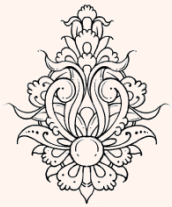
شناسایی آماری الگو
بخش دهم
(۴۱-۱۰-۷۱۱-۰۱)

مقدمه‌ای بر یادگیری عمیق
DNN

دانشگاه شهید بهشتی
پژوهشکده‌ی فضای مجازی
بهار ۱۳۹۶
احمد محمودی ازناوه

فهرست مطالب

- مقدمه‌ای بر یادگیری عمیق
- یادگیری خودآموخته
- یادگیری سلسله مراتبی خصیصه
- آشنایی با CNN
- مشکلات شبکه‌های عمیق



معرفی

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

- «یادگیری عمیق» باعث ایجاد پیشرفت‌های چشمگیری در زمینه‌ی 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 Hub5'00-SWB	WER RT03S-FSH
GMM, 40 mix DT 309h SI	29.4	23.6	27.4
NN 1 hidden-layer×4634 units	43.6	26.0	29.4
+ 2×5 neighboring frames	45.1	22.4	25.7
DBN-DNN 7 hidden layers×2048 units	45.1	17.1	19.6
+ updated state alignment	45.1	16.4	18.6
+ sparsification	15.2 nz	16.1	18.5
GMM 72 mix DT 2000h SA	102.4	17.1	18.6

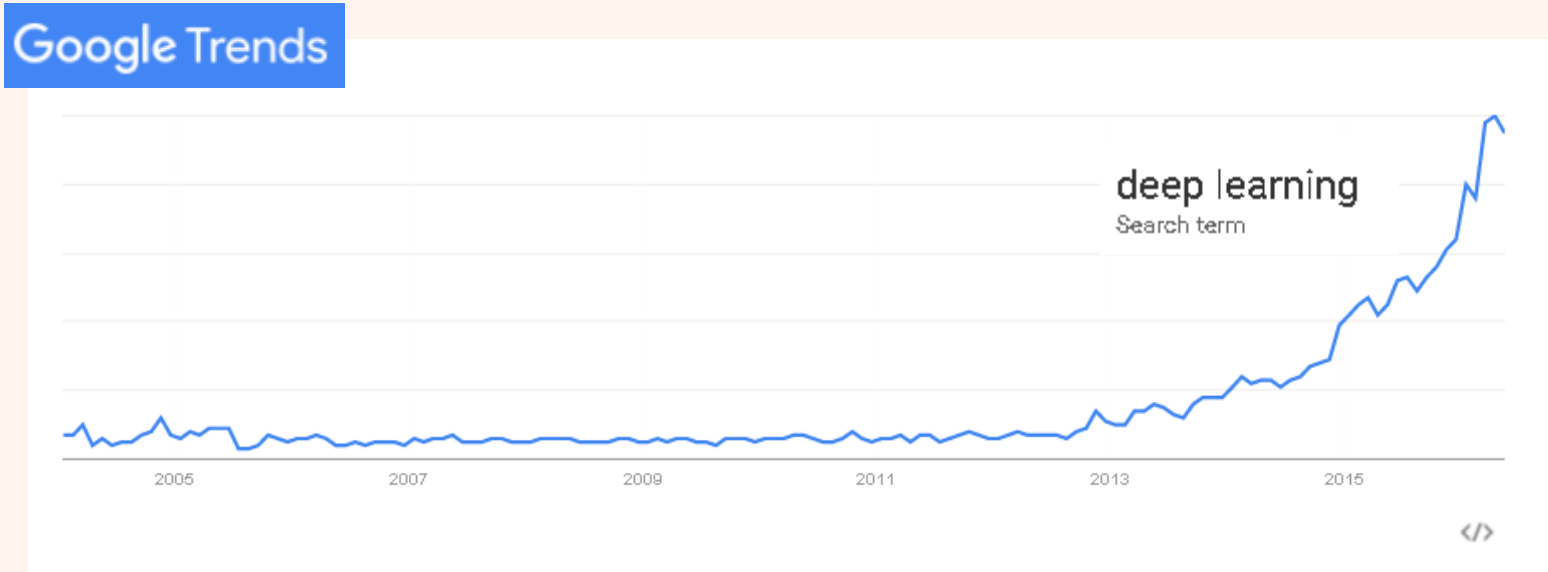
task	hours of training data	DNN-HMM	GMM-HMM with same data	GMM-HMM with more data
Switchboard (test set 1)	309	18.5	27.4	18.6 (2000 hrs)
Switchboard (test set 2)	309	16.1	23.6	17.1 (2000 hrs)
English Broadcast News	50	17.5	18.8	
Bing Voice Search (Sentence error rates)	24	30.4	36.2	
Google Voice Input	5,870	12.3		16.0 (>>5,870hrs)
Youtube	1,400	47.6	52.3	

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

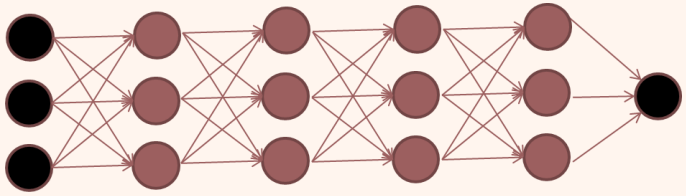
یادگیری عمیق



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

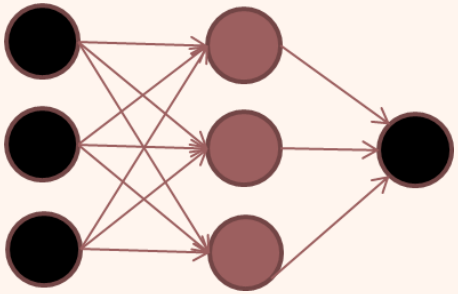
یادگیری عمیق چیست؟

- یک شبکه‌ی عصبی با چندین لایه بین ورودی و خروجی
- این شبکه‌های از نحوه‌ی پردازش اطلاعات در مغز 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!

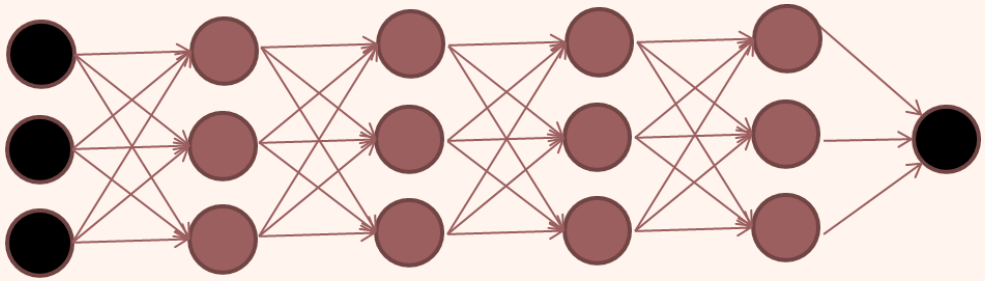


یادگیری عمیق

- شبکه‌ی عصبی با تعداد لایه‌های مخفی محدود به راحتی آموزش می‌بینند:

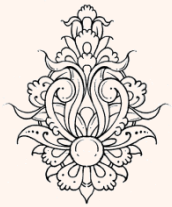


- اما در صورت افزایش تعداد لایه‌ها، الگوریتم‌های یادگیری کارایی لازم را نخواهند داشت.



یادگیری عمیق

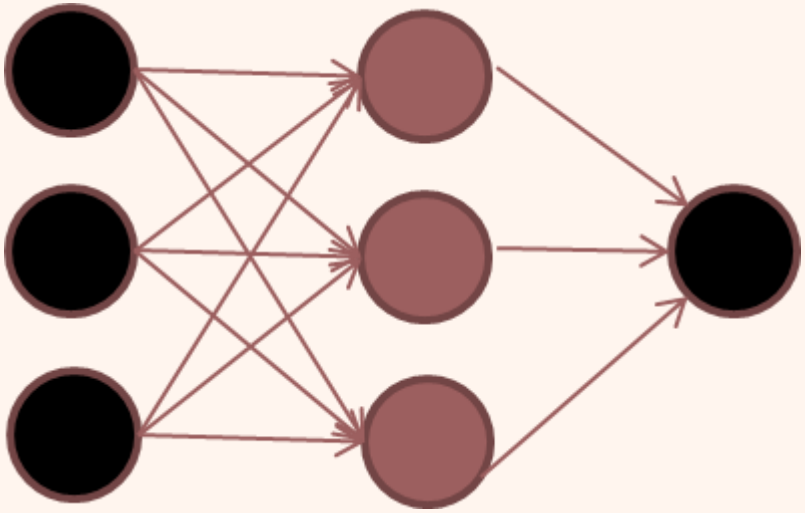
- در واقع، مهمترین تفاوت فوت و فن‌هایی است که برای آموزش این شبکه‌ها به کار گرفته شده است.
- یکی از این شیوه‌ها یادگیری بی‌نظارت خصیصه‌هاست. ***unsupervised feature learning***
- کنترل پارامترهای آزاد، نیز یکی از شیوه‌های مورد استفاده است.



آموزش در شبکه‌های چند لایه

مجموعه‌ی داده‌ها

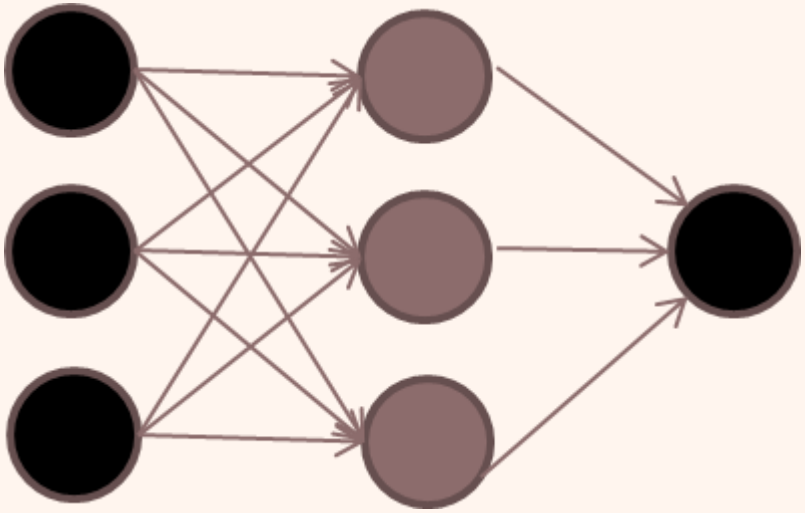
Fields			*class*
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			



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

آموزش در شبکه‌های چند لایه (ادامه...)

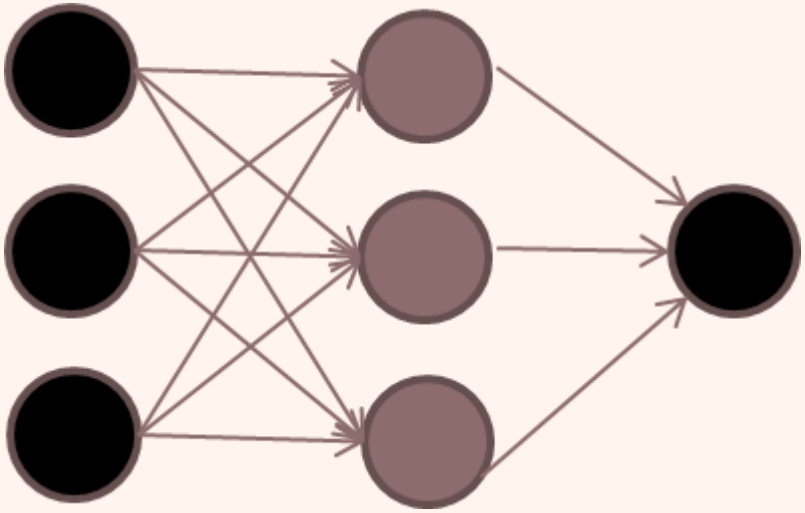
Fields			*class*
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			



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 …			

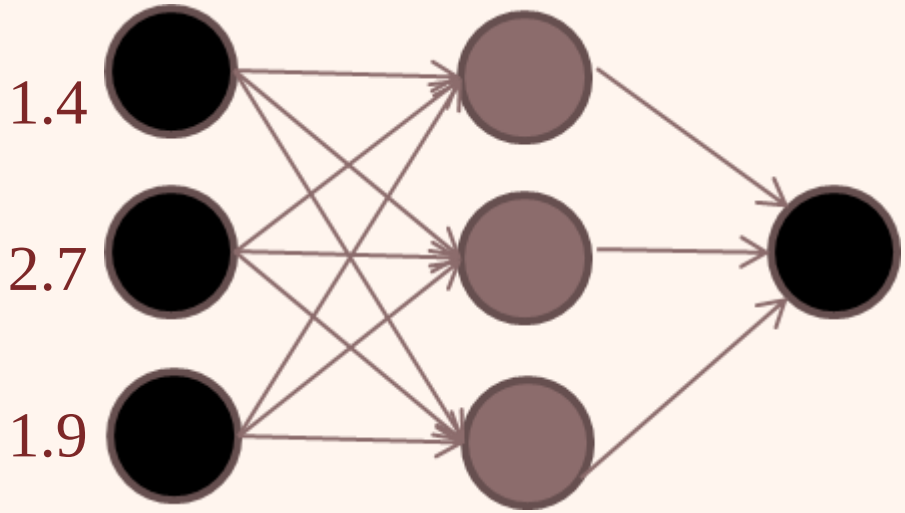
مقداردهی اولیه وزن‌ها به صورت تصادفی



آموزش در شبکه‌های چند لایه (ادامه...)

اعمال ورودی‌ها

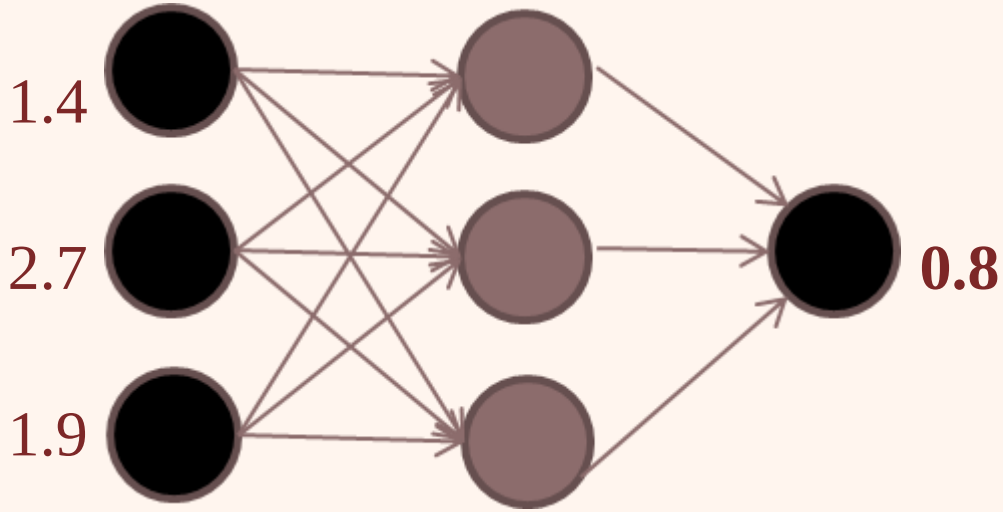
Fields			*class*
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			



آموزش در شبکه‌های چند لایه (ادامه...)

