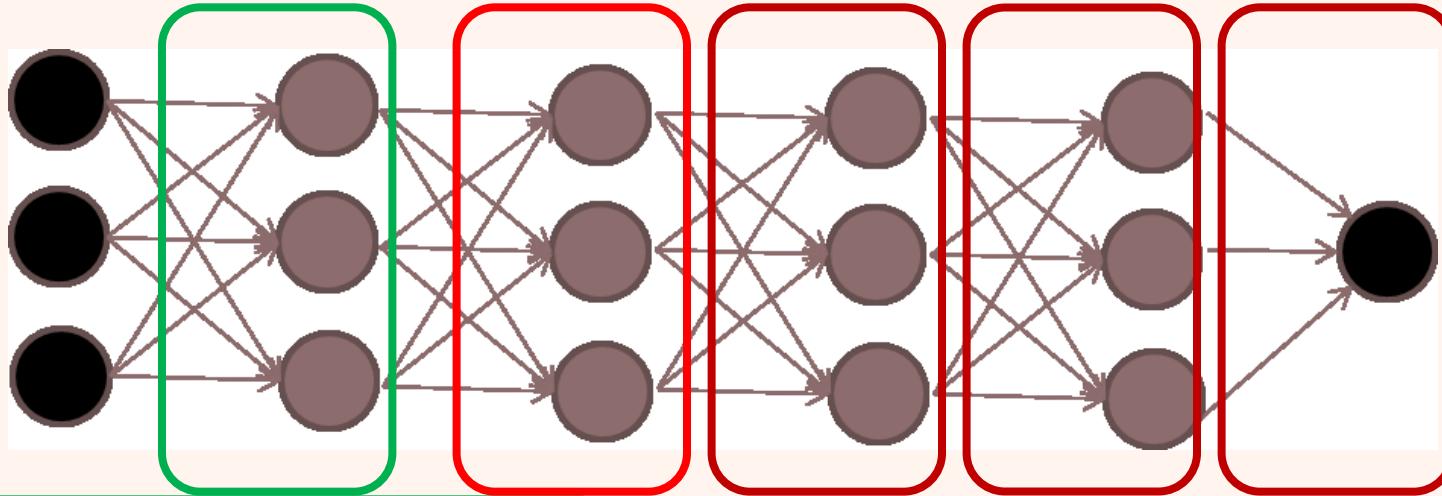
مشکل بیش برآری

دانشگاه
سینمایی

۱۳



شیوه‌ای نو برای آموزش



ابتدا این لایه آموزش می‌بینند

سپس این لایه و به
همین ترتیب

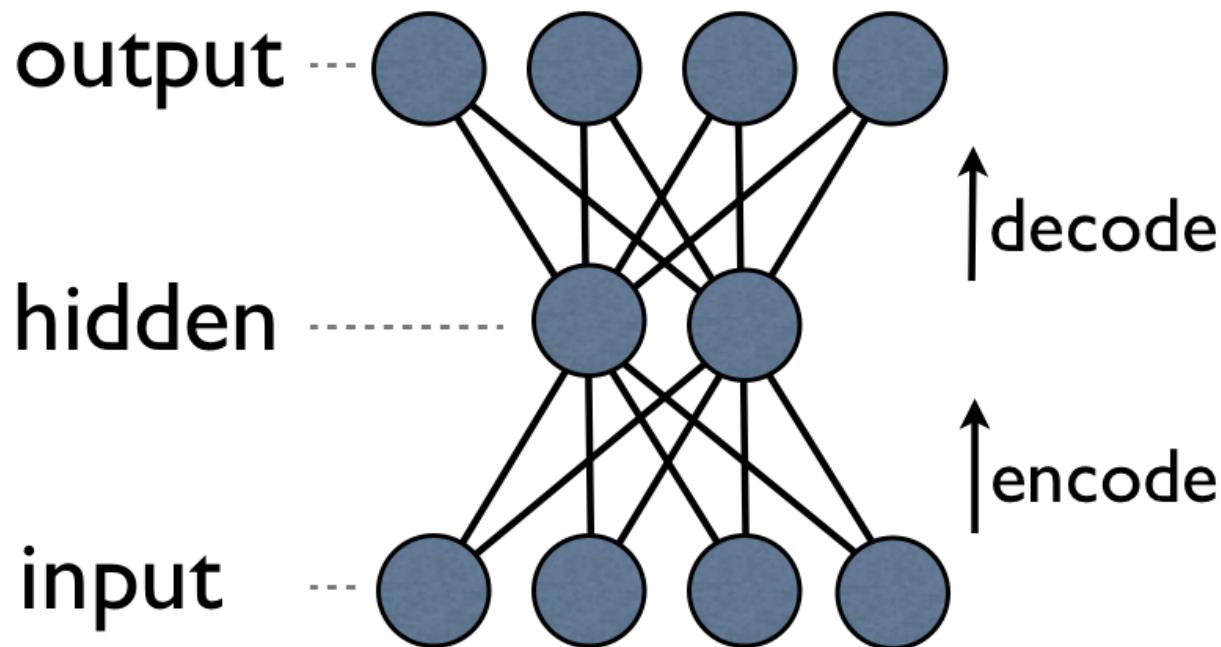
هر کدام از لایه‌های مخفی به صورت یک *auto-encoder* آموزش می‌بینند.
در واقع تلاش می‌شود داده‌های لایه‌ی داده را به بهترین نحو توصیف کنند.



دانش
سازمان
جمهوری اسلامی ایران

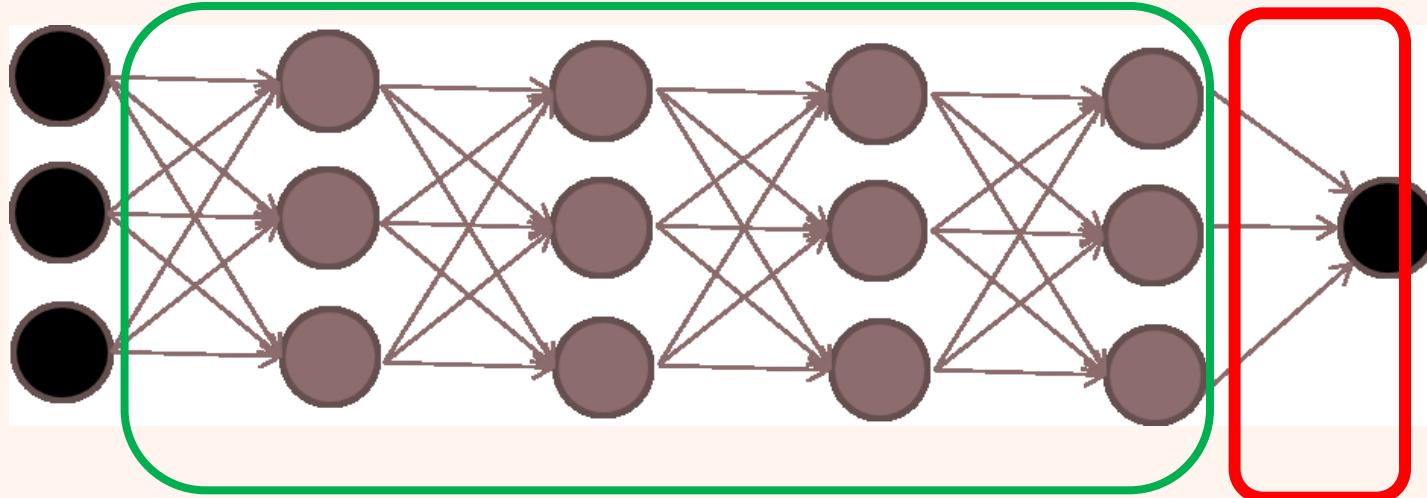
auto-encoder

از شیوه‌های استاندارد آموزش برای تنظیم وزن‌ها استفاده می‌کند به گونه‌ای که ورودی را بازتولید کند.



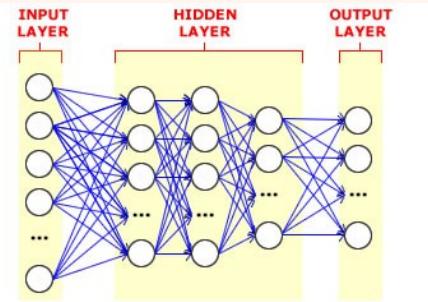
در این حالت تعداد نوون‌های لایه‌ی مخفی کمتر در نظر گرفته می‌شود، بدین‌ترتیب لایه‌ی مخفی برای استخراج خصیصه آموزش می‌بیند.

شیوه‌ای نو برای آموزش



لایه‌های میانی به عنوان auto-encoder آموزش می‌بینند.

لایه‌ی آن برای دسته‌بندی مورد استفاده قرار می‌گیرد.



شبکه عصبی

انواع مختلفی از شبکه‌های عمیق وجود دارد،
شبکه‌های متفاوت، ساختار و شیوه‌های
آموزش مختلف در شبکه‌های مختلف استفاده
می‌شود.



يادگيري خودآموزه

self-taught Learning



یادگیری خودآموزه

- دلایلی وجود دارد که نشان می‌دهد مخز انسان از یک شیوه‌ی یادگیری یکسان برای داده‌های مختلف استفاده می‌کند.
 - در یک آزمایش سیگنال‌های بینایی هوش‌فرما به بخش شنوایی مخزش هدایت شد و بخش شنوایی دیدن را آموزت!