محاسبه‌ی خروجی

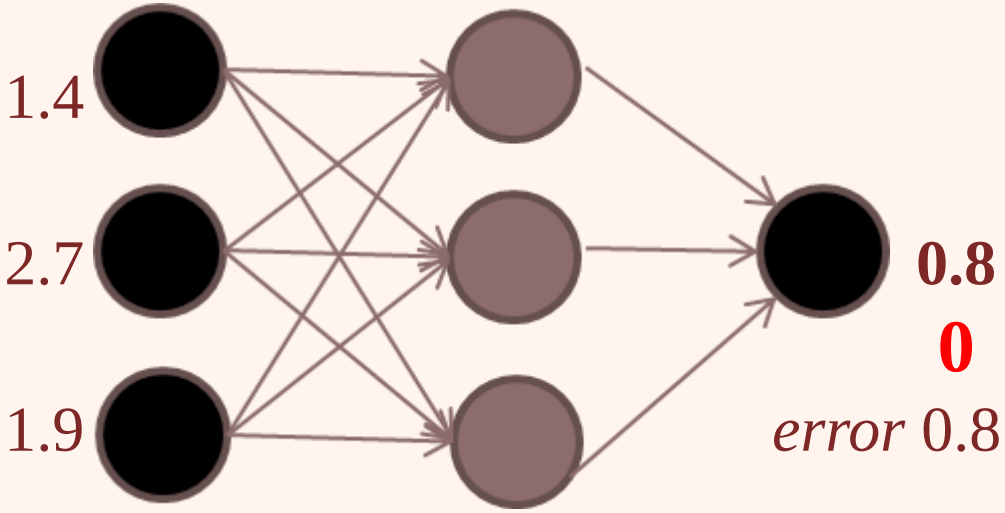
Fields			*class*
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			



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 …			

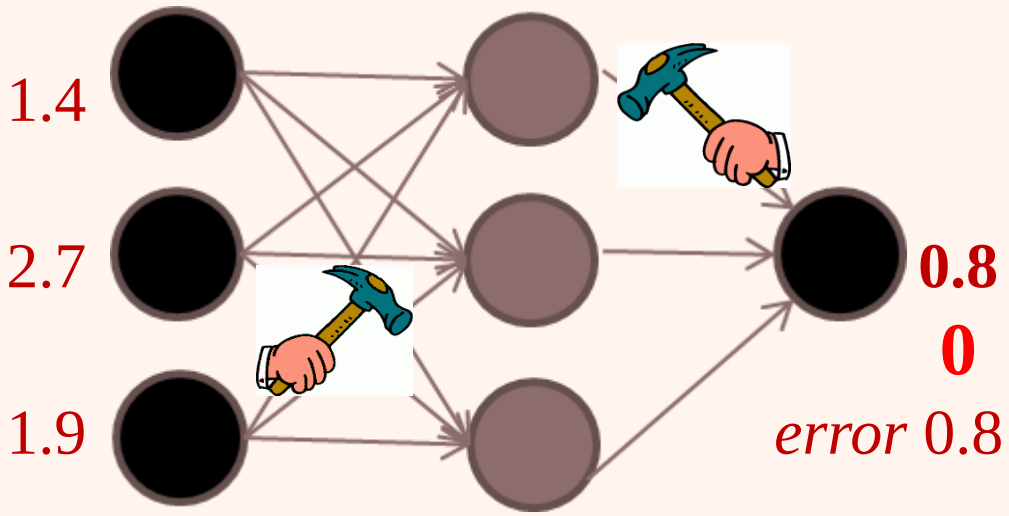
مقایسه با مقدار مطلوب



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 …			

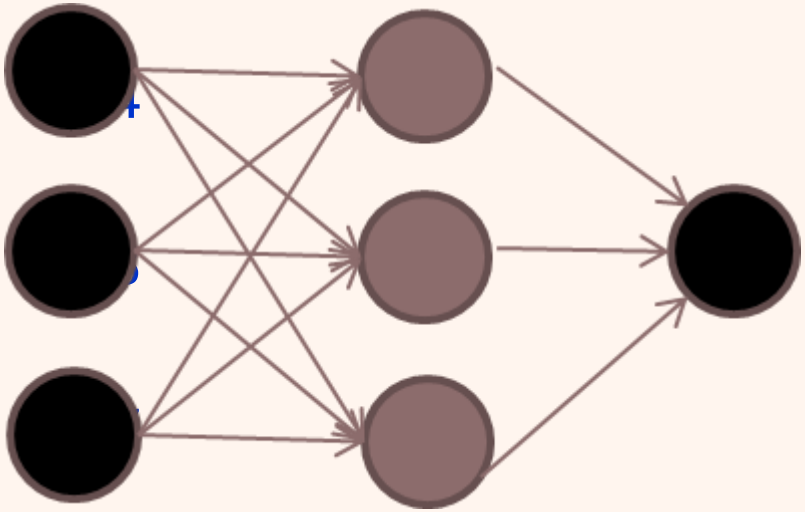
تنظیم وزن‌ها براساس خطا



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 …			

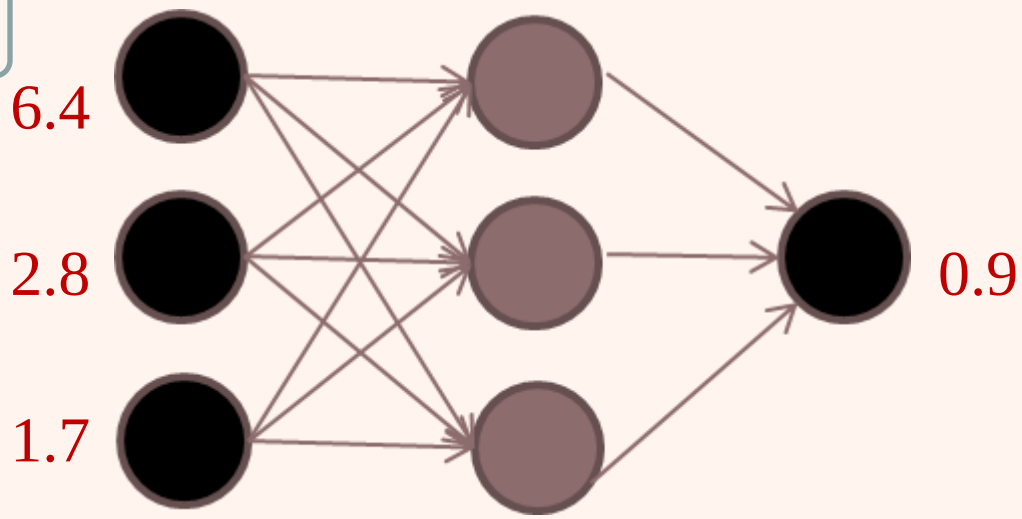
اعمال ورودی‌ها



آموزش در شبکه‌های چند لایه (ادامه...)

محاسبه‌ی خروجی

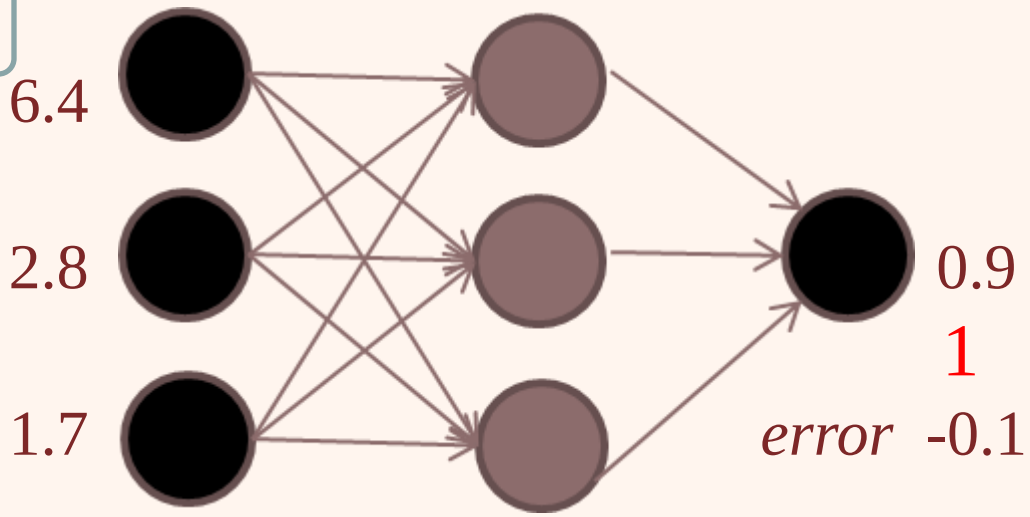
Fields			*class*
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			



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 …			

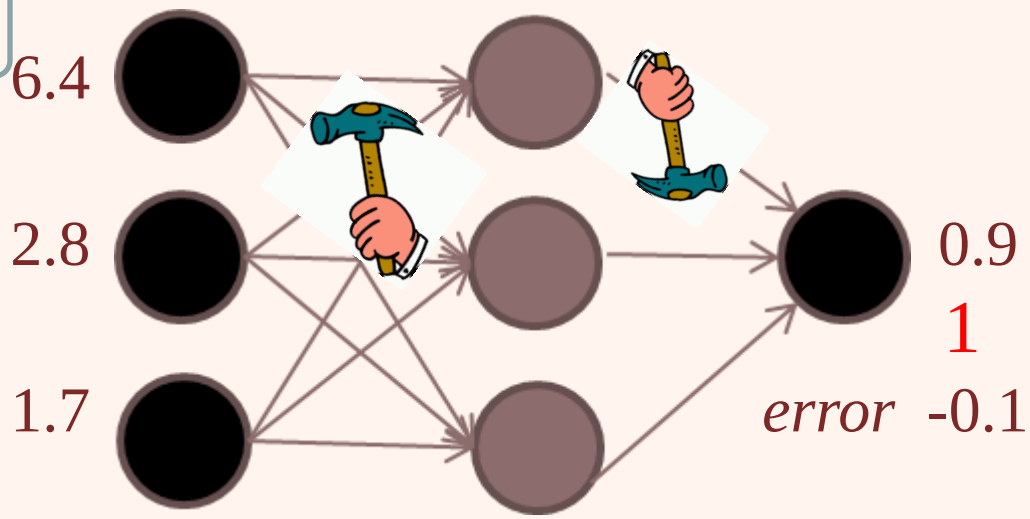
مقایسه با مقدار مطلوب



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 …			

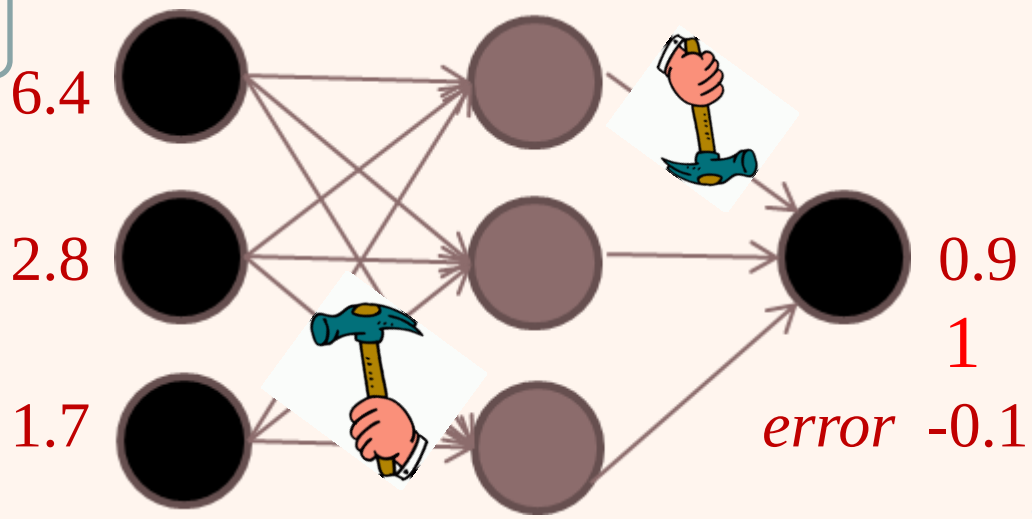
تنظیم وزن‌ها براساس خطا



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 …			

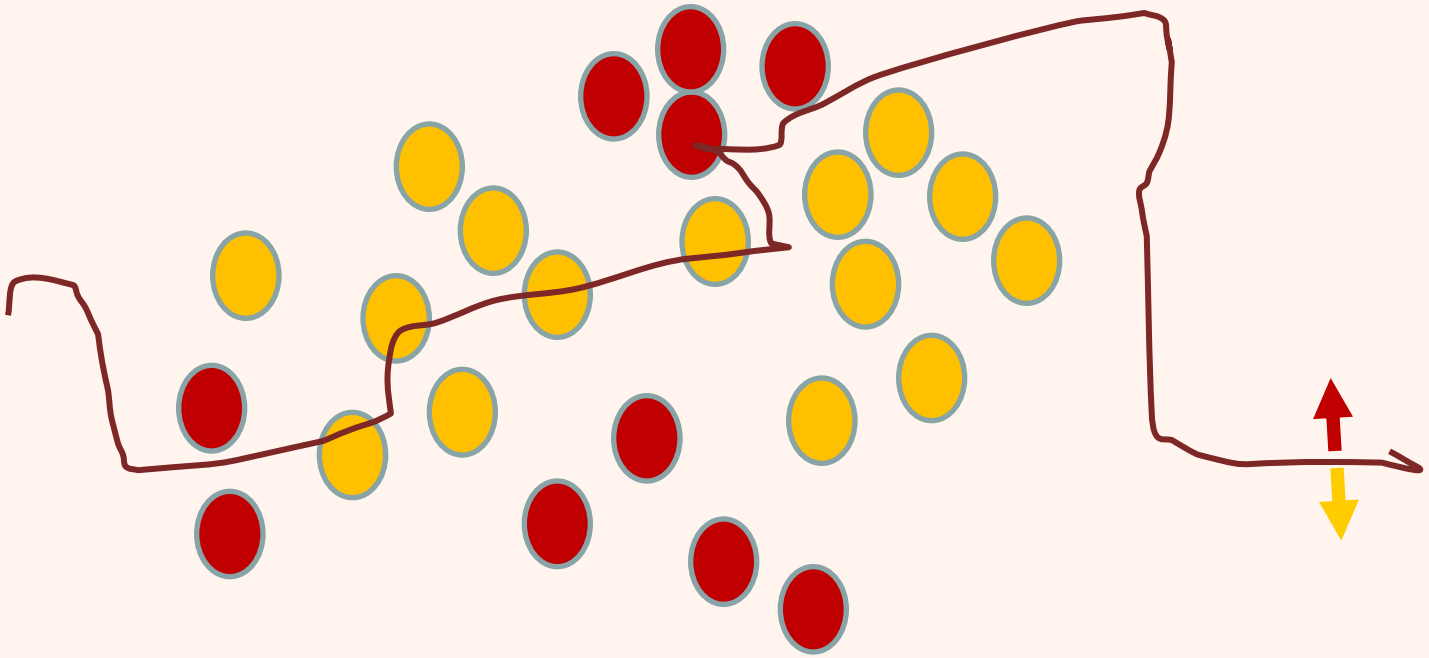
و به همین‌ترتیب ....



این روند هزاران بار و بلکه میلیون‌ها بار(بسته به پیچیدگی مدل) تکرار می‌شود.
در روش ترتیبی هربار داده‌ای به صورت تصادفی انتخاب شده و به شبکه اعمال می‌شود.
وزن‌ها به گونه‌ای تنظیم می‌شوند که خطا کمینه شود.
تعداد زیاد پارامترهای آزاد می‌تواند منجر به کند شدن شبکه شود.

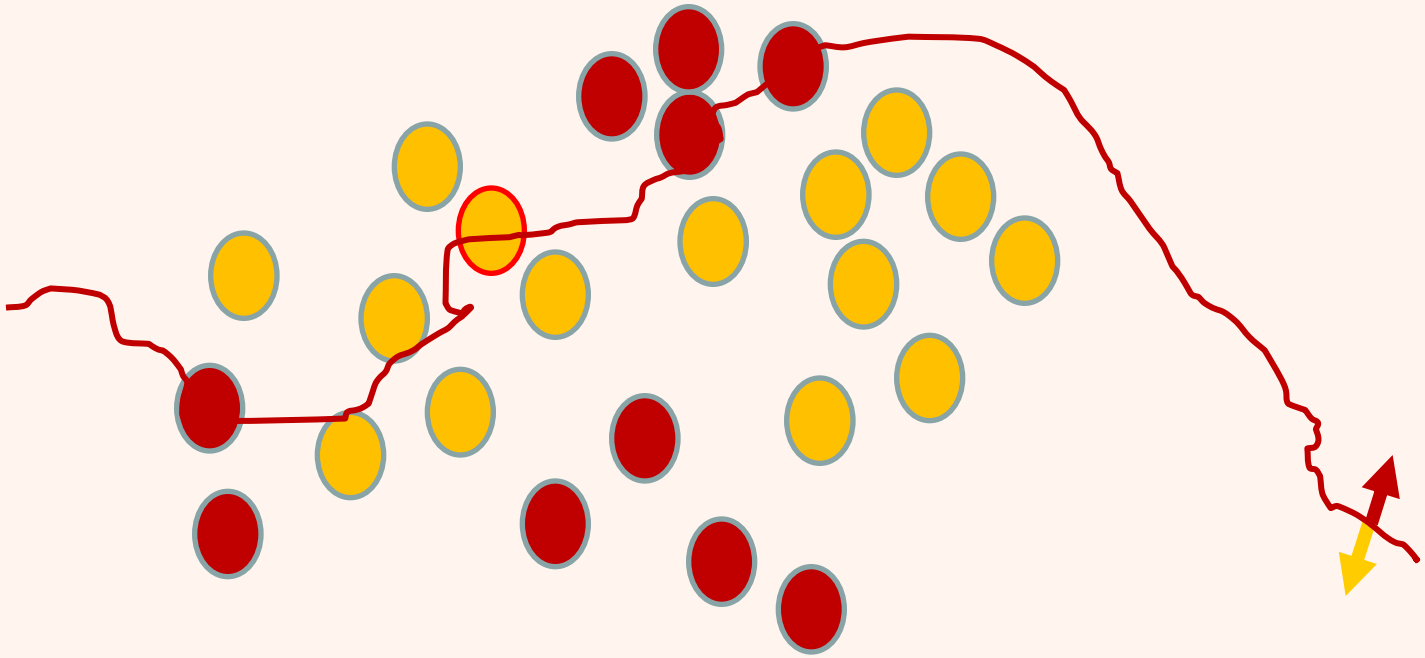
شبکه‌های پیچیده و مرز جداساز

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



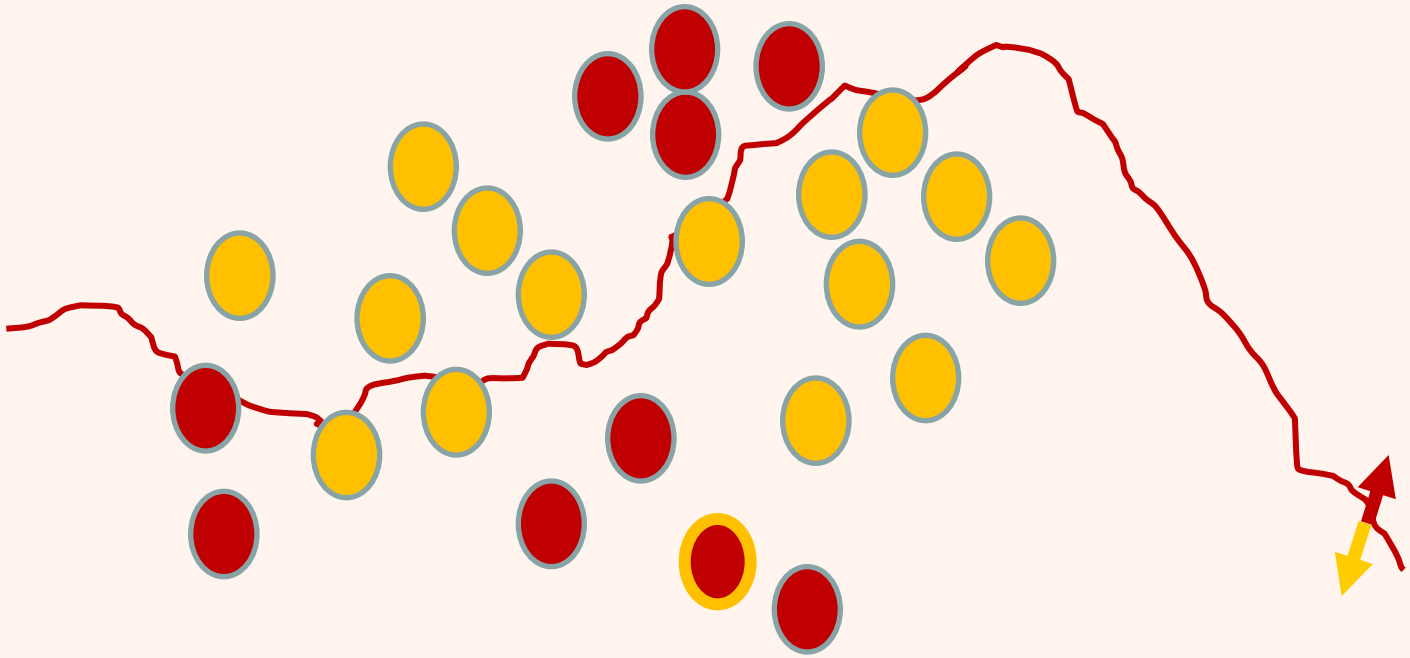
وزن‌های تصادفی اولیه

شبکه‌های پیچیده و مرز جداساز (ادامه...)