**Visual Projections Routed to the Auditory Pathway in Ferrets:
Receptive Fields of Visual Neurons in Primary Auditory Cortex**

Anna W. Roe,^a Sarah L. Pallas,^b Young H. Kwon, and Mriganka Sur

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139



اگر این شیوه‌ی یادگیری کشف شود، می‌توان گفت به هدفمان رسیده‌ایم!

آموزش در مخز

- مخز انسان محدود 10^{14} ارتباط سیناپسی دارد.
- هر انسان در مرتبه 10^9 ثانیه زندگی می‌کند.
- اگر نمایش هر ارتباط سیناپسی به یک بیت احتیاج داشته باشد، برای استفاده از ظرفیت کل ذهن به نفع آموزش 10^{14} در 10^9 ثانیه نیاز است.
 - یعنی 10^5 بیت در ثانیه
- از این مثال می‌توان نتیجه گرفت ساختار مخز به صورت بی‌نظرات آموزش می‌بیند و بیشتر از داده‌های بدون برچسب استفاده می‌کند.



دانشگاه
تهران
بهشتی

یادگیری با نظارت

مجموعه‌ی آموزشی



Cars



Motorcycles

داده‌های آزمایش



در صورتی که داده‌های آموزشی محدود باشد، ممکن یادگیری با نظارت خوب عمل نکند.



آموزش به شیوه‌ی مخفی

مجموعه‌ی آموزشی



Cars



Motorcycles

داده‌های آزمایش



تصاویر بدون برچسب



دانشکده
سینما
بهمیتی

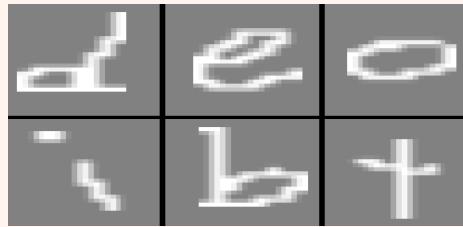
Self-taught Learning

آموزش به شیوه‌ی مخفی



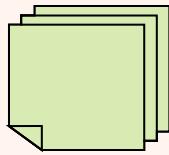
Labeled Digits

+



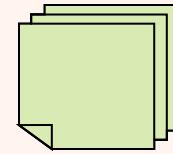
Unlabeled English characters

?



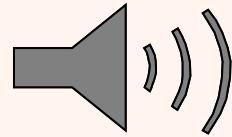
Labeled Webpages

+



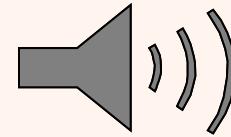
Unlabeled newspaper articles

?

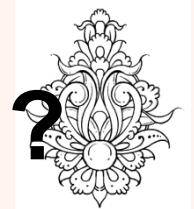


Labeled Russian Speech

+



Unlabeled English speech



تاریخچه‌ی یادگیری ماشین

• یادگیری بانظرارت (محدود بیست سال پیش)



Cars



Motorcycles

Supervised learning

• یادگیری نیمه‌نظرارت (محدود ده سال پیش)



Cars



Motorcycles



Semi-supervised learning

• انتقال یادگیری (محدود ده سال پیش)



Bus



Tractor



Aircraft



Helicopter

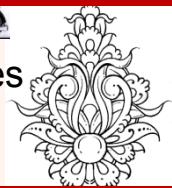


Cars



Motorcycles

Transfer learning



• همان‌گون: یادگیری فوداً موقته



Natural scenes



Car



Motorcycle

Self-taught learning

دانشکده
سینما
و تئاتر

یادگیری خودآموزنده

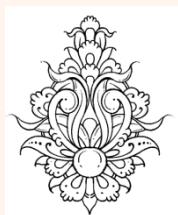
- مجموعه داده‌های آموزشی (برچسب خودده)

$\{(x_l^{(i)}, y^{(i)})\}_{i=1}^m \quad x_l^{(i)} \in R^n, y^{(i)} \in \{1, \dots, T\}$

- مجموعه داده‌های بدون برچسب

$\{x_u^{(i)}\}_{i=1}^k \quad x_u^{(i)} \in R^n, k \gg m$

- لزومی ندارد که داده‌های از یک نوع باشند.



دانشکده
سینما
بعلبکی

یادگیری خودآموقته

- هر داده به صورت ترکیب فطی یک مجموعه بردار پایه نمایش داده می‌شود.

$$x \approx \sum_j a_j b_j \quad b_j \in R^n, a_j \in R$$

$$x = 0.8 * \text{[basis image]} + 0.3 * \text{[basis image]} + 0.5 * \text{[basis image]}$$
$$= 0.8 * b_{87} + 0.3 * b_{376} + 0.5 * b_{411}$$

- از داده‌های بدون برعسب برای آموقتن بردارهای پایه استفاده می‌شود. در واقع، داده‌های بدون برعسب برای ارائه یک خصیصه‌ی سطح بالا استفاده می‌شود.

- پس از آن از داده‌های برعسب‌دار برای دسته‌بندی با نظرارت استفاده می‌شود.



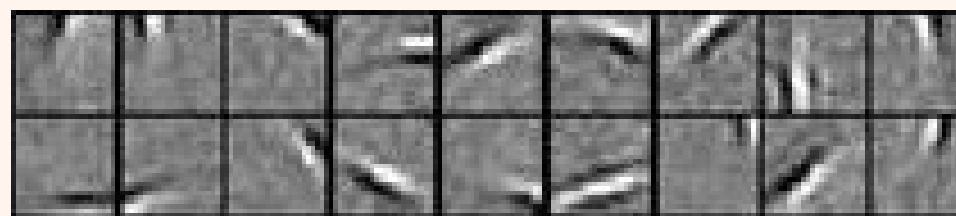
دانشکده
سینمای
بهره‌بری

مثالی از بردارهای پایه

Natural images.



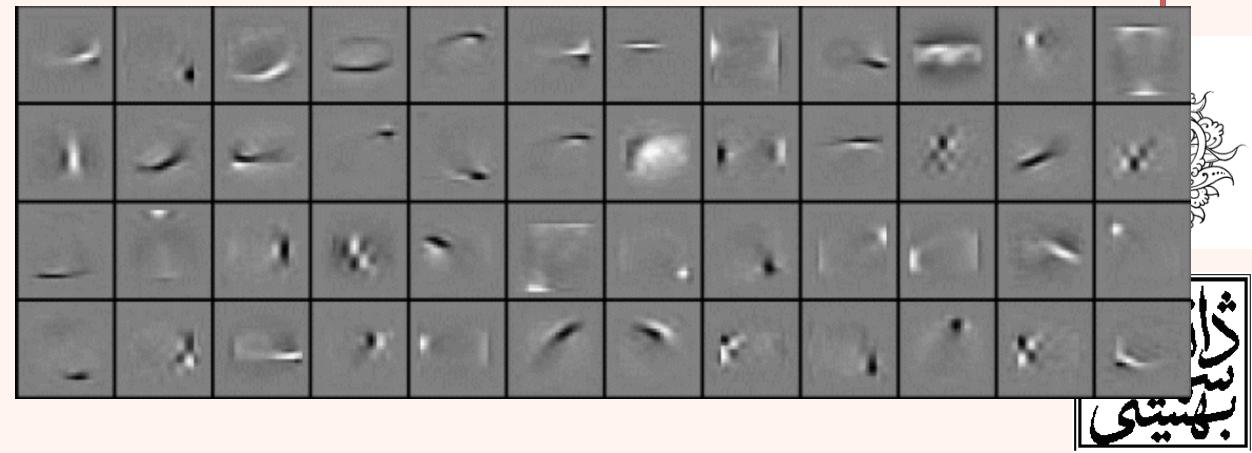
Learnt bases: “Edges”



Handwritten characters.



Learnt bases: “Strokes”



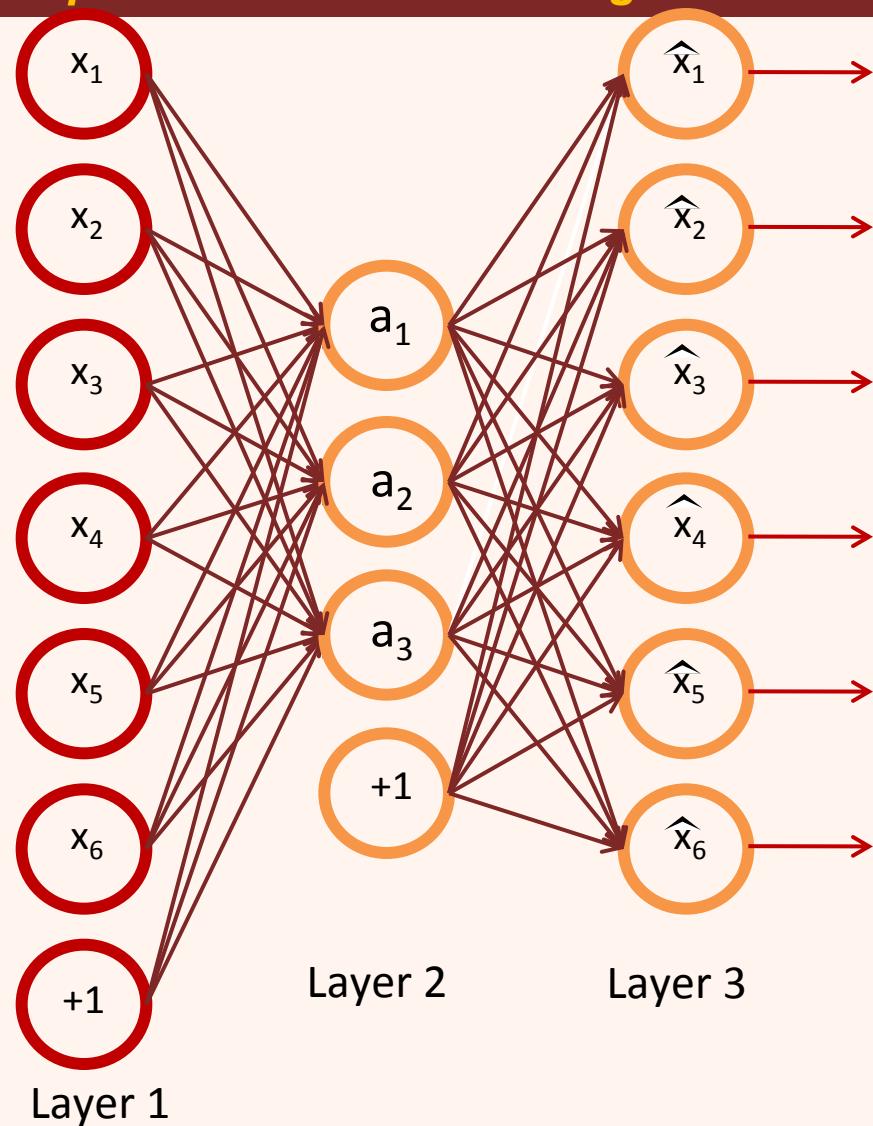
دانشگاه
سینمایی

auto-encoder



auto-encoder

Unsupervised feature learning with a neural network



شبکه عصبی

<http://ufldl.stanford.edu/eccv10-tutorial/>

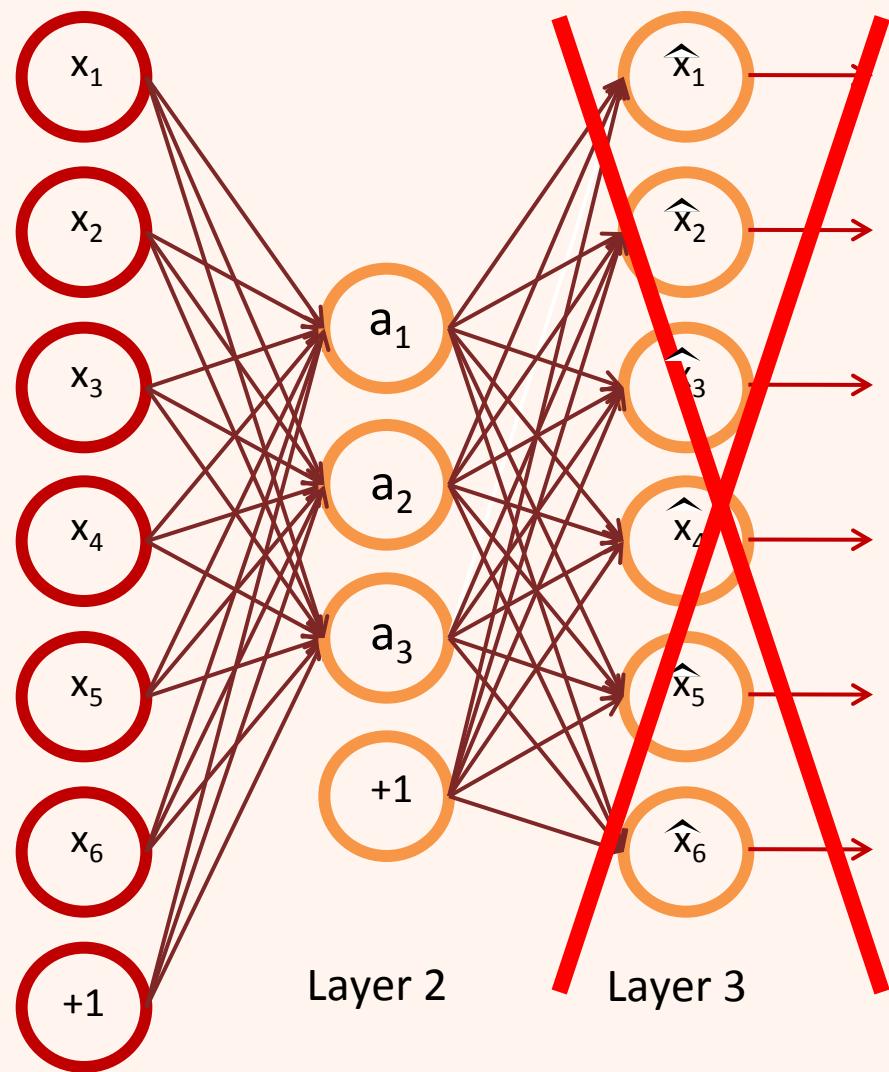
شبکه
گونه‌ای آموزش مبتنی‌نده

$$\hat{x} \simeq$$



دانشگاه
سینه‌پوشی

auto-encoder

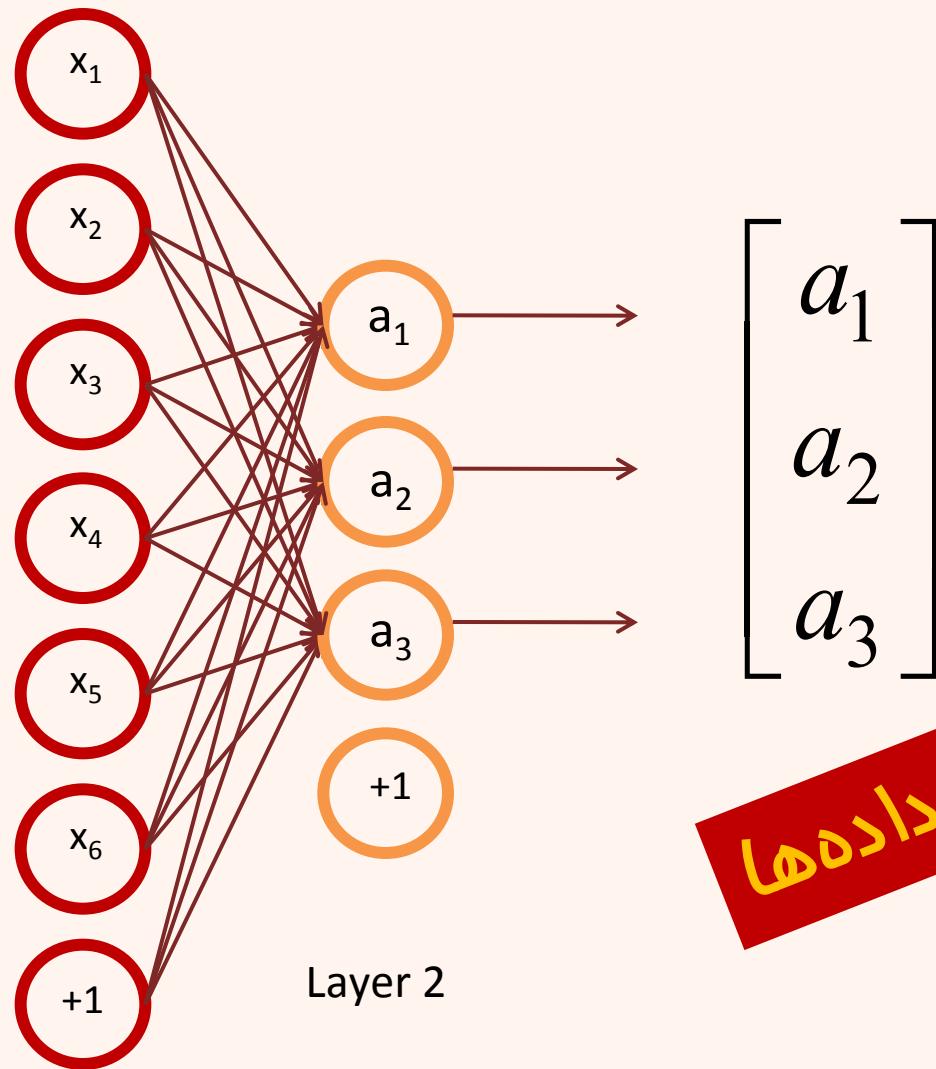


٥٠



دانشکده
سینمایی

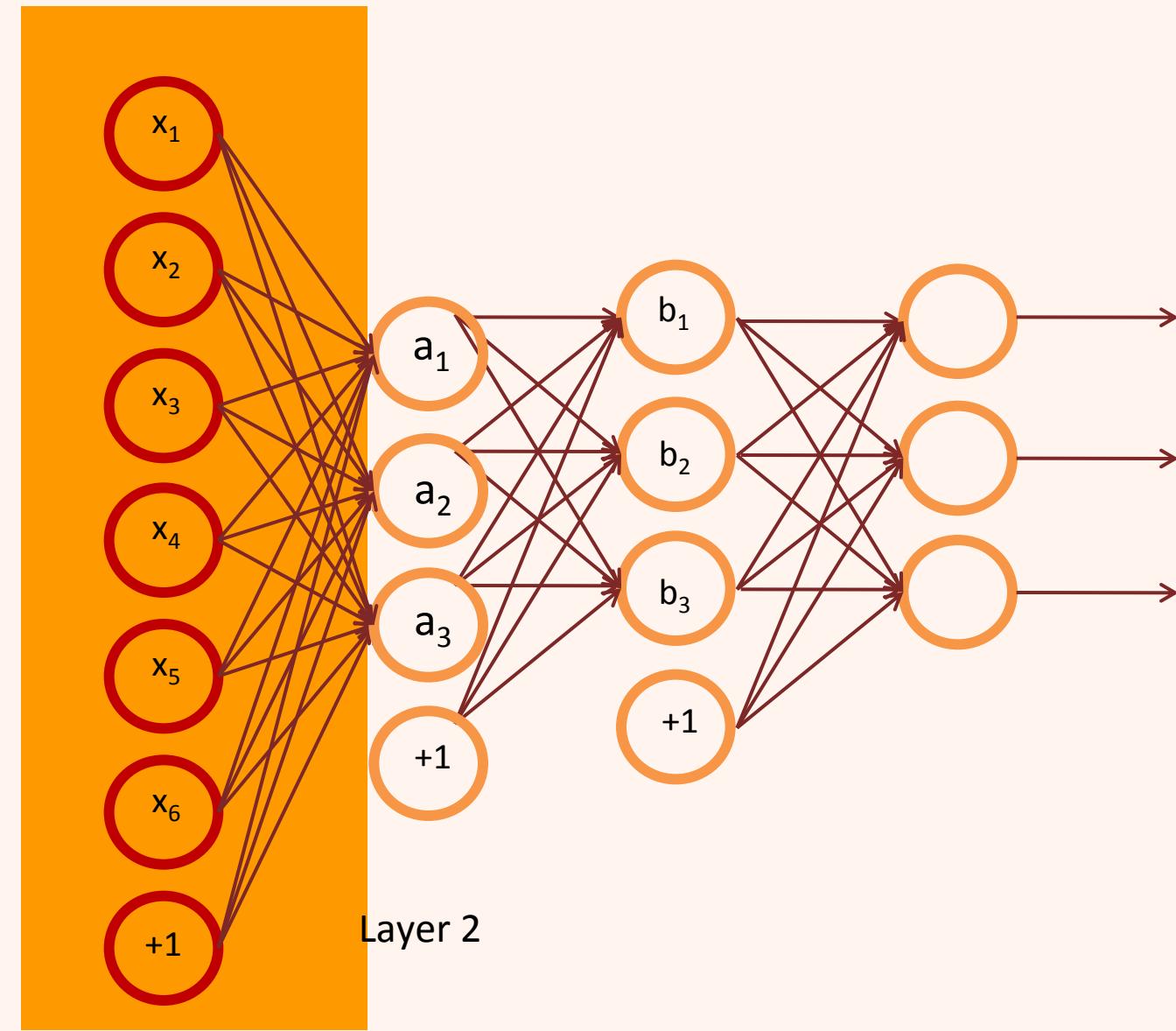
auto-encoder



نمایش جدید برای دادهها



auto-encoder



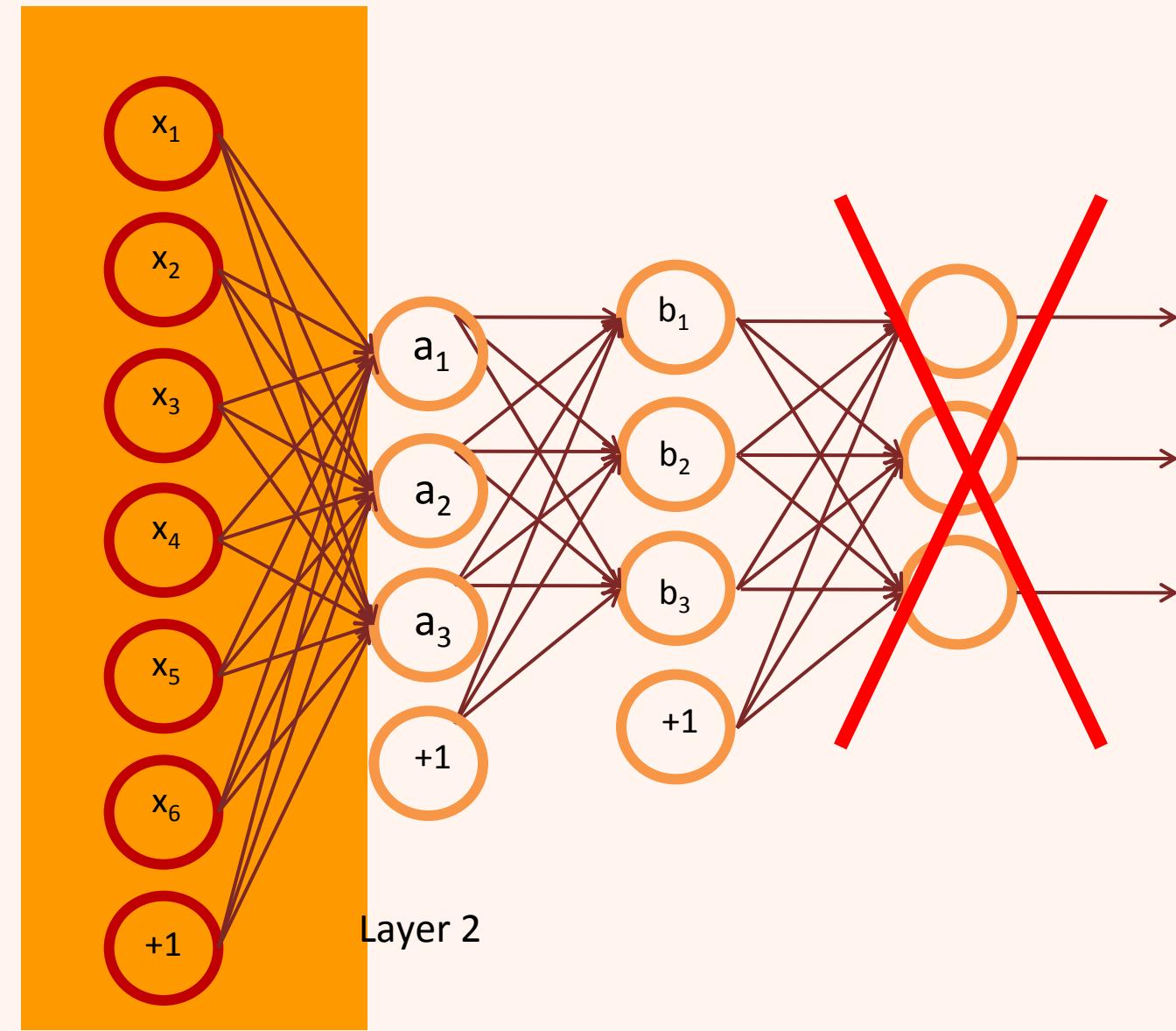