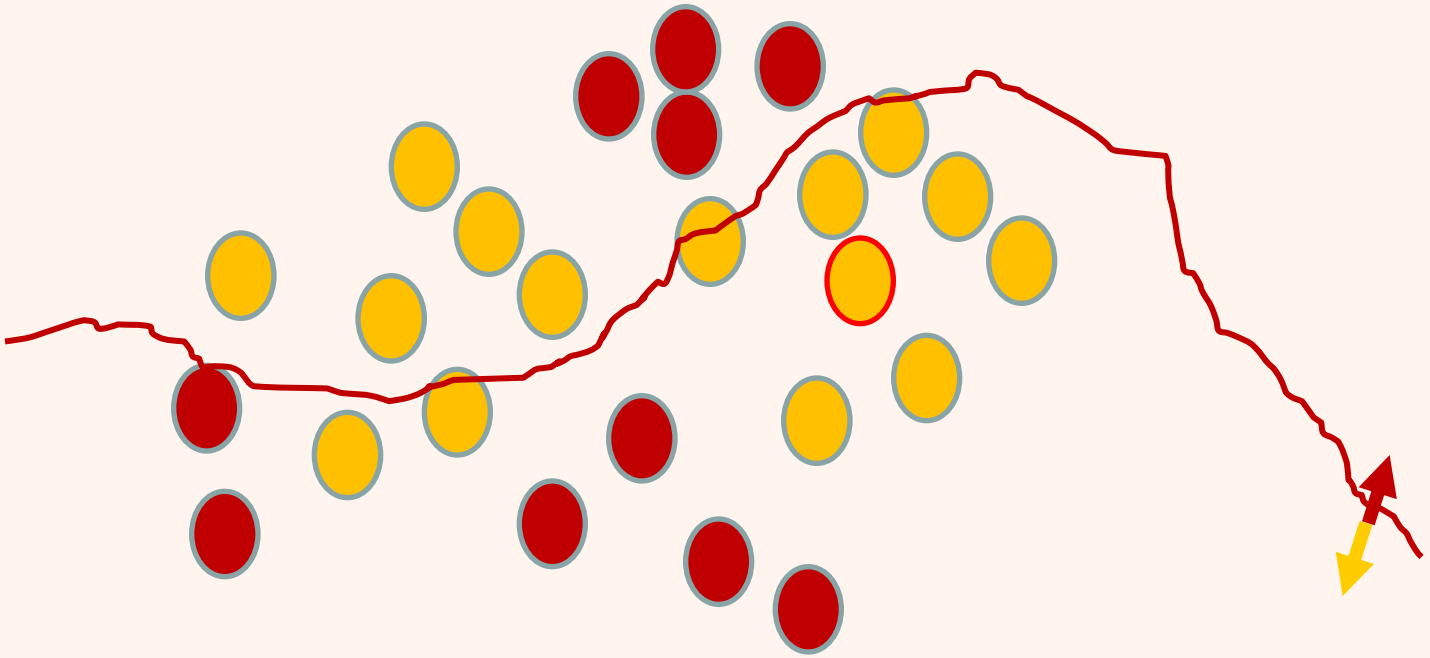
یک نمونه انتخاب و بر اساس آن وزن‌ها تنظیم می‌شوند

شبکه‌های پیچیده و مرز جداساز (ادامه...)



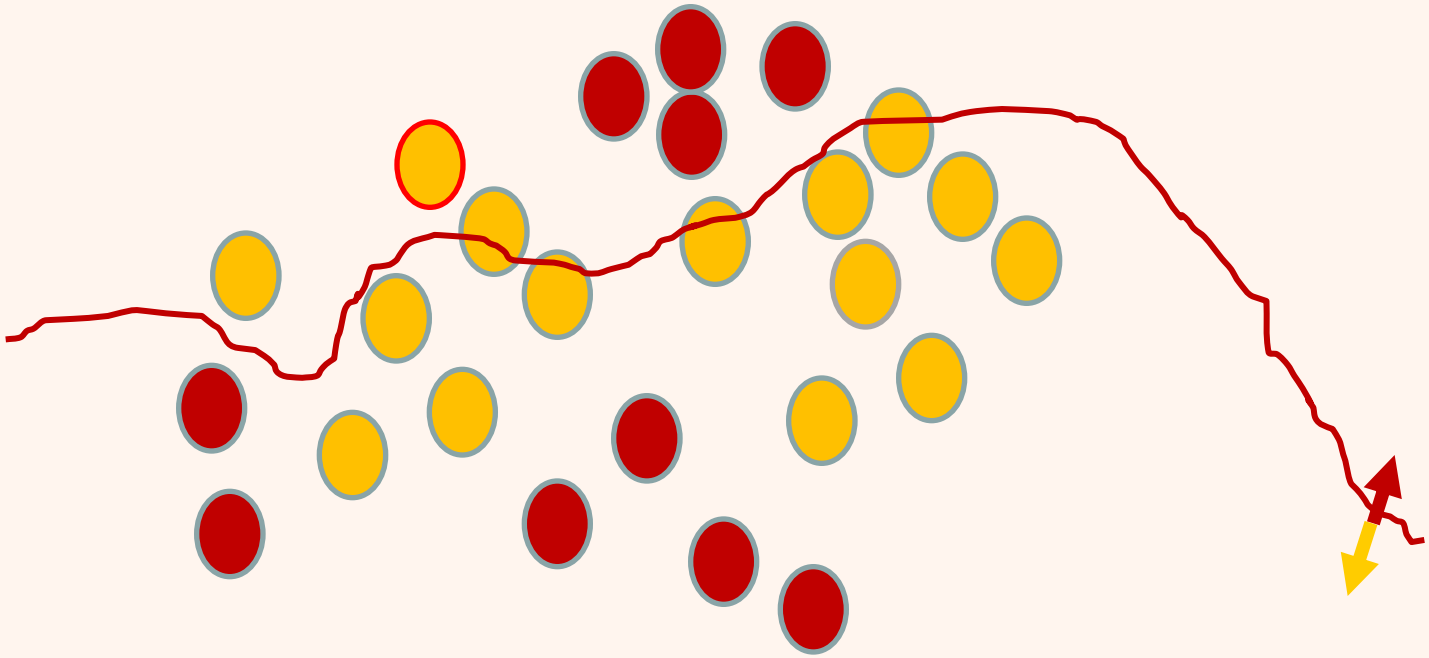
یک نمونه انتخاب و بر اساس آن وزن‌ها تنظیم می‌شوند

شبکه‌های پیچیده و مرز جداساز (ادامه...)



یک نمونه انتخاب و بر اساس آن وزن‌ها تنظیم می‌شوند

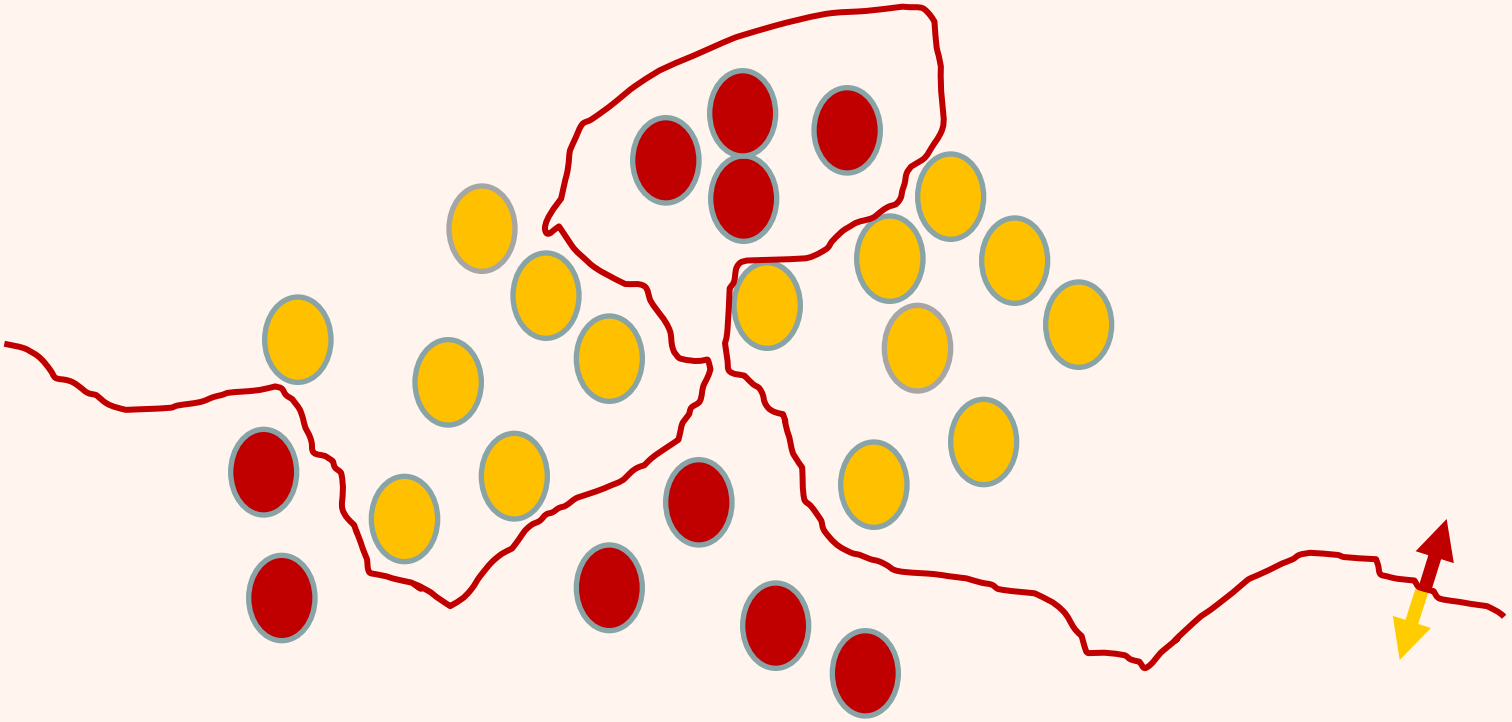
شبکه‌های پیچیده و مرز جداساز (ادامه...)



یک نمونه انتخاب و بر اساس آن وزن‌ها تنظیم می‌شوند

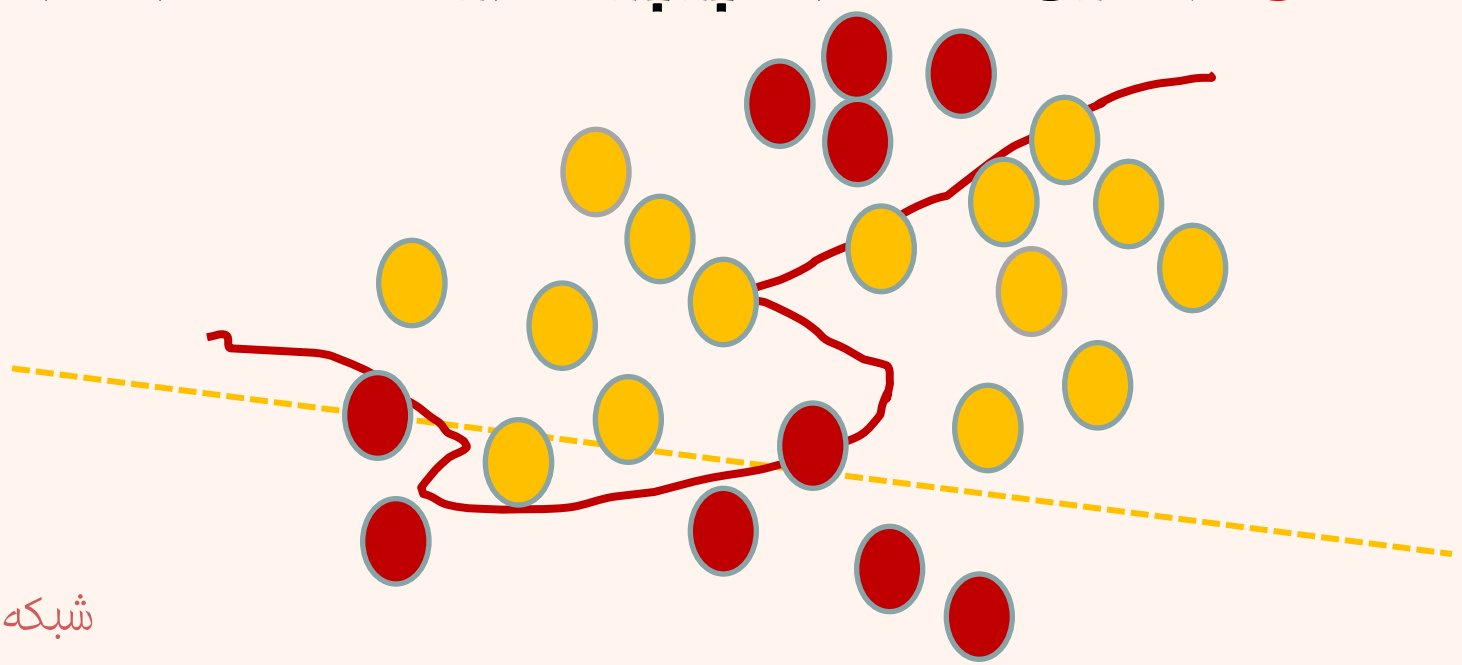
شبکه‌های پیچیده و مرز جداساز (ادامه...)

و درنهایت:

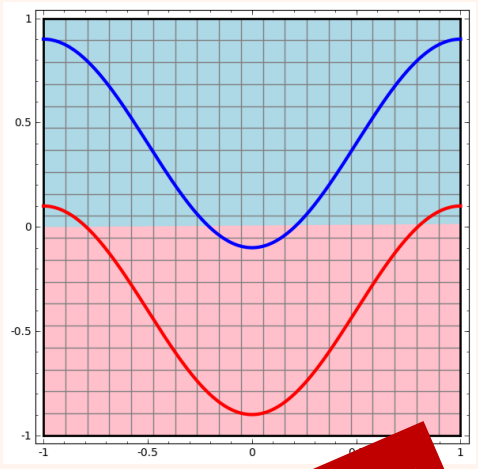


چند نکته

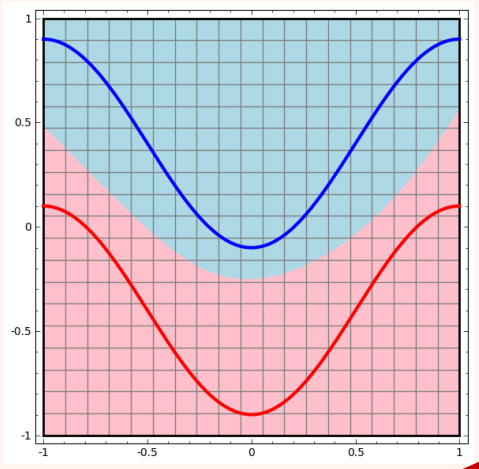
- وزن‌های آموزش دیده، تفسیرپذیر نیستند.
- یک شبکه‌ی تک‌لایه توانایی دسته‌بندی داده‌های جدایی‌پذیر خطی را دارد.
- می‌دانیم که یک شبکه‌ی عصبی با تنها یک لایه‌ی مخفی توانایی جداسازی پیچیده‌ترین اشکال را دارد.