شبکه عصبی

۵۲



دانشگاه
سینمایی

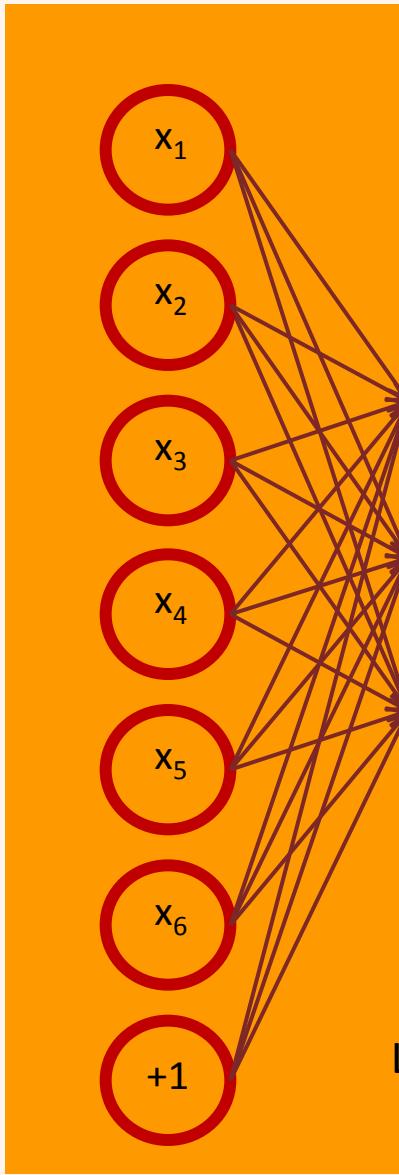
auto-encoder



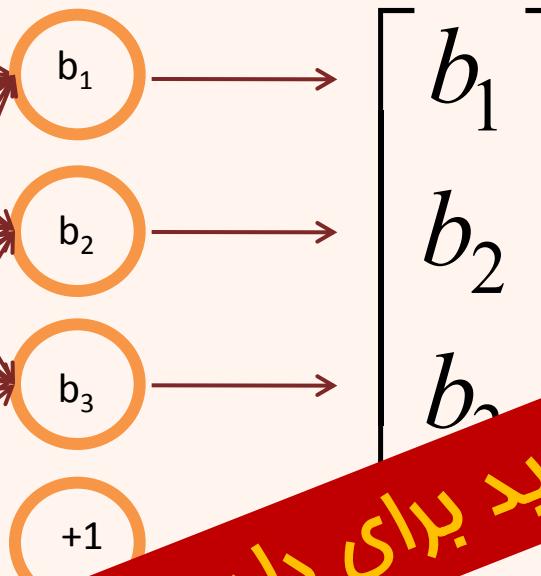
۵۴



auto-encoder



Layer 2

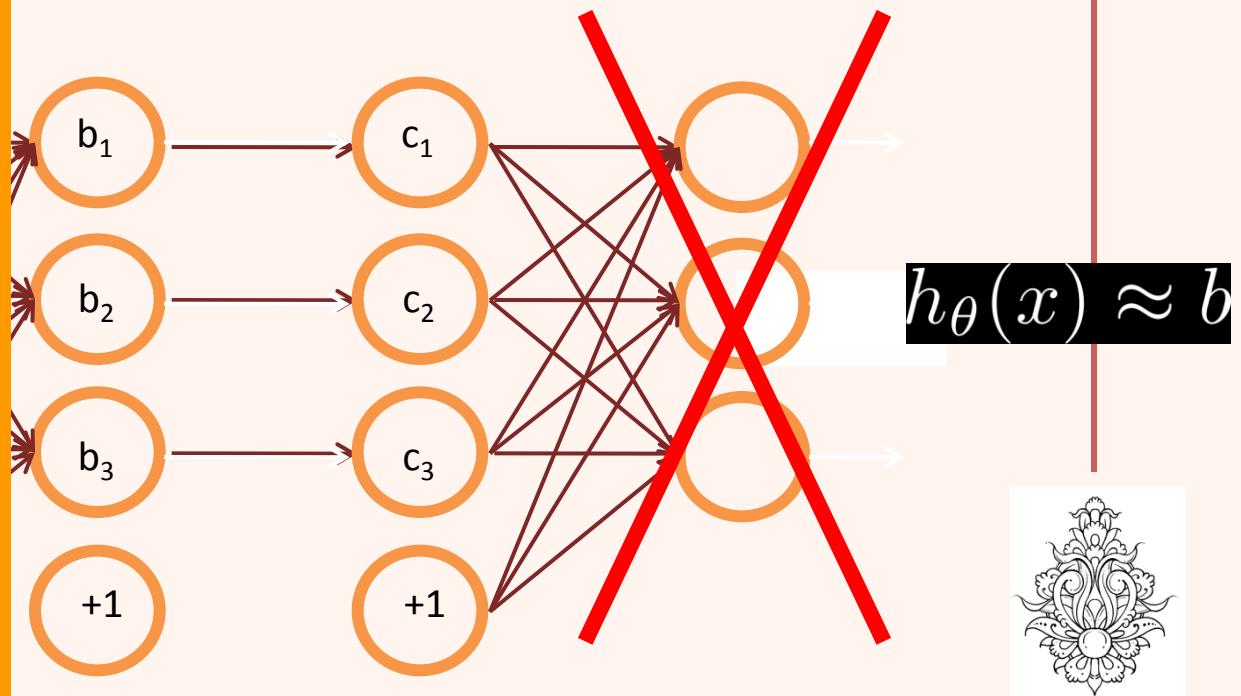


نمایش جدید برای دادهها



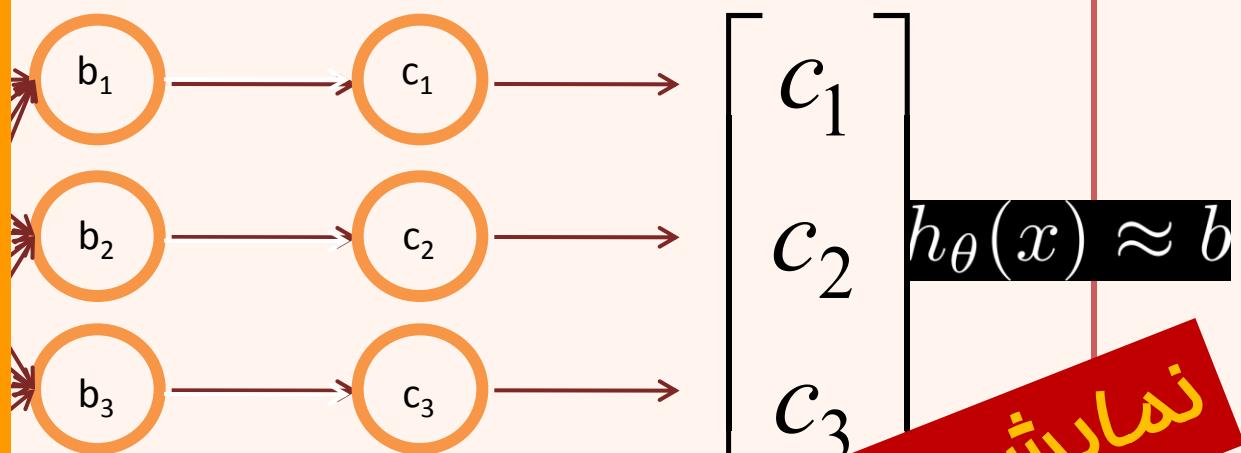
دانشکده
سینمایی
بهشتی

auto-encoder



دانشکده
سینمایی

auto-encoder



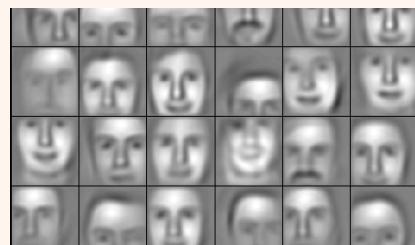
نمایش جدید برای دادهها



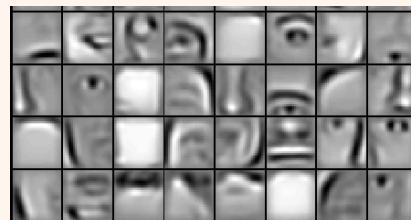
خصوصیتی به دست آمده در لایه‌ی آفر برای دسته‌بندی مورد استفاده قرار می‌گیرد.

آموزش سلسله مراتبی خصیصه ها

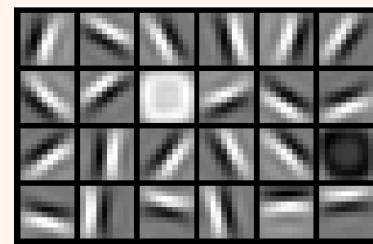
object models



*object parts
(combination
of edges)*



edges



pixels



دانشکده
سینما و
بهاشتی

آموزش سلسله‌مراتبی فضیمها

Faces



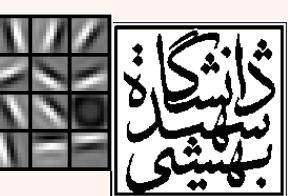
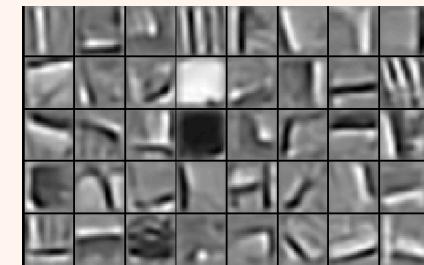
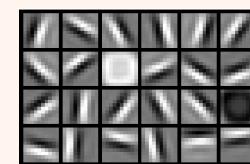
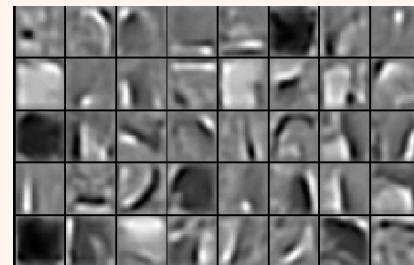
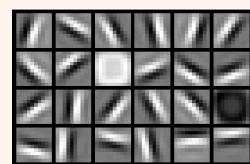
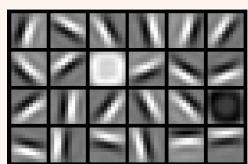
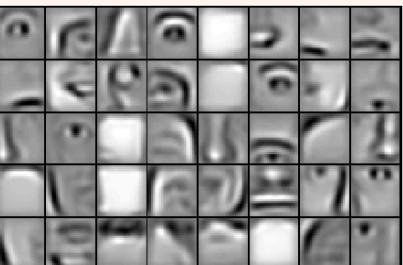
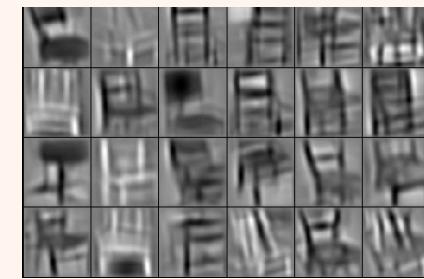
Cars