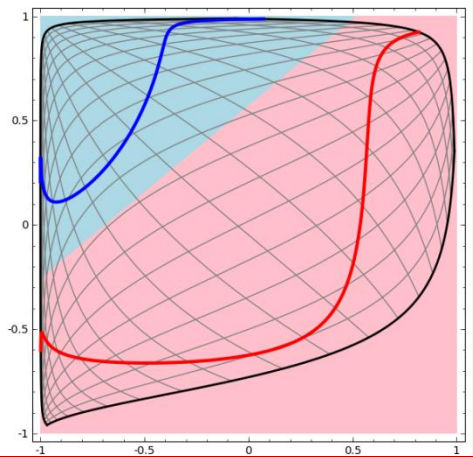
نقش لایه‌های مخفی



پرسپترون



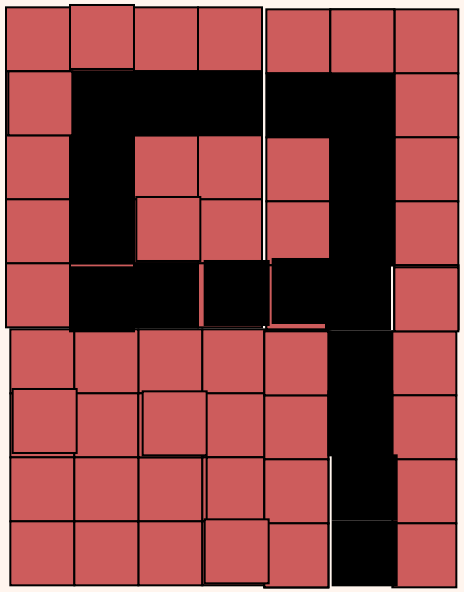
شبکه‌ی چندلایه

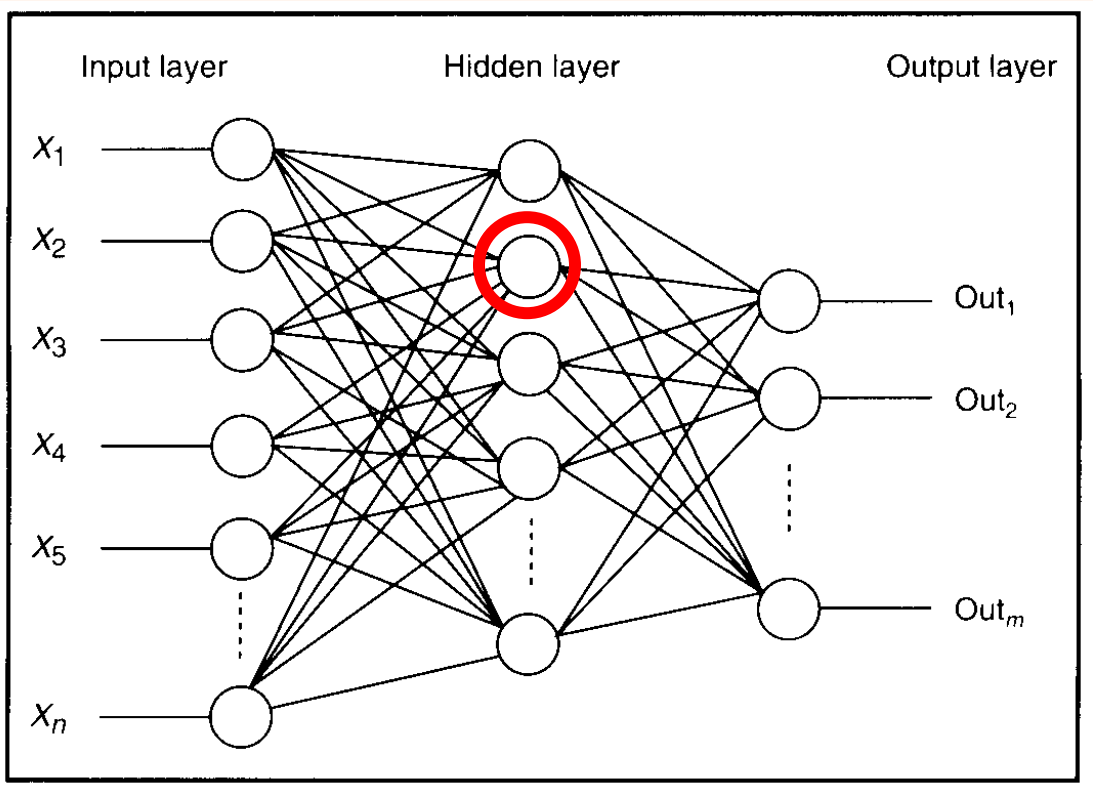


داده‌ی ورودی به لایه‌ی آخر

http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

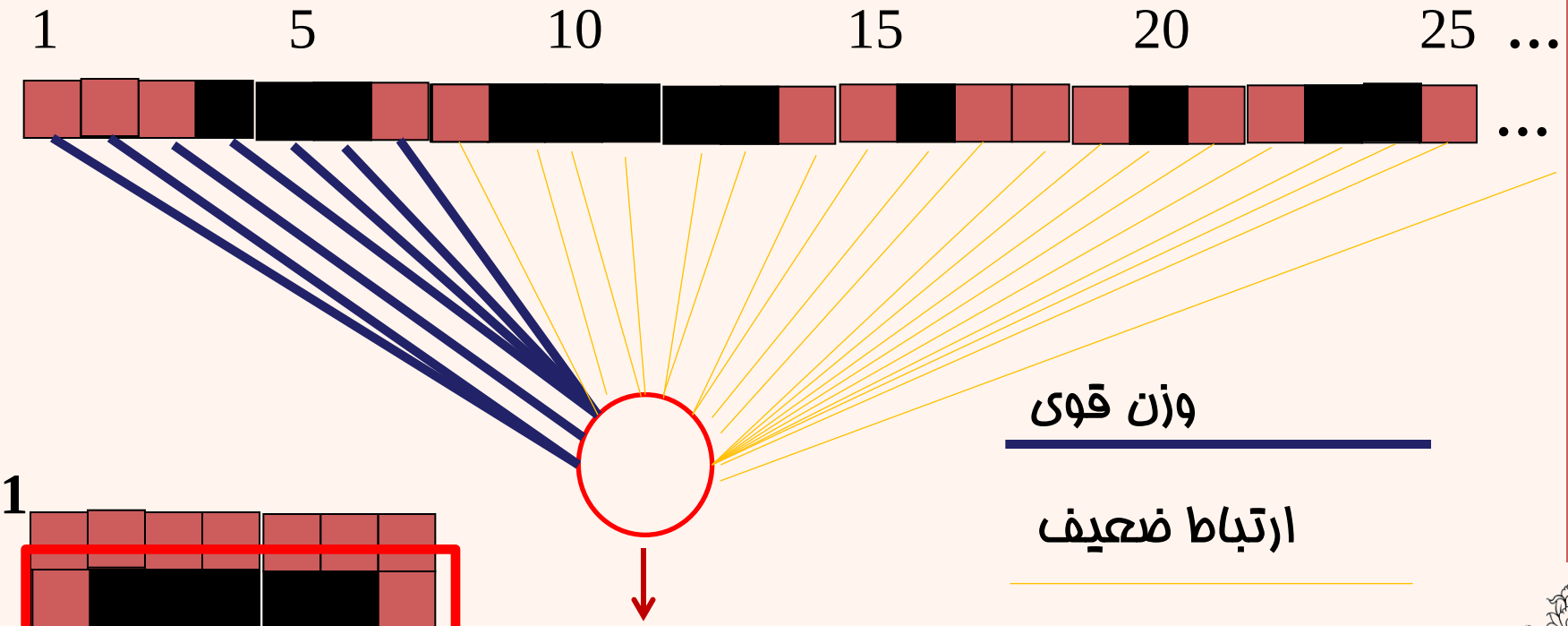
استخراج ویژگی

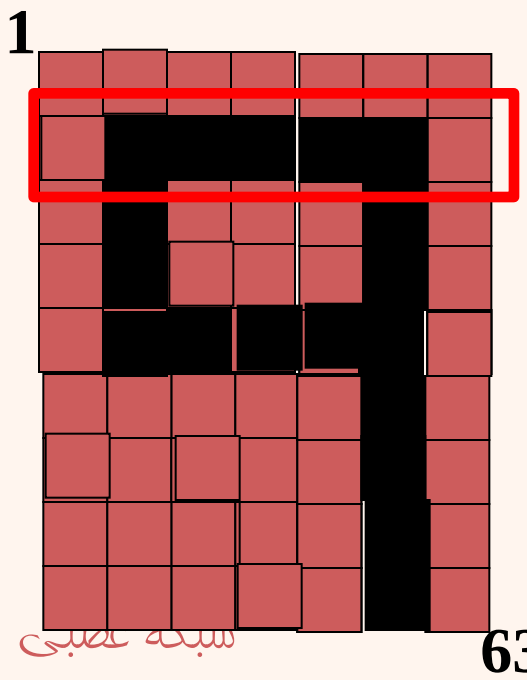




نقش نورون‌های لایه‌ی مخفی چیست؟

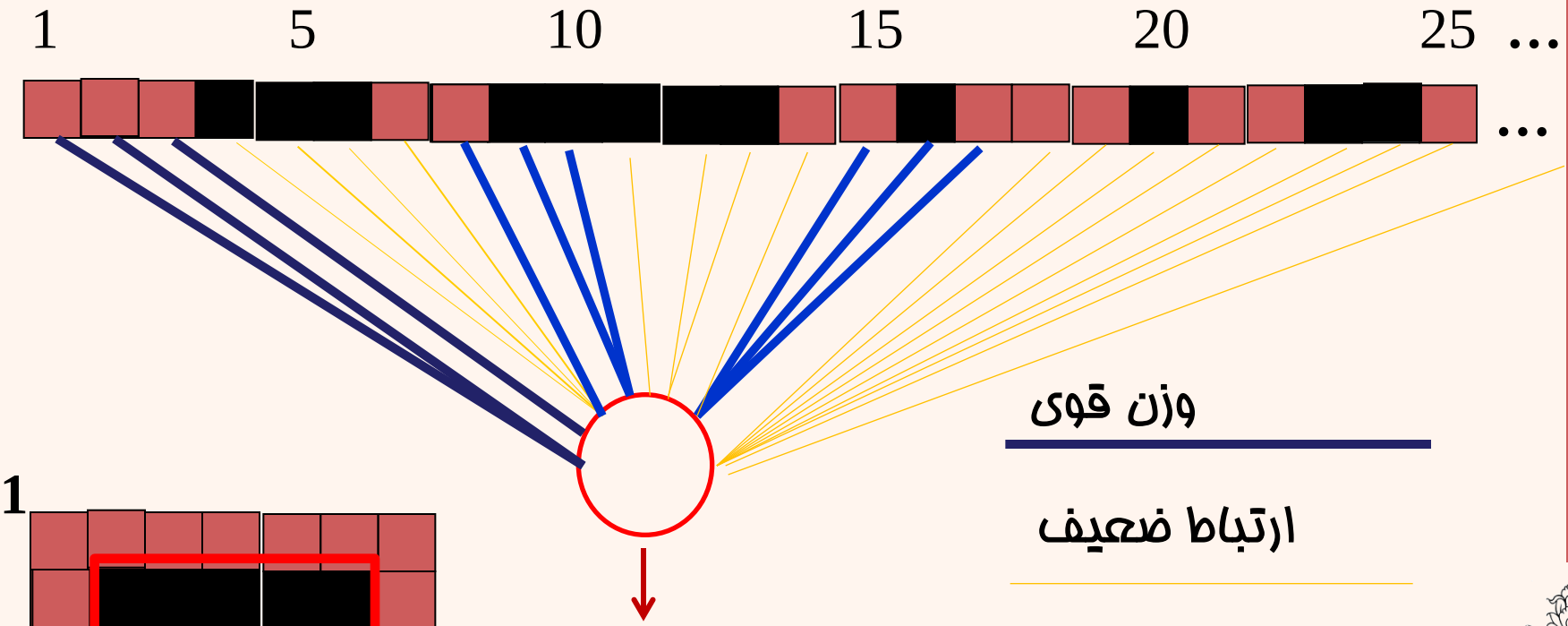
استخراج ویژگی(ادامه...)

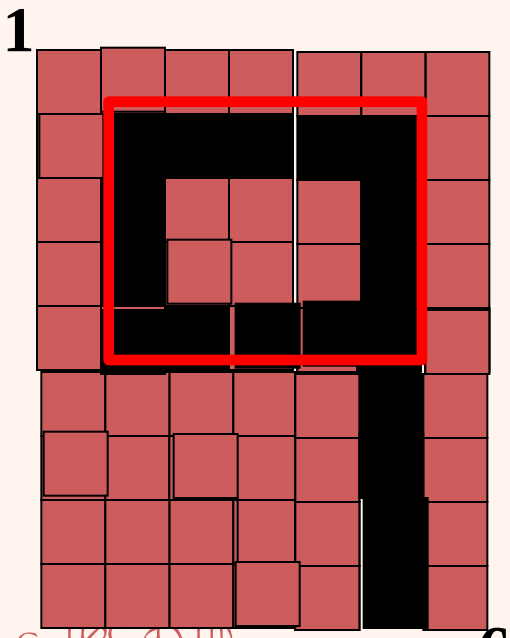




بیانگر یک لبه‌ی قوی افقی

استخراج ویژگی(ادامه...)





یک ناحیه مدور در بالای صفحه

شبکه عصبی

استخراج ویژگی(ادامه...)

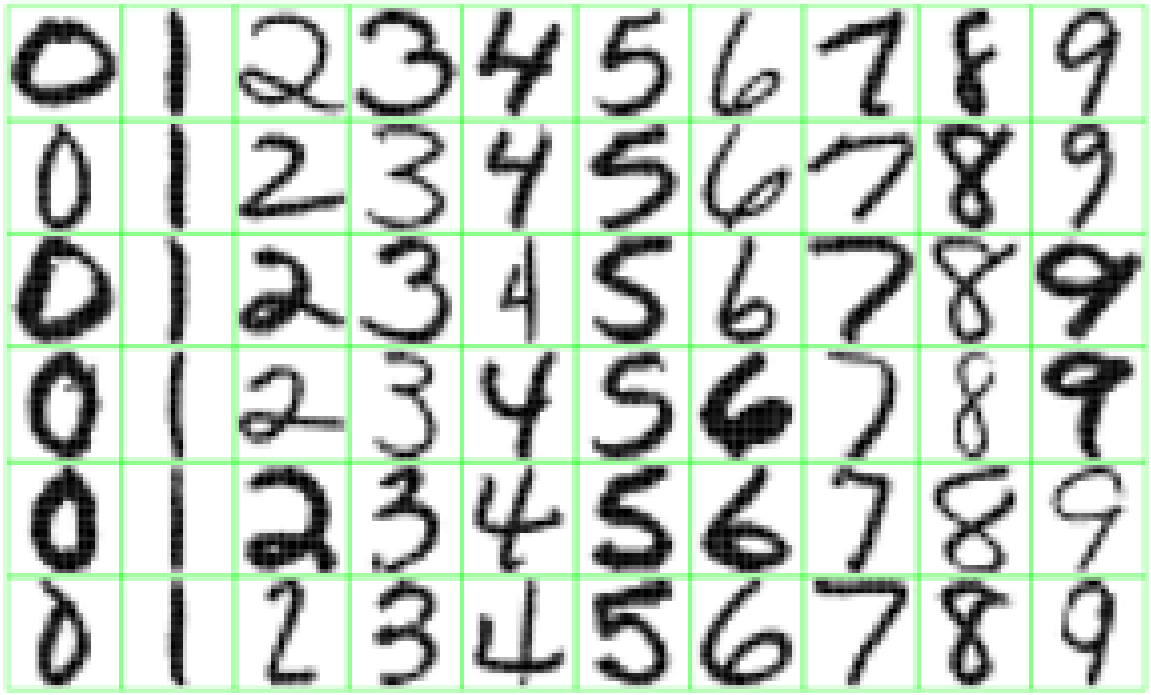


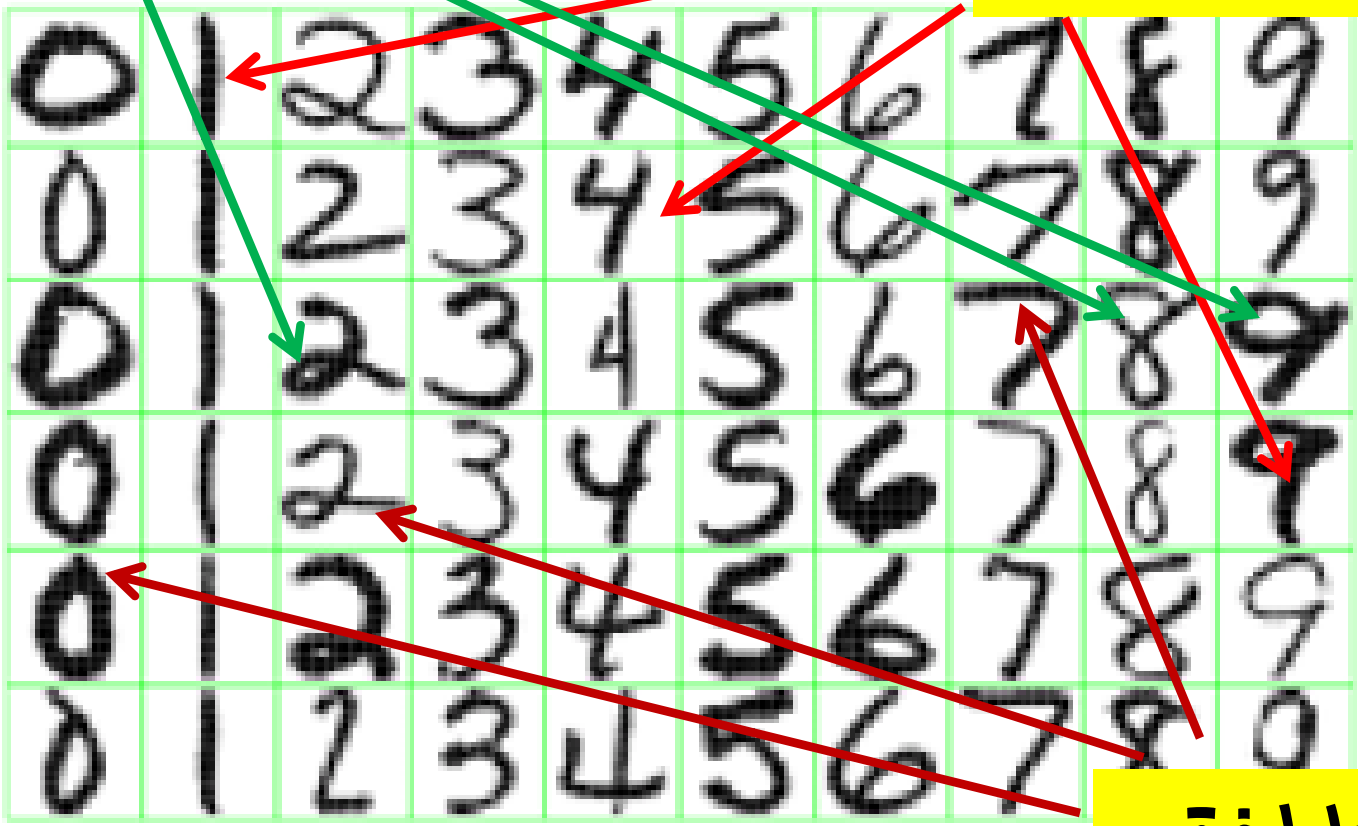
Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

شبکه‌ی عصبی چه نوع خصیصه‌ای برای تشخیص کاراکتر یاد می‌گیرد؟

استخراج ویژگی(ادامه...)