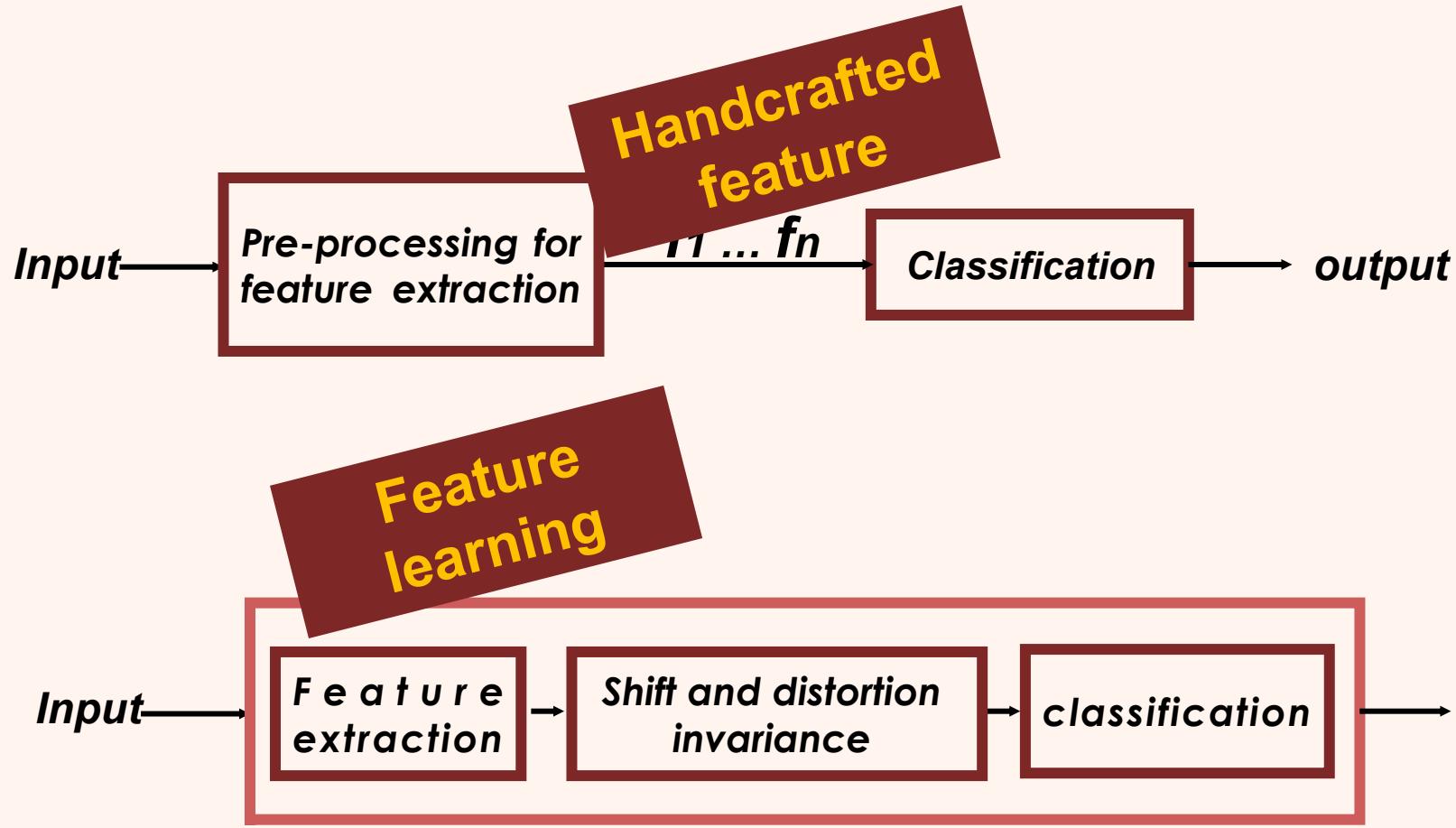
Elephants



Chairs



اسندرائی فصلیں



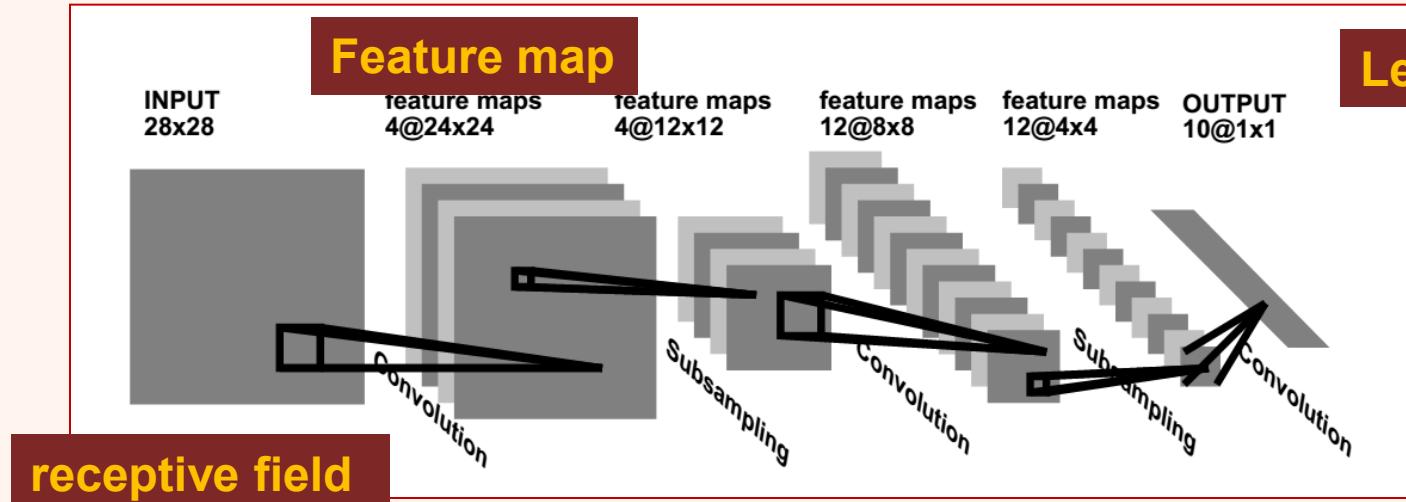
CNN



In 1995, Yann LeCun and Yoshua Bengio introduced the concept of convolutional neural networks.

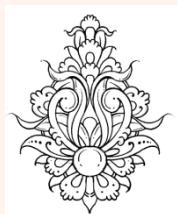


Convolutional Neural Networks



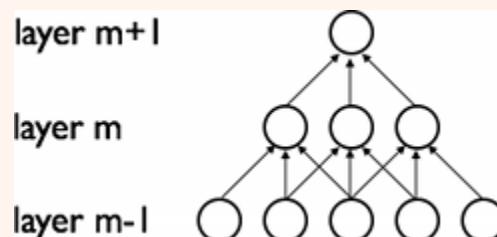
LeNet1(3000/100000)

- این شبکه عمدها برای داده‌های دوبعدی طراحی شده است.
- برای داده‌های مناسب است که به صورت محلی ساختار منظمی دارند.
 - اتصالات بین نوronها به صورت محلی است.



دانشکده
سینمایی
بهرامی

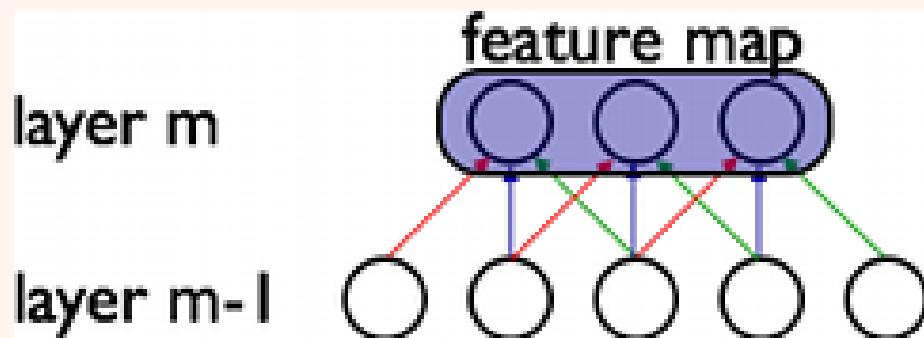
Sparse Connectivity



- با توجه به این که وزن‌های متصول به لایه‌ی feature-map یکسان است، خصیصه‌های به دست آمده نسبت به جابجایی مقاوم است.

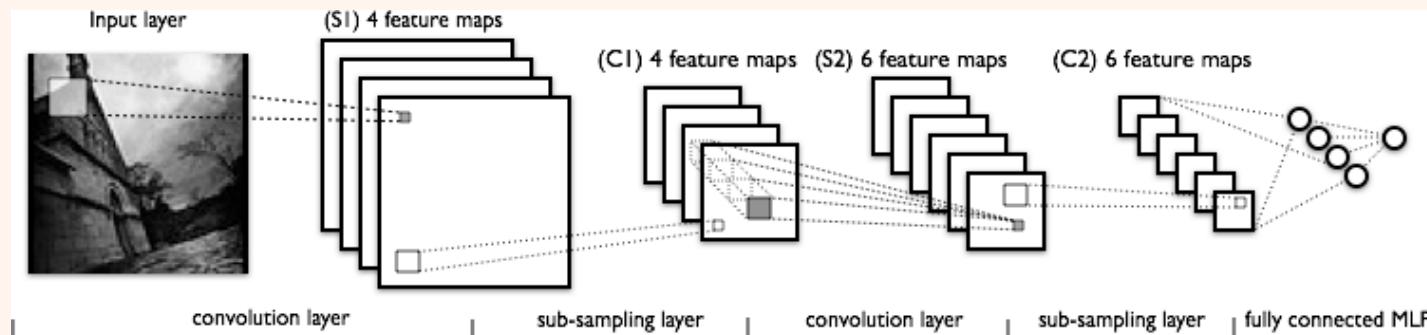
- از سوی دیگر این کار باعث می‌شود پارامترهای آزاد شبکه به شدت کاهش یابد.