دایره‌های کوچک

خطوط عمودی



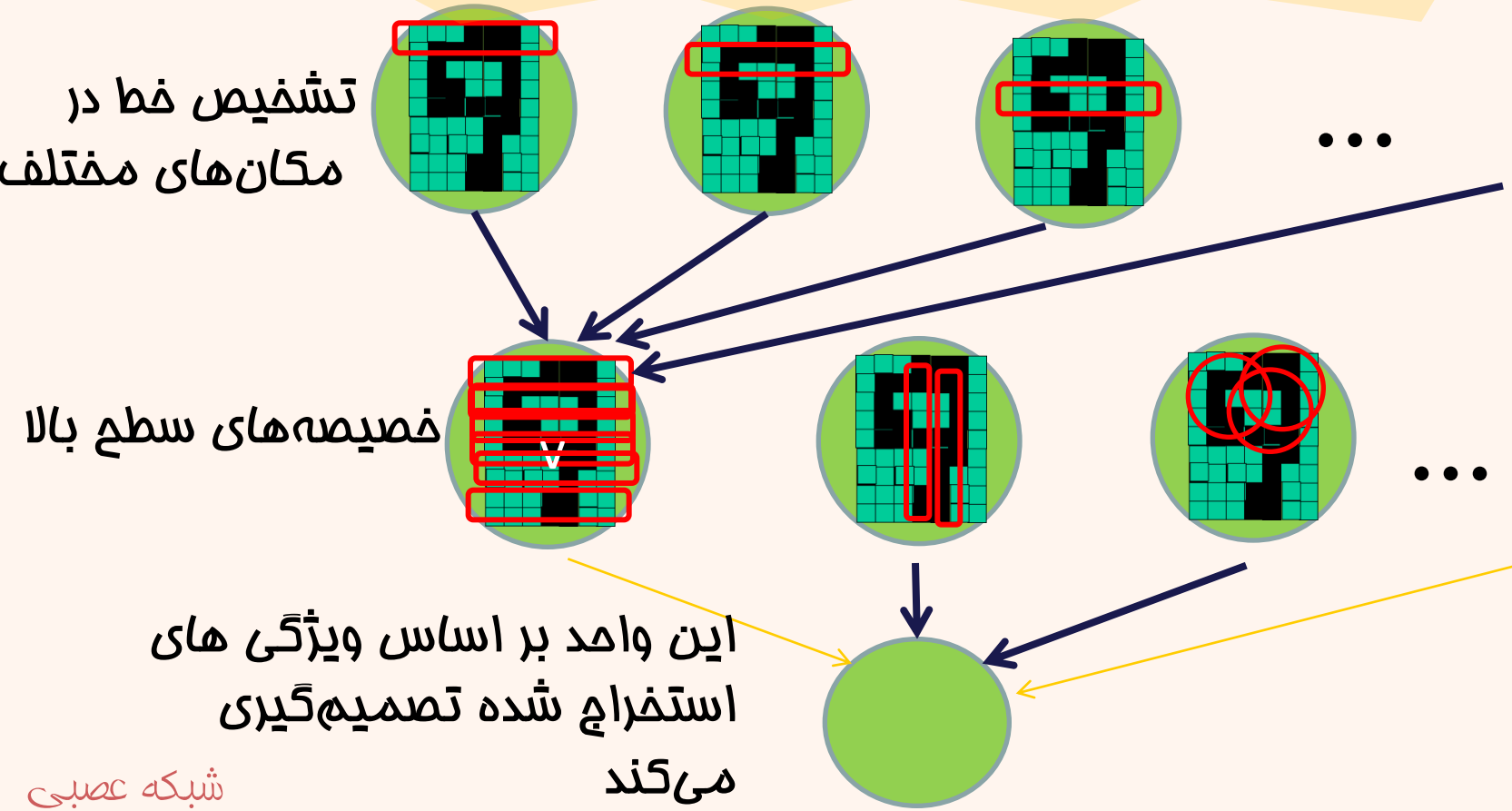
خطوط افقی

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

خصیصه‌های استخراج شده نسبت به مکان حساس هستند!

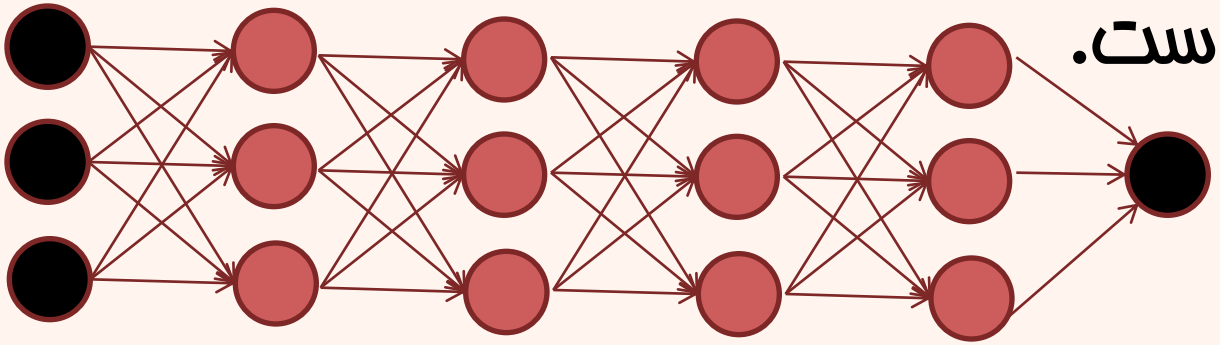
استخراج ویژگی(ادامه...)

- لایه‌های مخفی می‌توانند خصیصه‌های سطح بالا استخراج کنند.

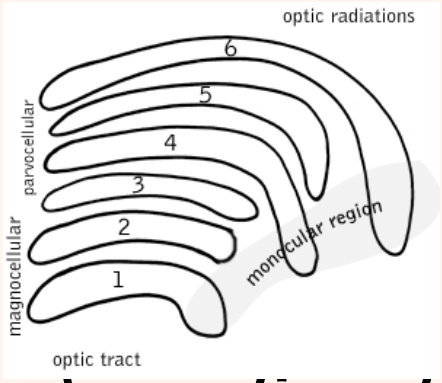


استخراج ویژگی(ادامه...)

- در این حالت استفاده از چندین لایه‌ی مختلف توجیه‌پذیر است.

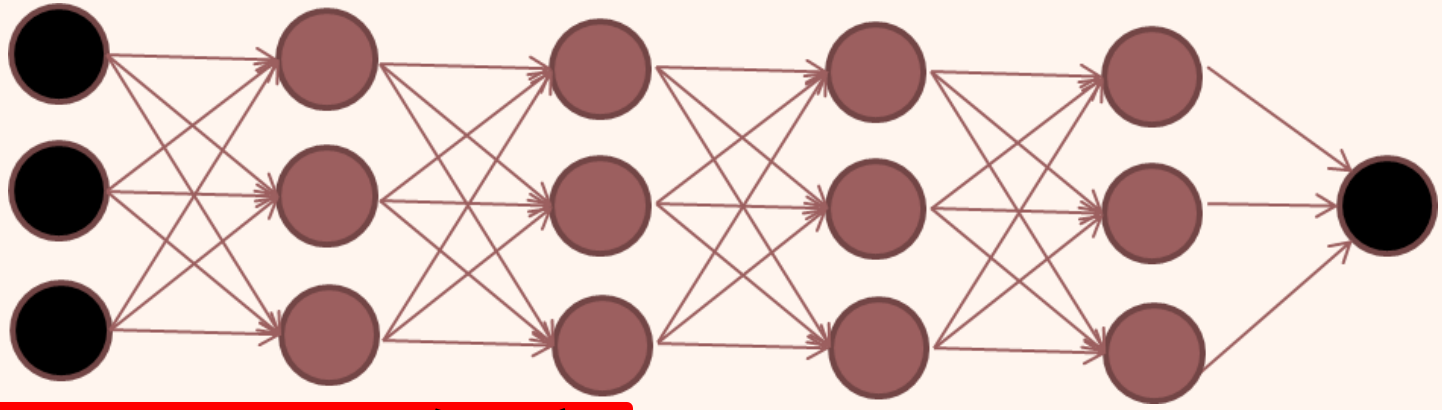


- درست مانند مغز

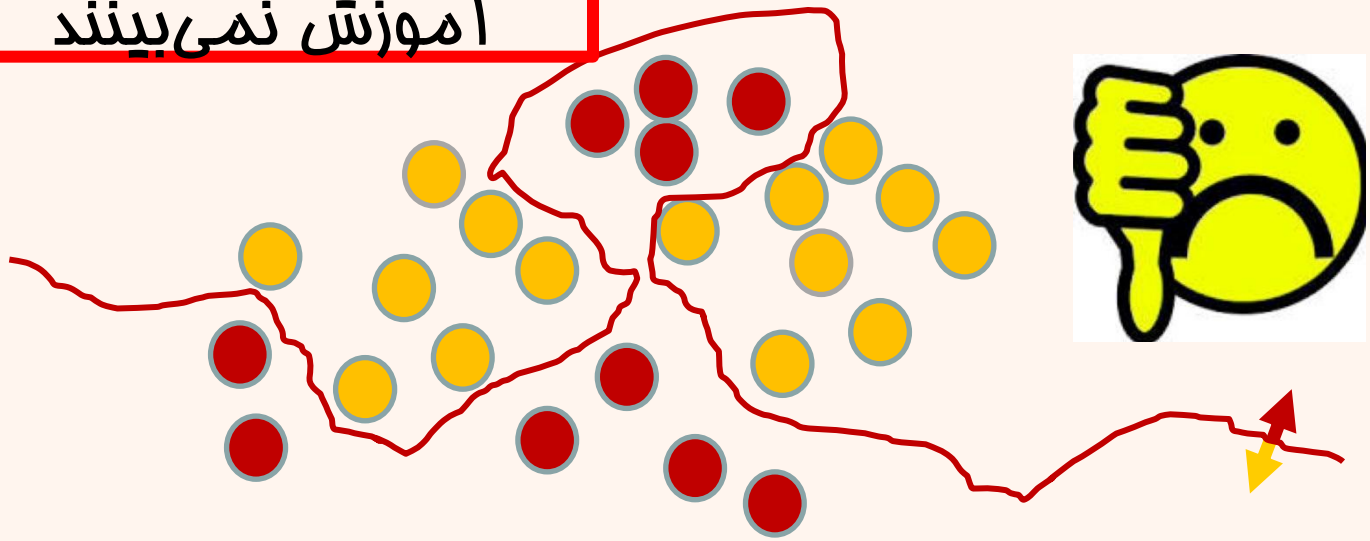


- لایه‌های مخفی باید بتوانند خصیصه‌های مناسب را استخراج کنند.

استخراج ویژگی(ادامه...)

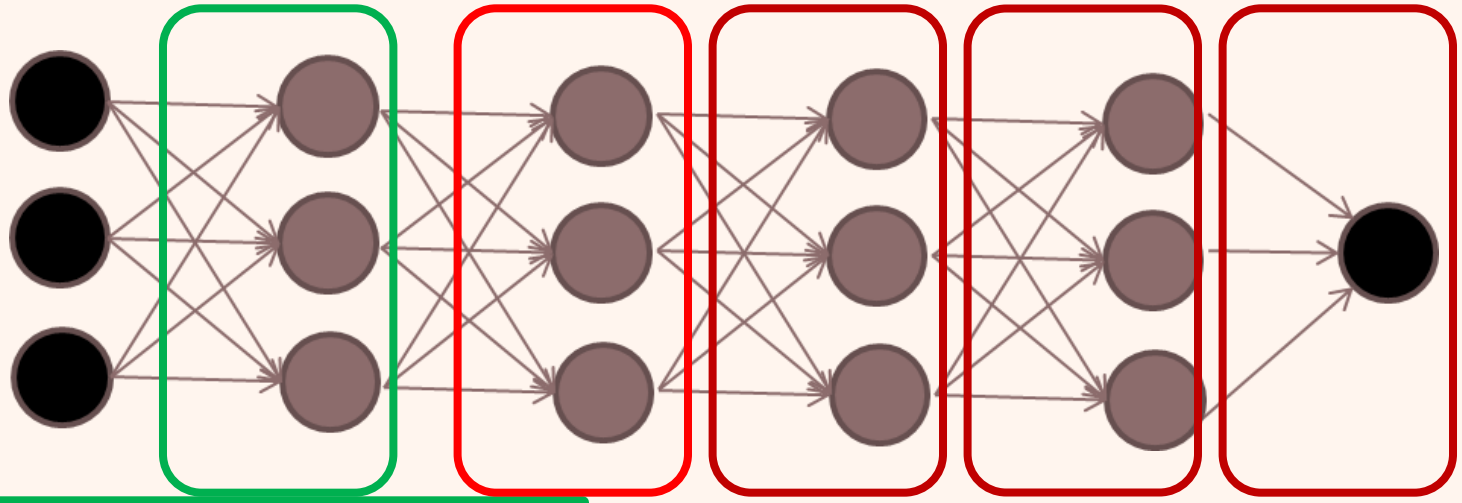


لایه‌های نخست خوب آموزش نمی‌بینند



مشکل بیش‌برازش

شیوه‌ای نو برای آموزش



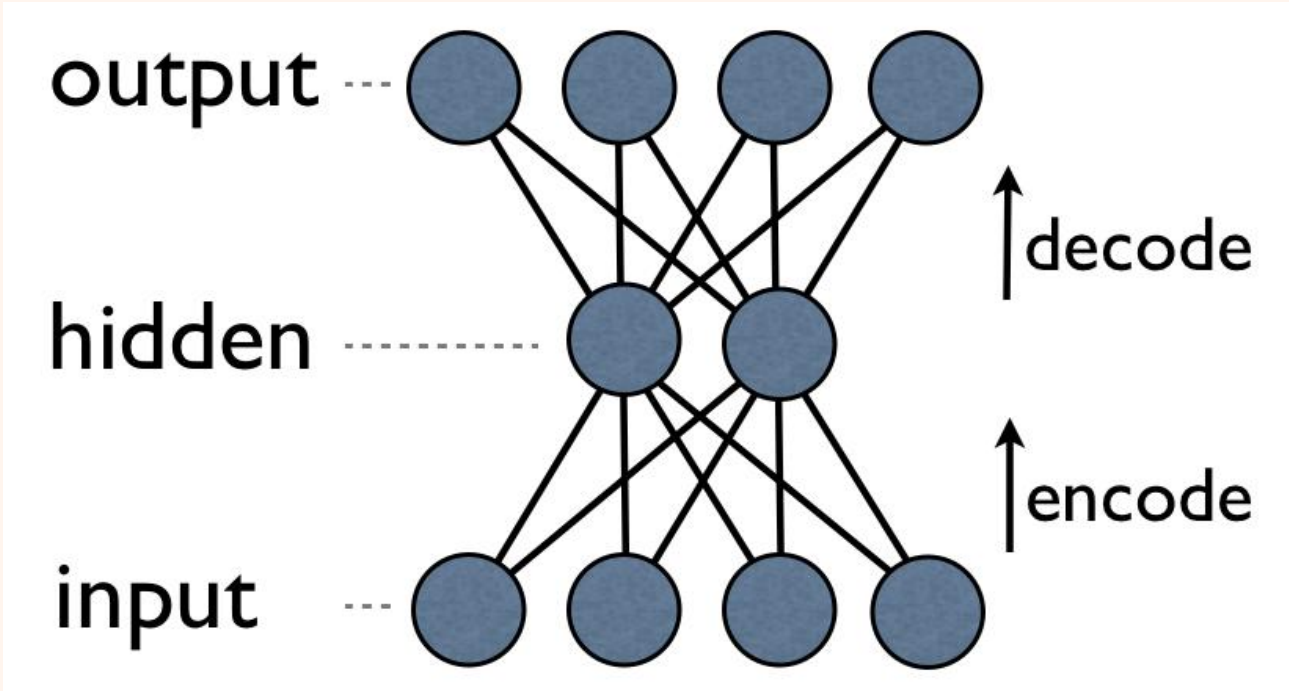
ابتدا این لایه آموزش می‌بیند

سپس این لایه و به همین‌ترتیب

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

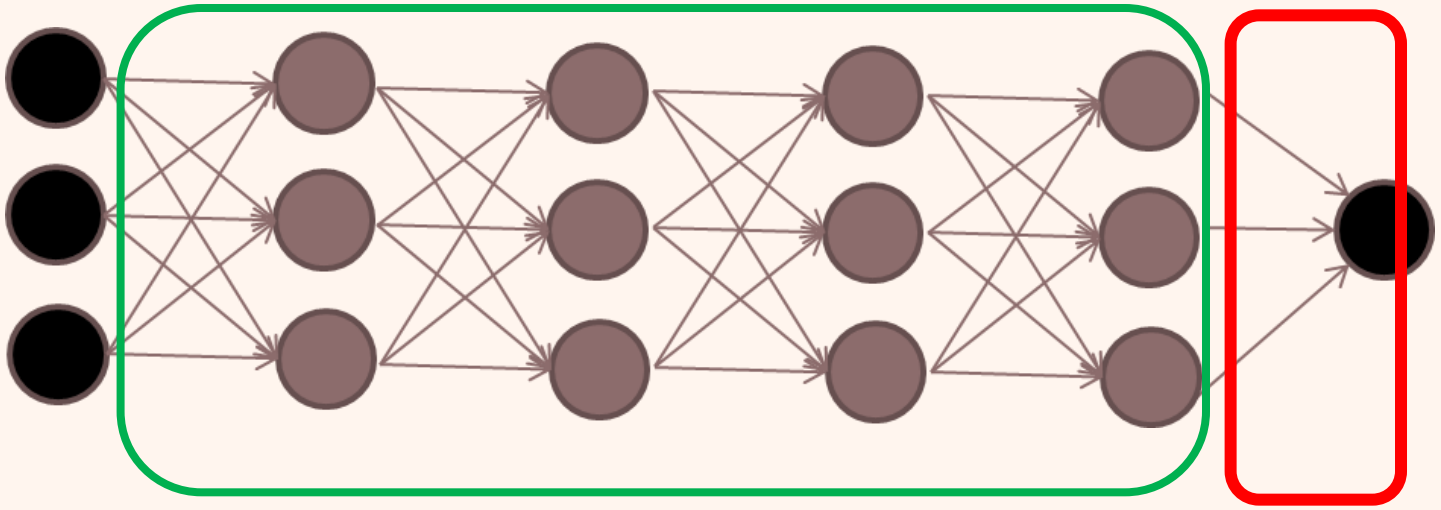
auto-encoder

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



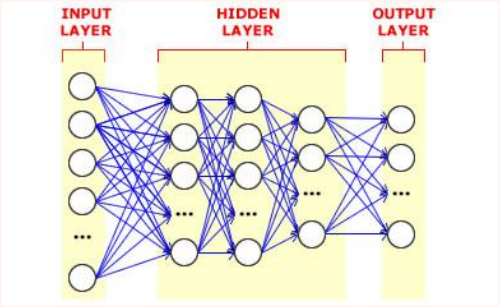
در این حالت تعداد نورون‌های لایه‌ی مخفی کم‌تر در نظر گرفته می‌شود، بدین‌ترتیب لایه‌ی مخفی برای استخراج خصیصه آموزش می‌بیند.

شیوه‌ای نو برای آموزش



لایه‌های میانی به عنوان auto-encoder آموزش می‌بینند.

لایه‌ی آخر برای دسته‌بندی مورد استفاده قرار می‌گیرد.



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

یادگیری خودآموخته

Self-taught Learning

یادگیری خودآموخته

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

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!

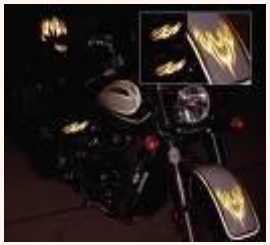
آموزش در مغز

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

(Geoffrey Hinton, personal communication)

یادگیری بانظارت

مجموعه‌ی آموزشی



Cars

Motorcycles

داده‌های آزمایش

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

آموزش به شیوه‌ی مغز

مجموعه‌ی آموزشی

Cars Motorcycles

تصاویر بدون برچسب

داده‌های آزمایش

Self-taught Learning

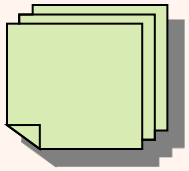
آموزش به شیوه‌ی مغز

Labeled Digits

+

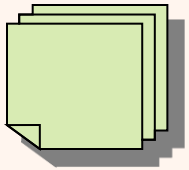
Unlabeled English characters

?



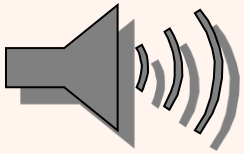
Labeled Webpages

+



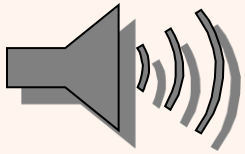
Unlabeled newspaper articles

?



Labeled Russian Speech

+



Unlabeled English speech

?

تاریخچه‌ی یادگیری ماشین

- یادگیری بانظارت(حدود بیست‌سال پیش)

Supervised learning

Cars

Motorcycles

- یادگیری نیمه‌نظارتی(حدود ده سال پیش)

Semi-supervised learning

Cars

Motorcycles

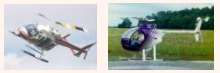
- انتقال یادگیری (حدود ده سال پیش)

Transfer learning

Bus

Tractor

Aircraft

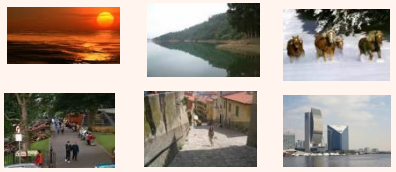

Helicopter

Cars

Motorcycles

- هم‌اکنون: یادگیری خودآموخته

Self-taught learning


Natural scenes

Car

Motorcycle

شبکه عصبی

یادگیری خودآموخته

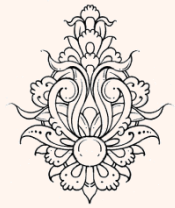
- مجموعه‌ی داده‌های آموزشی(برچسب‌خورده)

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

- مجموعه‌ی داده‌های بدون برچسب

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

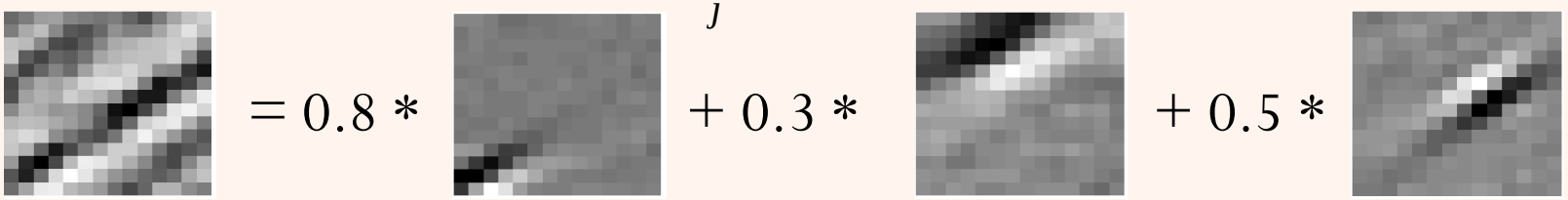
- لزومی ندارد که داده‌های از یک نوع باشند.



یادگیری خودآموخته

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

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



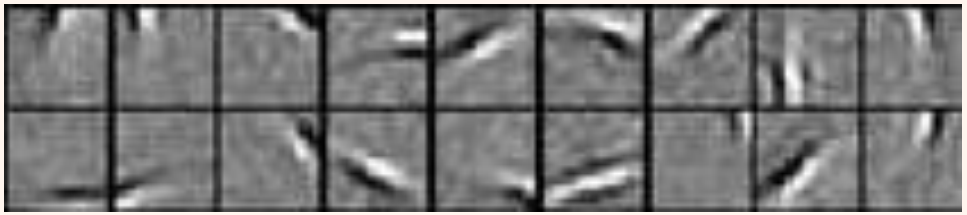
$$x = 0.8 * b_{87} + 0.3 * b_{376} + 0.5 * b_{411}$$

- از داده‌های بدون برچسب برای آموختن بردارهای پایه استفاده می‌شود. در واقع، داده‌های بدون برچسب برای ارائه‌ی یک خصیصه‌ی سطح بالا استفاده می‌شود.
- پس از آن از داده‌های برچسب‌دار برای دسته‌بندی بانظارت استفاده می‌شود.

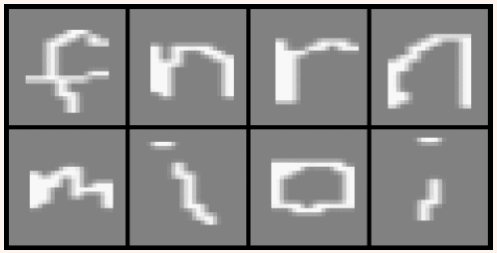
مثالی از بردارهای پایه

Natural images.

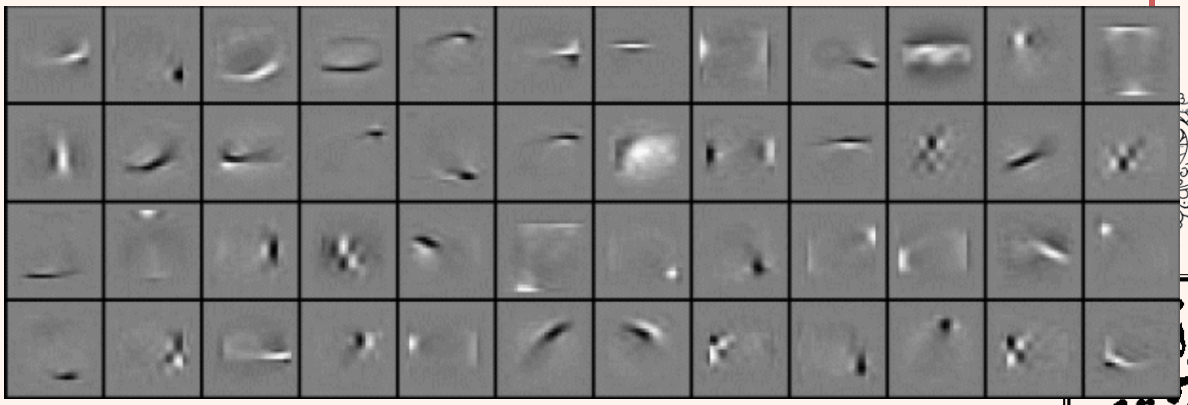
Learnt bases: “Edges”



Handwritten characters.



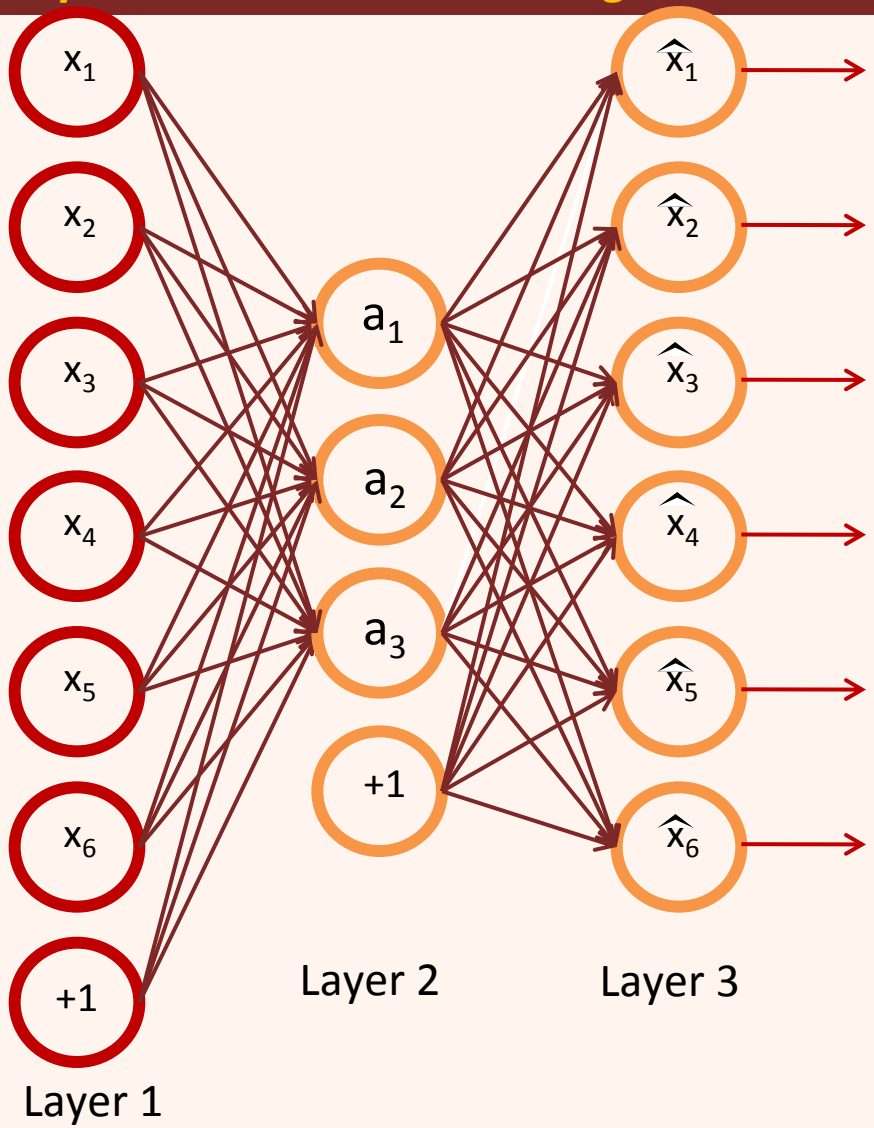
Learnt bases: “Strokes”