Shared Weights



Convolutional Neural Networks

- در این مالت فروجی این لایه در مواردی که یک الگو تکرار شود، مقدار بالایی خواهد بود.
- در لایه‌های بعدی، براساس این خصیصه‌ها، خصیصه‌های جدید استخراج می‌شود.
- با توجه به مشابهت محاسبه‌ی هر گره در این لایه با کانولوشن معمولی، شبکه‌ی کانولوشنی نامیده شده‌اند.



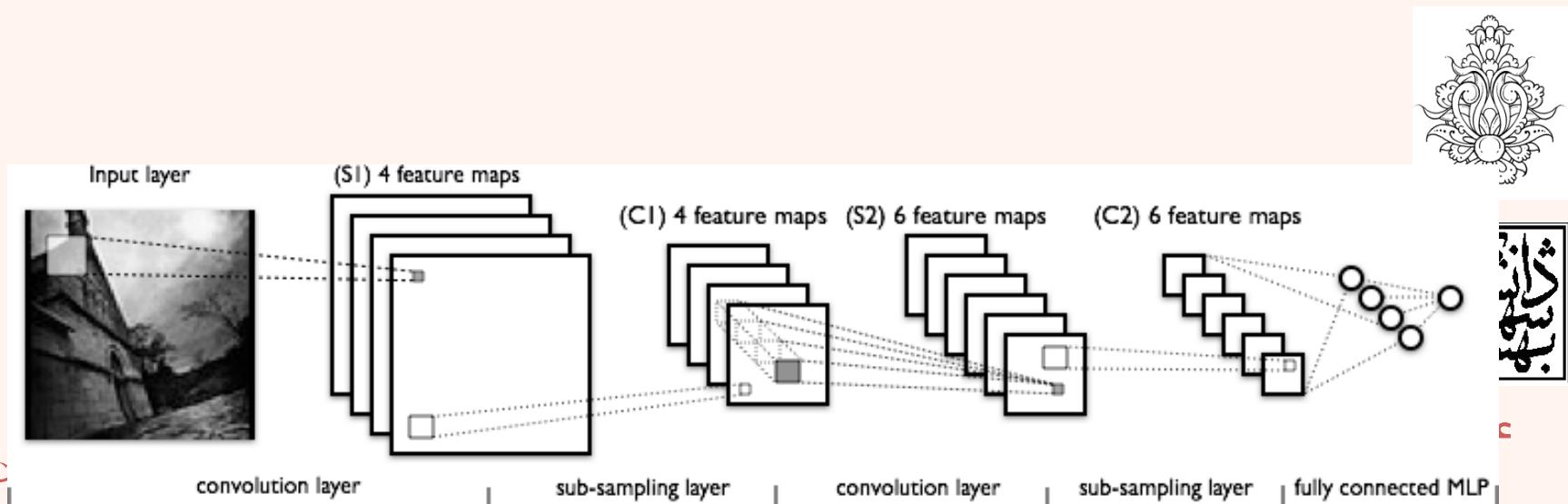
شبکه عصبی



دانشکده
سینمای
بهرستانی

Sub-Sampling (Pooling)

- عملاً لایهی کانولوشن باعث می‌شود تعداد خصیصه‌ها افزایش یابد.
- در این لایه ابعاد با روش‌هایی پون میانگین‌گیری یا جایگزین کردن مقدار بیشینه کاهش می‌یابد.
 - بدین ترتیب وابستگی خصیصه‌ی استخراج شده به مکان کاهش می‌یابد.



LeNet-5(60000/340000)

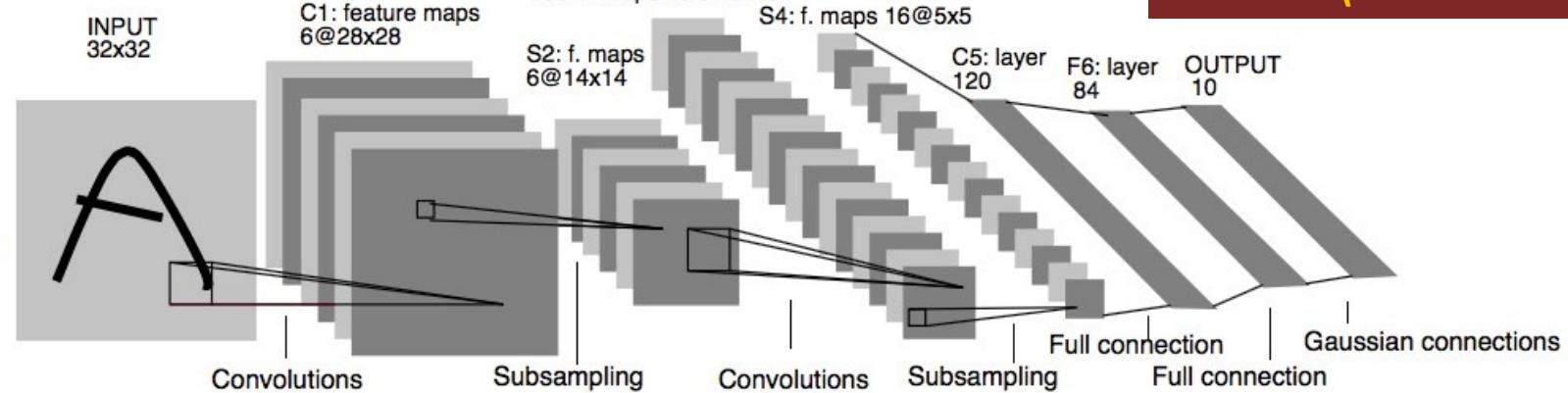


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Layer	Trainable Weights	Connections
C1	$(25+1)*6 = 156$	$(25+1)*6*28*28 = 122,304$
S2	$(1+1)*6 = 12$	$(4+1)*6*14*14 = 5880$ (2x2 links and bias)
C3	$6*(25*3+1) + 9*(25*4+1) + 1*(25*6+1) = 1516$	$1516*10*10 = 151,600$
S4	$16*2 = 32$	$16*5*5*5 = 2000$ (2x2 links and bias)
C5	$120*(5*5*16+1) = 48,120$	Same since fully connected MLP at this point
F6	$84*(120+1) = 10,164$	Same
Output	$10*(84+1) = 850$ (RBF)	Same



دانشگا
سوسیتی
پژوهشی

آموزش(پس انتشار خطا)

در این حالت قانون پس انتشار خطا به صورت زیر اصلاح می شود:

To constrain: $w_1 = w_2$

we need: $\Delta w_1 = \Delta w_2$

compute: $\frac{\square E}{\square w_1}$ and $\frac{\square E}{\square w_2}$

use $\frac{\square E}{\square w_1} + \frac{\square E}{\square w_2}$ for w_1 and w_2

تمام پارامترهای آزاد، با کمک الگوریتم پس انتشار خطا قابل آموزش می باشند.

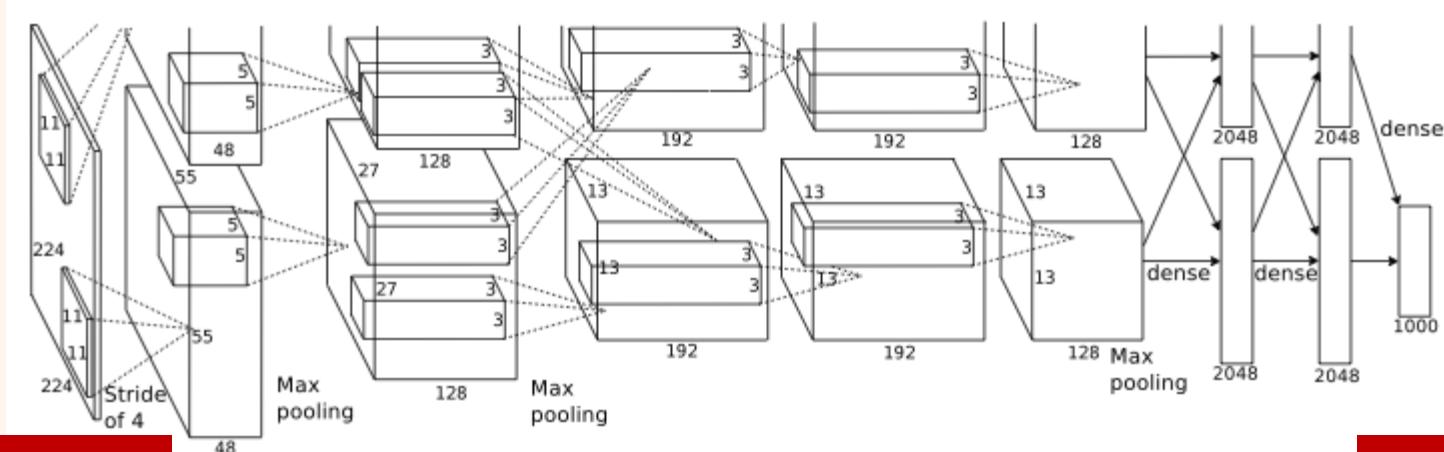


آموزش(پس انتشار خطا)

- استفاده از پیشآموزش به صورت بینظارت و سپس آموزش با الگوریتم پس انتشار خطا می‌تواند باعث افزایش کاری شود.
- بعد از فاز پیشآموزش(auto-encoder)، یک لایه به شبکه برای دسته‌بندی اضافه می‌شود.
- سپس شبکه با استفاده از شیوه‌ی بانظارت همه‌ی وزن‌ها آموزش می‌بینند.