auto-encoder

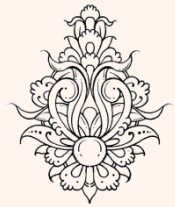
auto-encoder

Unsupervised feature learning with a neural network



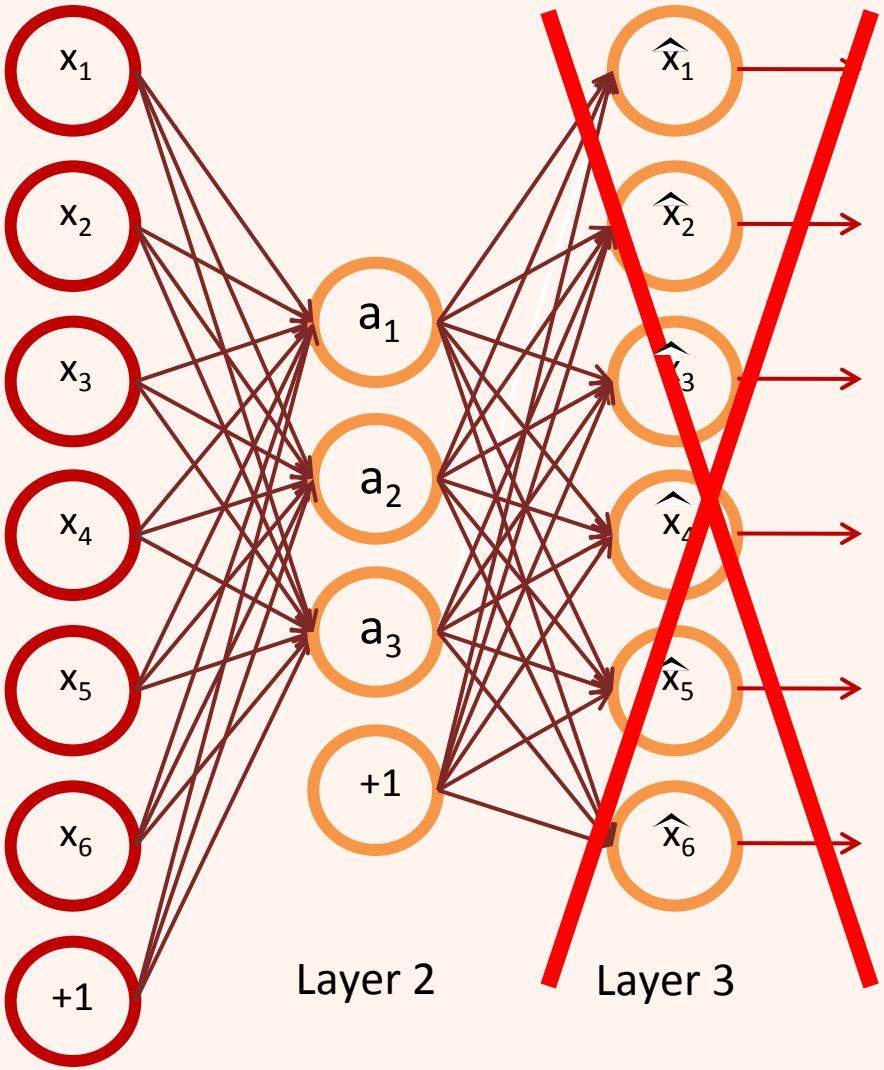
شبکه به
گونه‌ای آموزش می‌بیند که

$$\hat{x} \approx x$$

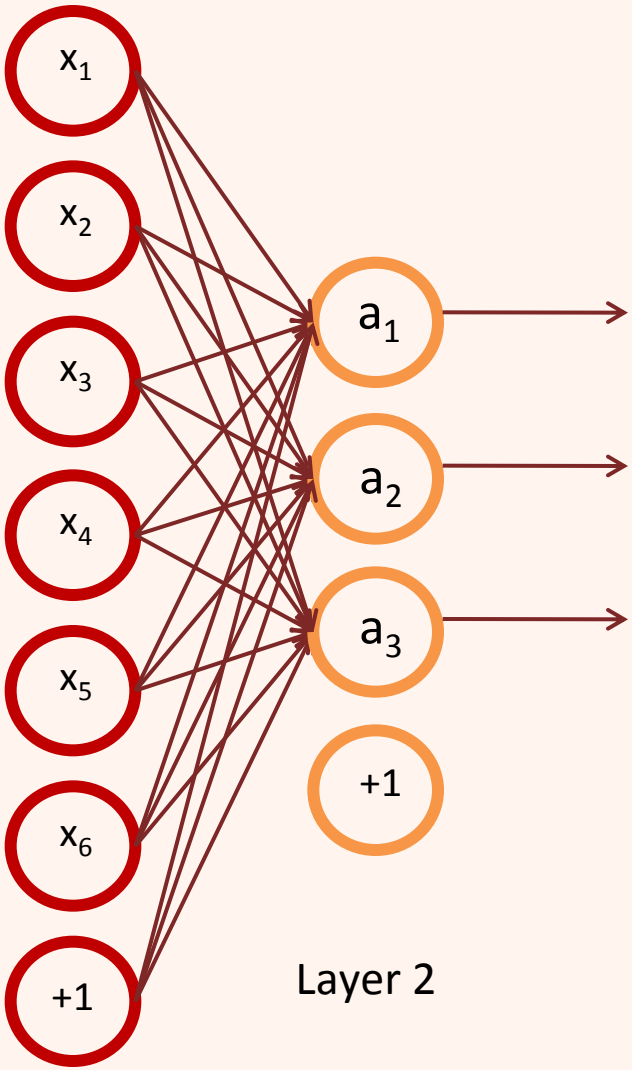


http://ufldl.stanford.edu/eccv10-tutorial/

auto-encoder



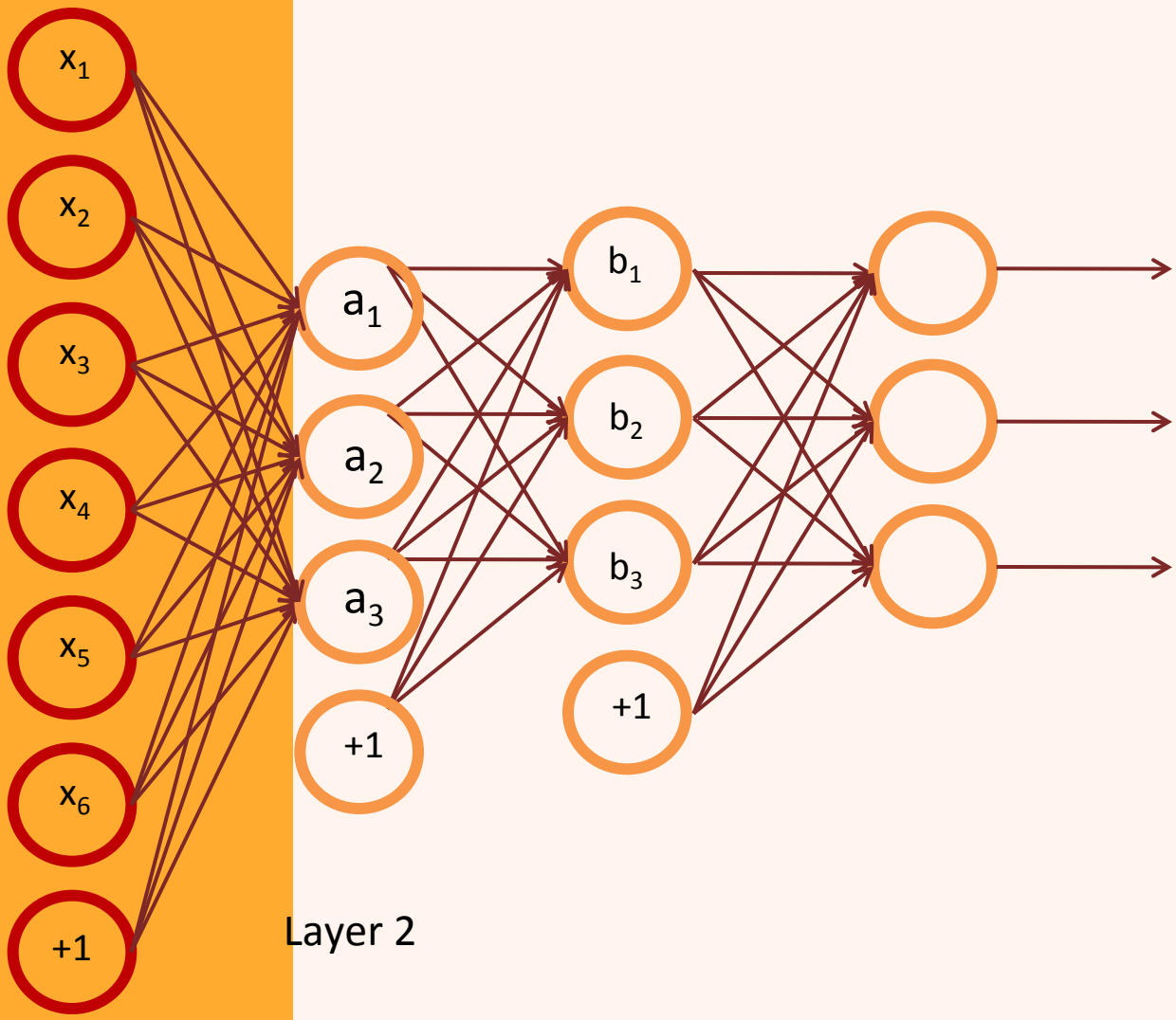
auto-encoder



$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

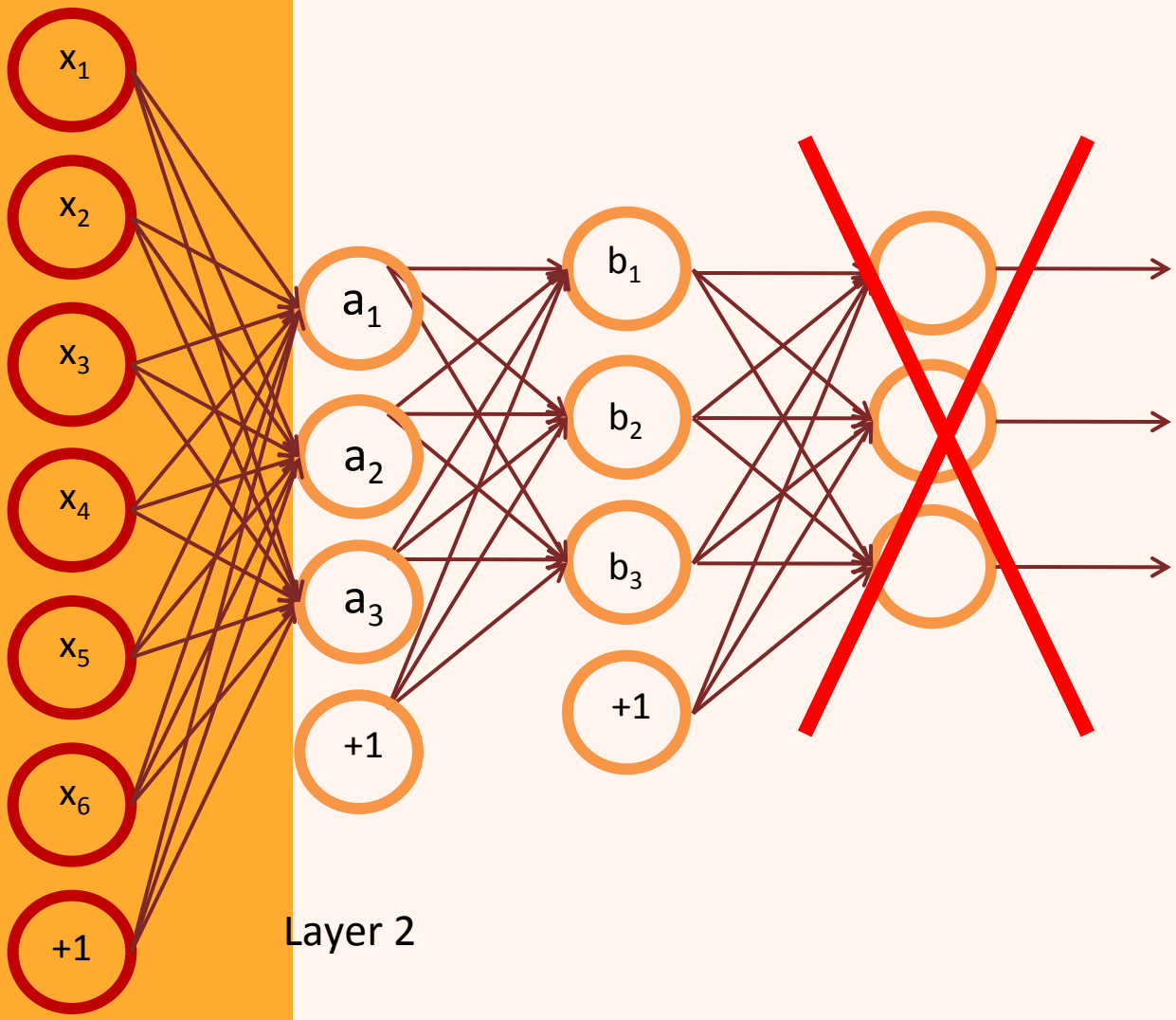
نمایش جدید برای داده‌ها

auto-encoder



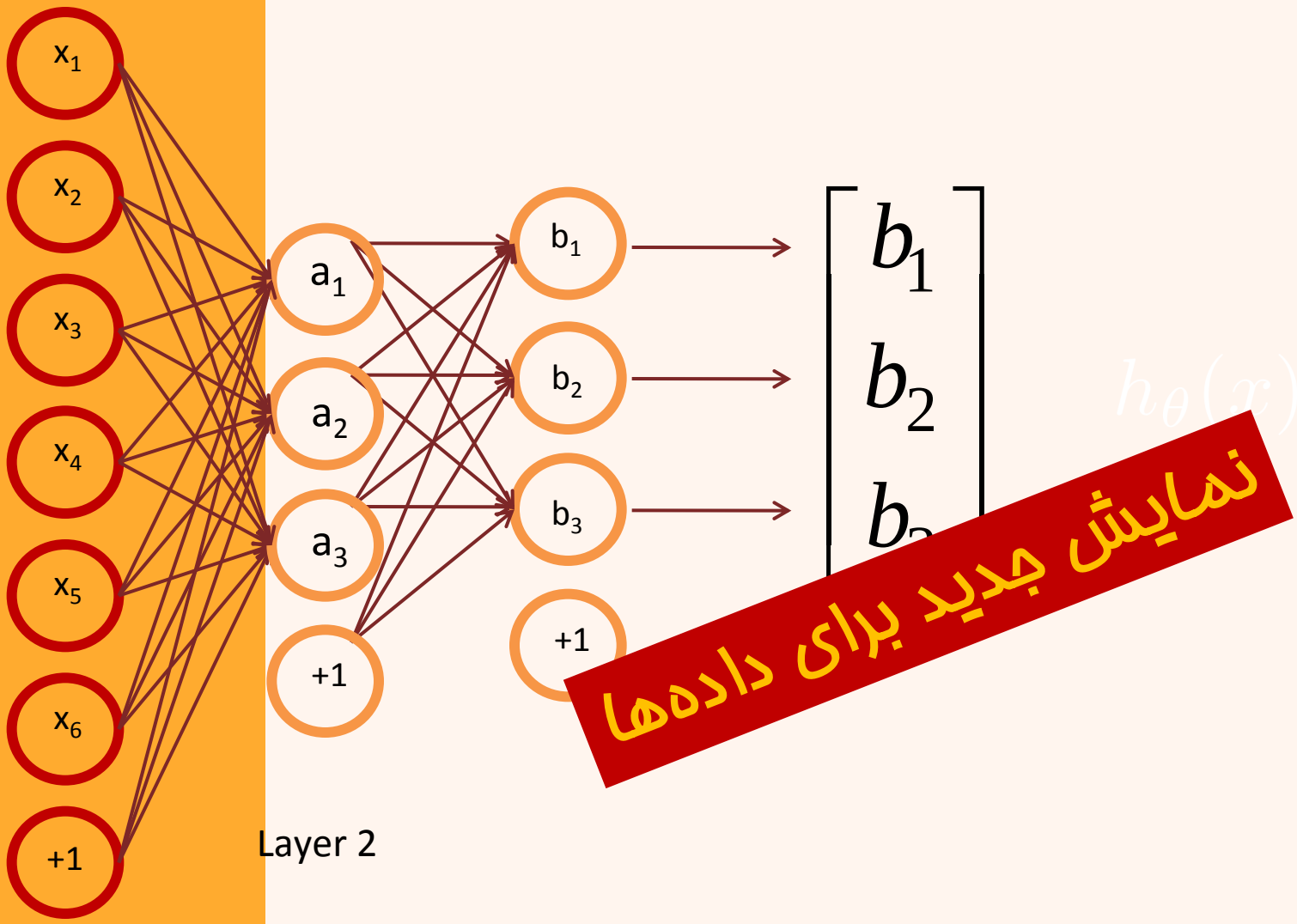
$h_\theta(x)$

auto-encoder

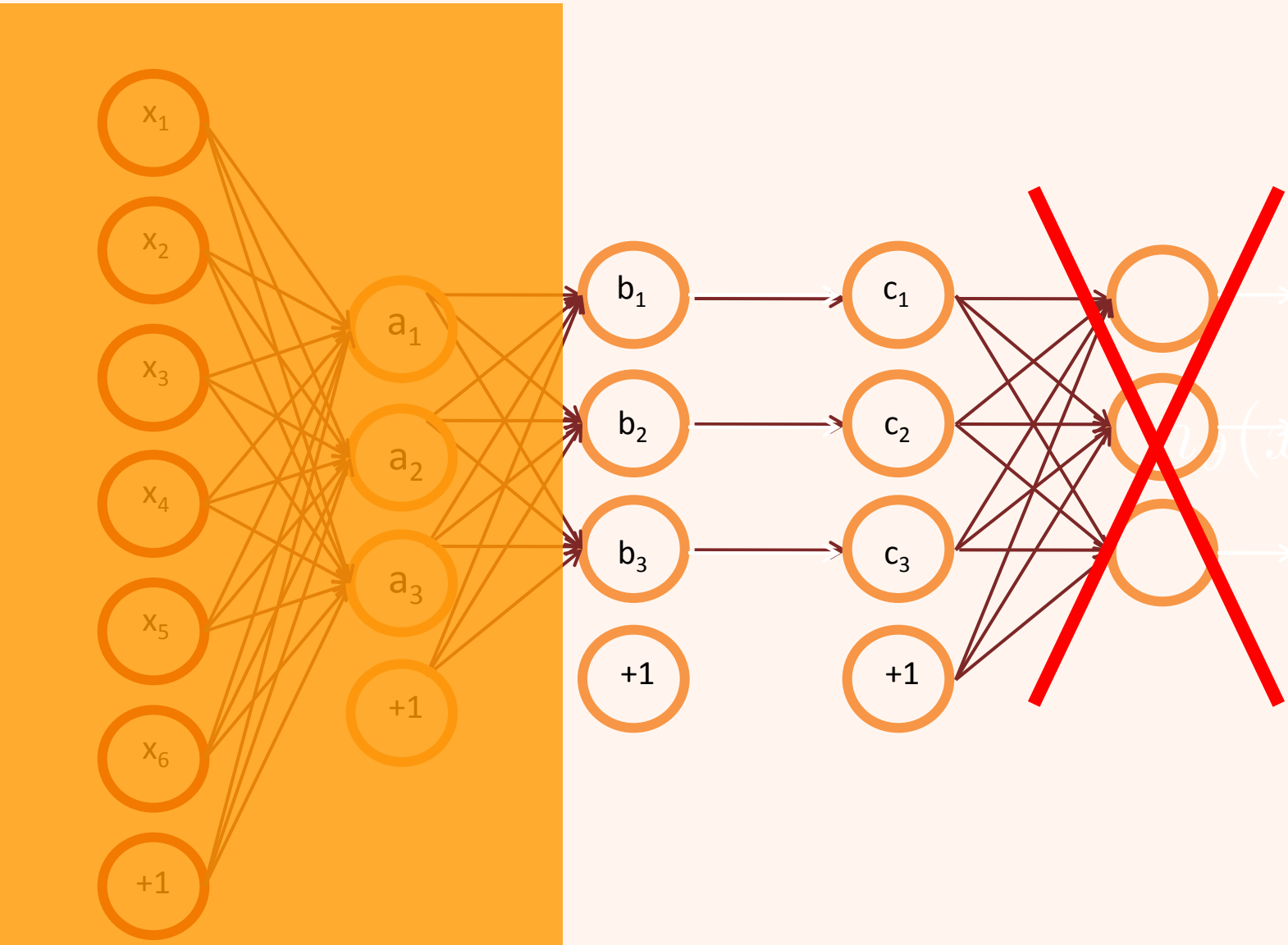


$h_\theta(x)$

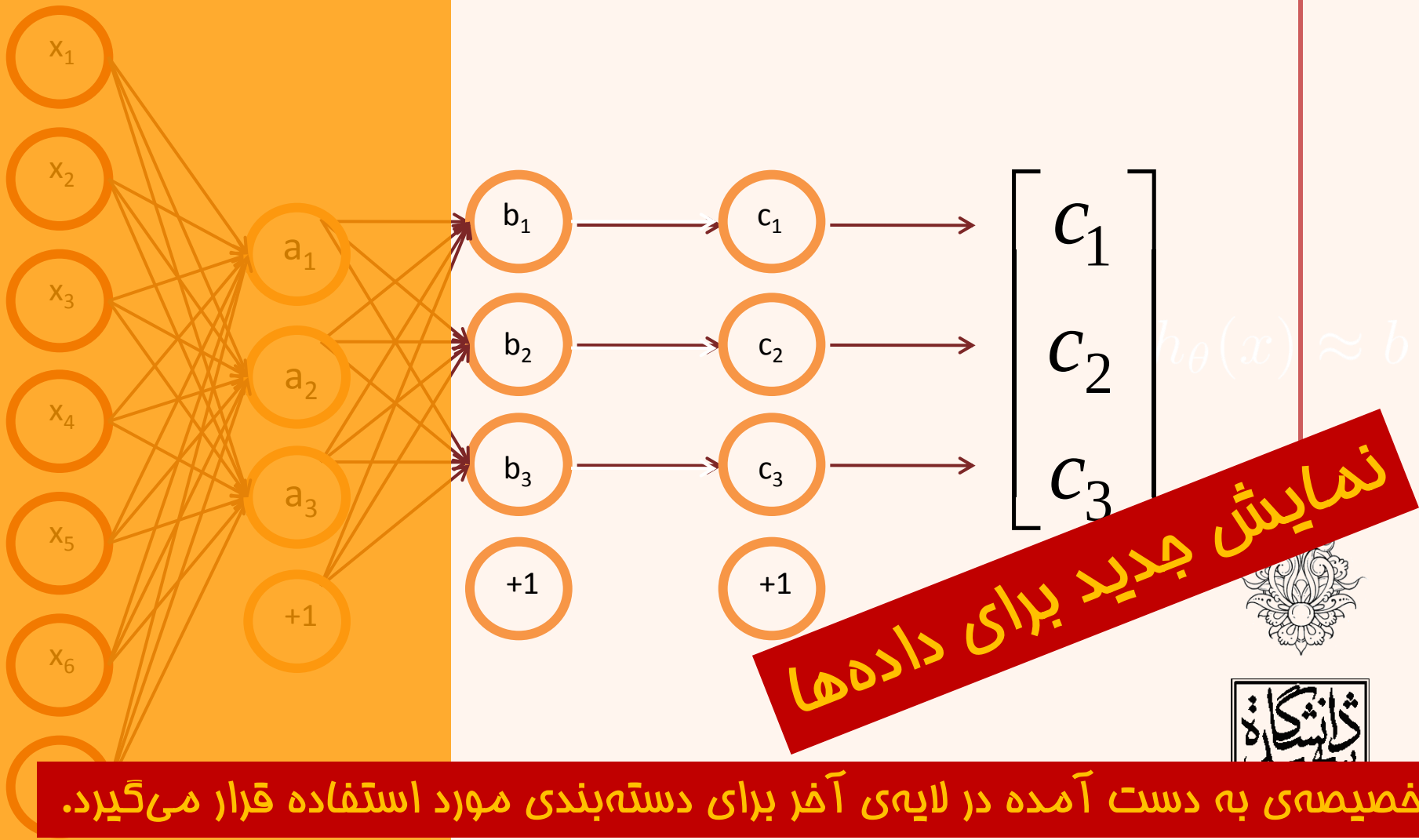
auto-encoder



auto-encoder



auto-encoder

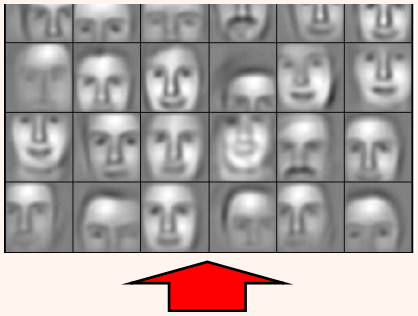


نمایش جدید برای داده‌ها