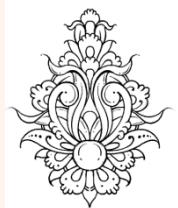
دانشکده
سینما
و تئاتر



AlexNet شبکه

LSVRC-2010

برای دسته‌بندی ۱.۲ میلیون تصویر
۱۰۰۰ دسته‌ی مختلف
۶۵۰۰۰ نورون و ۴۰ میلیون پارامتر آزاد
پنج لایه‌ی کانولوشن و سه لایه‌ی متصل
top1 و top5 و ۳۷ و ۱۷ درصد خطای
از drop out استفاده می‌گردد.



دانشکده
سینما و تئاتر
بهشتی

فریب شبکه‌های عمیق



We demonstrated that deep neural networks have counter-intuitive properties both with respect to the semantic meaning of individual units and with respect to their discontinuities(Szegedy2013).

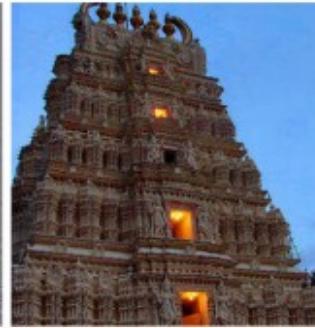
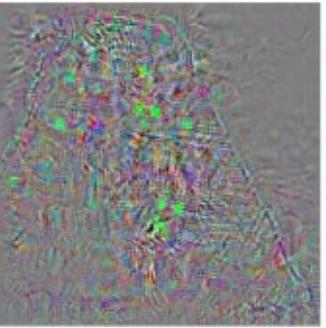
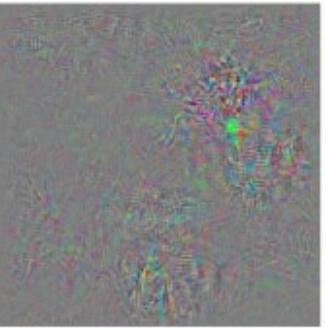
adversarial examples



Differences



image predicted incorrectly



correctly predicted sample



دانشکده
سینمایی
بهریتی

Szegedy, C., et al., *Intriguing properties of neural networks.*
arXiv preprint arXiv:1312.6199, 2013.

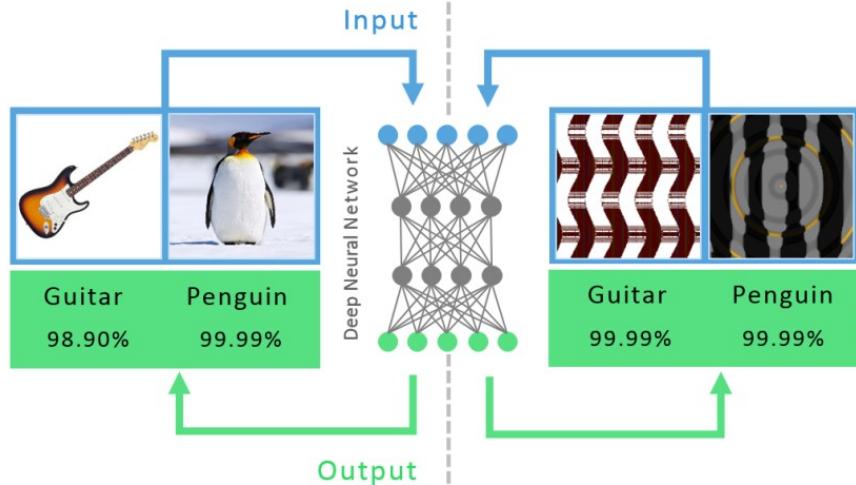
v.



فریب شبکه‌های عصبی عمیق

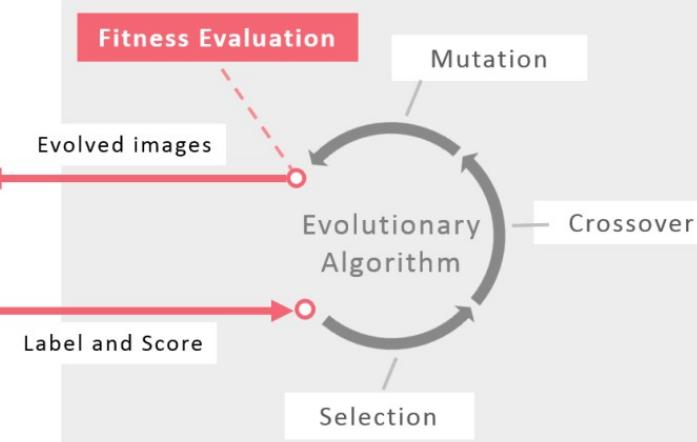
1

State-of-the-art DNNs can recognize real images with high confidence



2

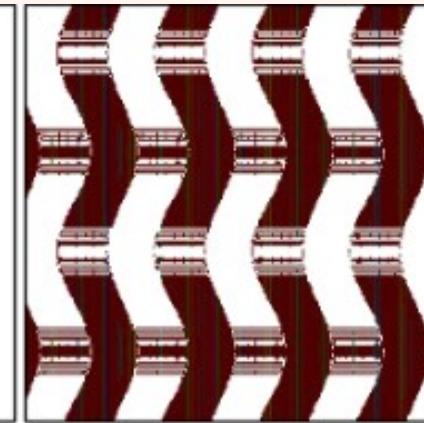
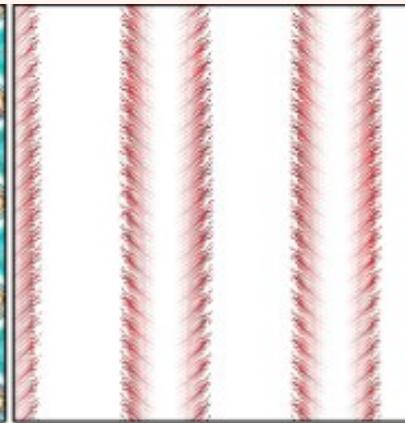
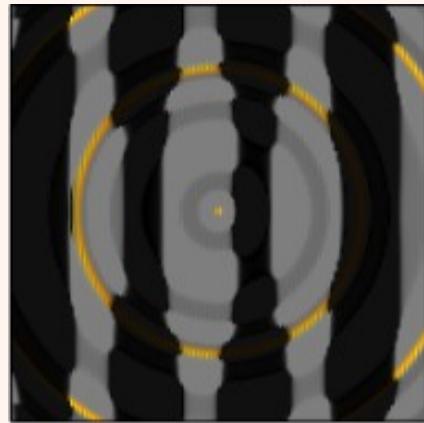
But DNNs are also easily fooled: images can be produced that are unrecognizable to humans, but DNNs believe with 99.99% certainty are natural objects



Nguyen, A., J. Yosinski, and J. Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.



فریب شبکه‌های عصبی عمیق

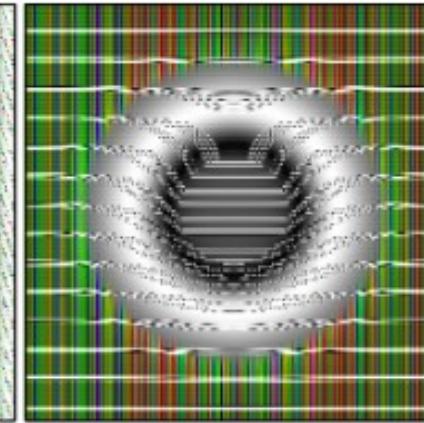
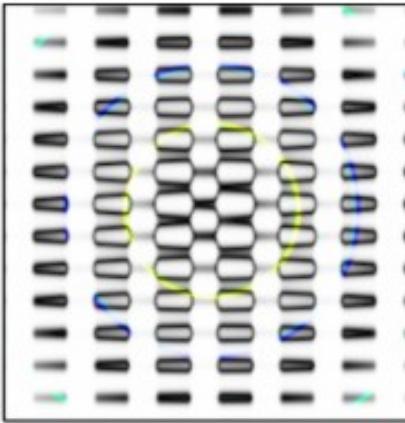
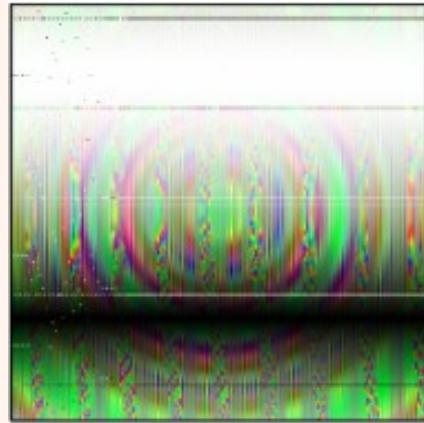


king penguin

starfish

baseball

electric guitar



freight car

remote control

peacock

African grey



دانش
سینمایی
بهشتی

منابع

- Raina, R., et al., *Self-taught learning: transfer learning from unlabeled data*, in *Proceedings of the 24th international conference on Machine learning*. 2007, ACM: Corvalis, Oregon, USA. p. 759-766.
- Szegedy, C., et al., *Intriguing properties of neural networks*. arXiv preprint arXiv:1312.6199, 2013.
- Nguyen, A., J. Yosinski, and J. Clune. *Deep neural networks are easily fooled: High confidence predictions for unrecognizable images*. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
- <https://www.macs.hw.ac.uk/~dwcorne/Teaching/dmml.htm>



۷۱۶