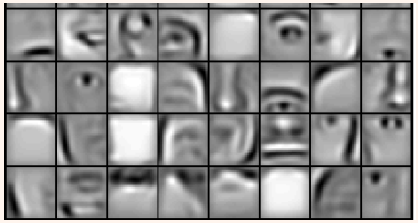
خصیصه‌ی به دست آمده در لایه‌ی آخر برای دسته‌بندی مورد استفاده قرار می‌گیرد.

آموزش سلسله‌مراتبی خصیصه‌ها

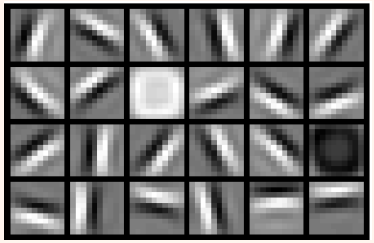
object models



object parts (combination of edges)

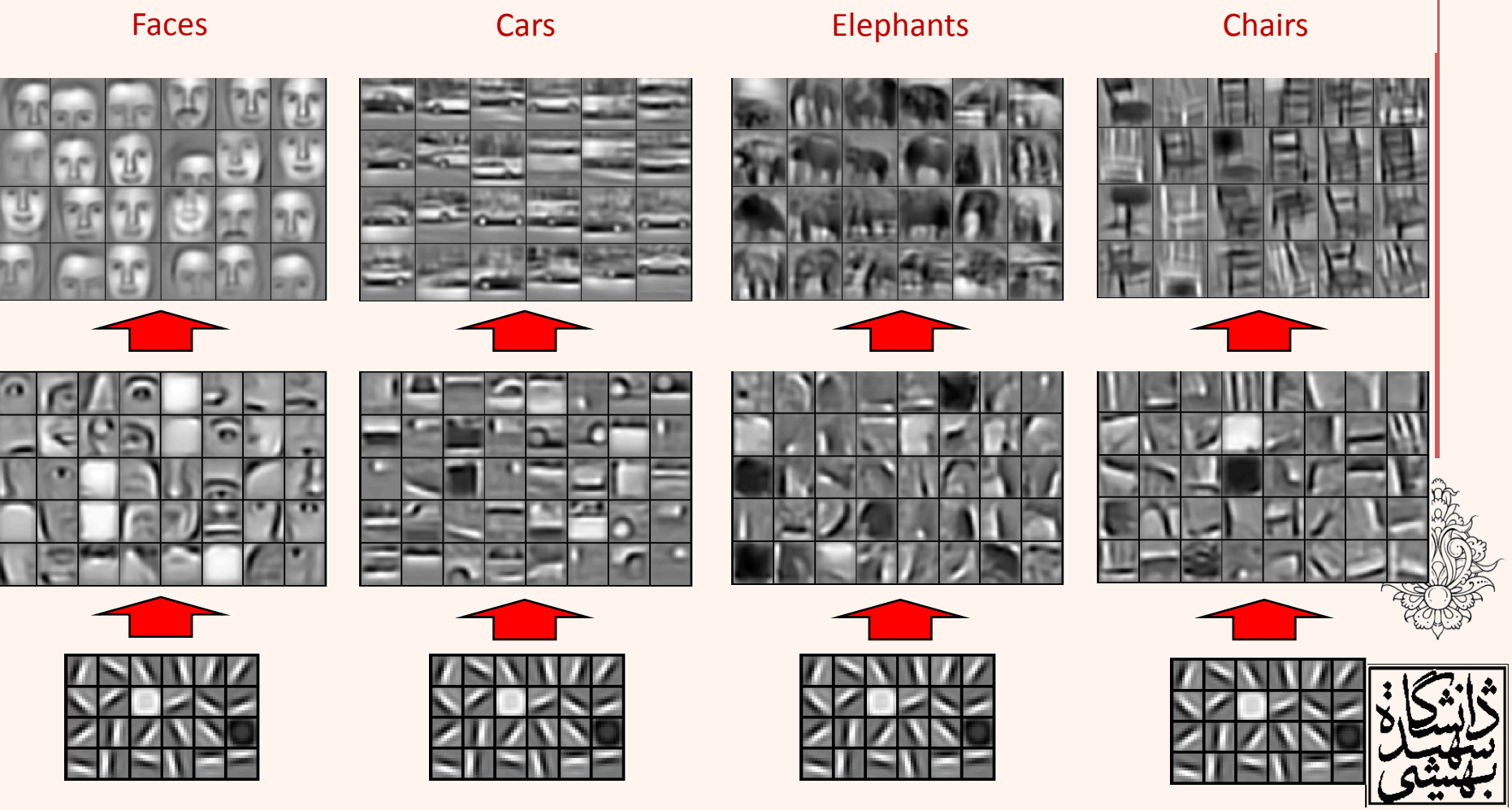


edges

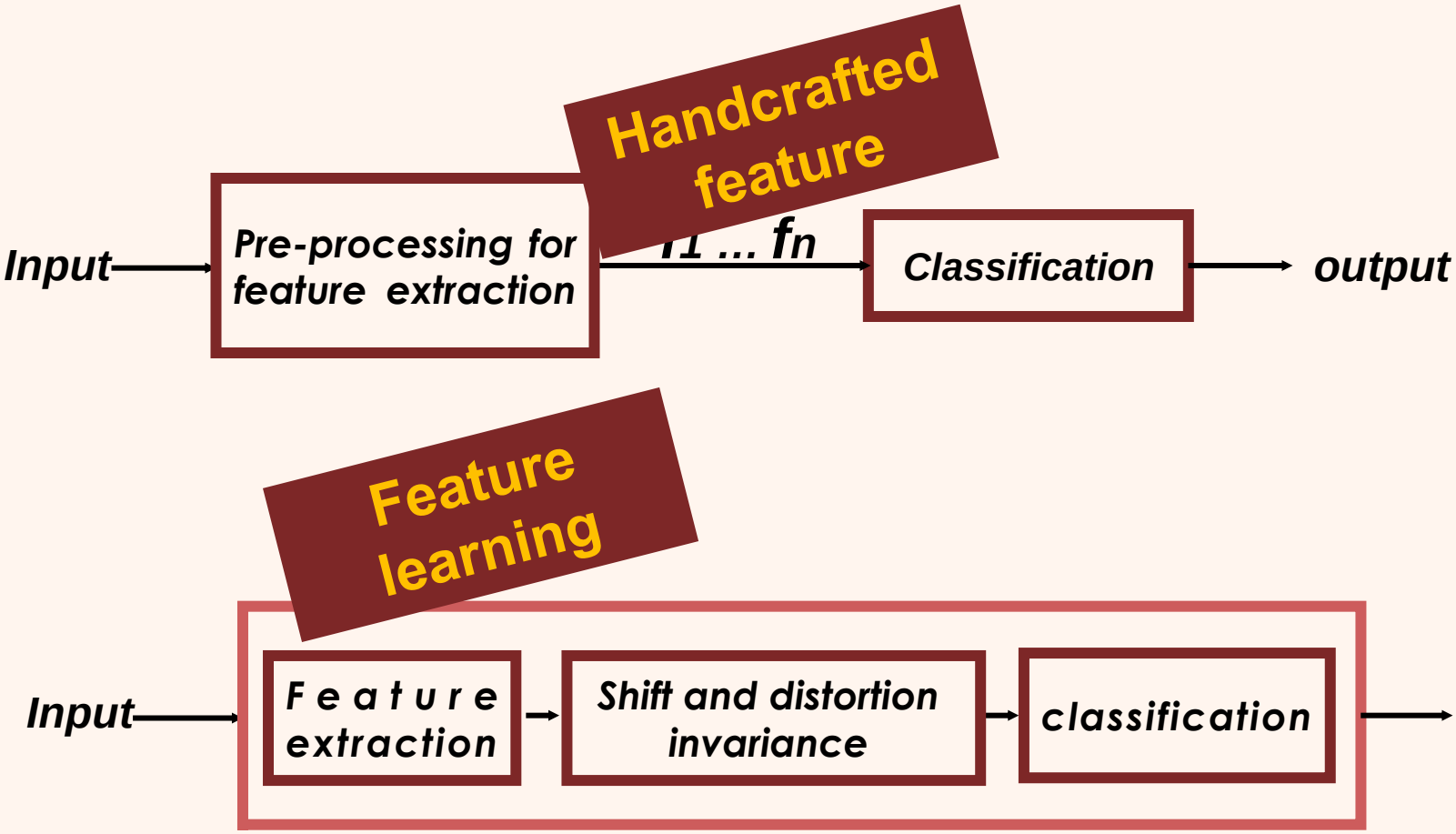


pixels

آموزش سلسله‌مراتبی خصیصه‌ها



استخراج خصیصه



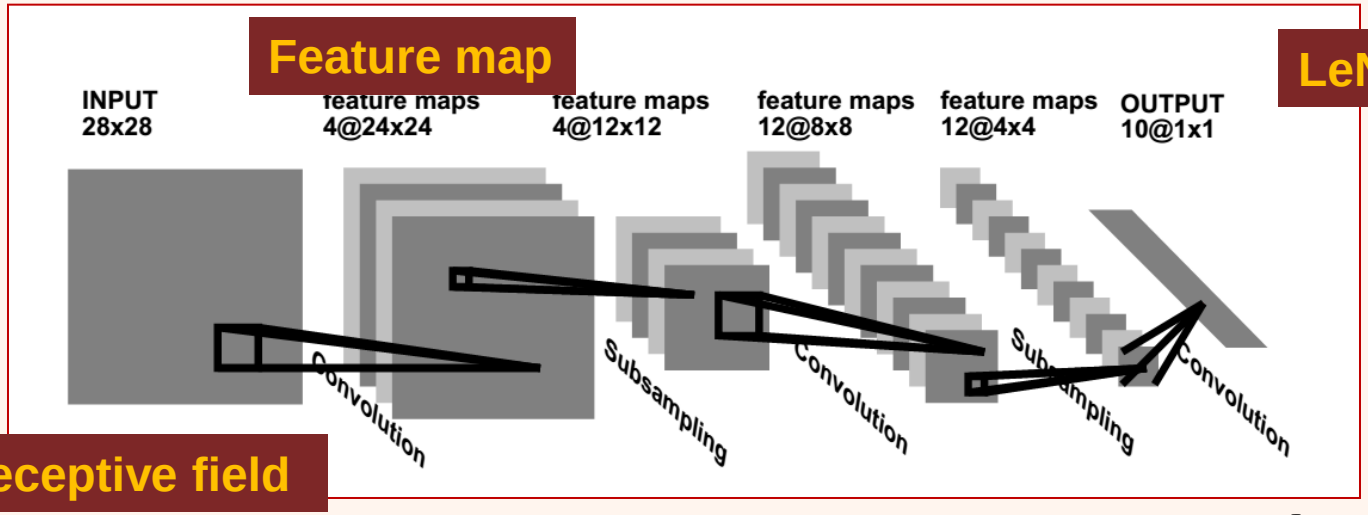
CNN

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

Convolutional Neural Networks

Feature map

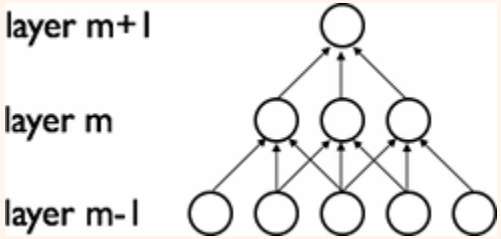
LeNet1(3000/100000)



receptive field

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

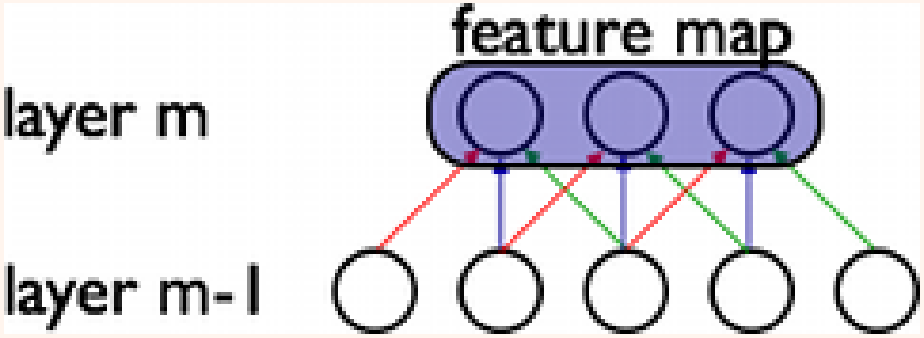
Sparse Connectivity



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

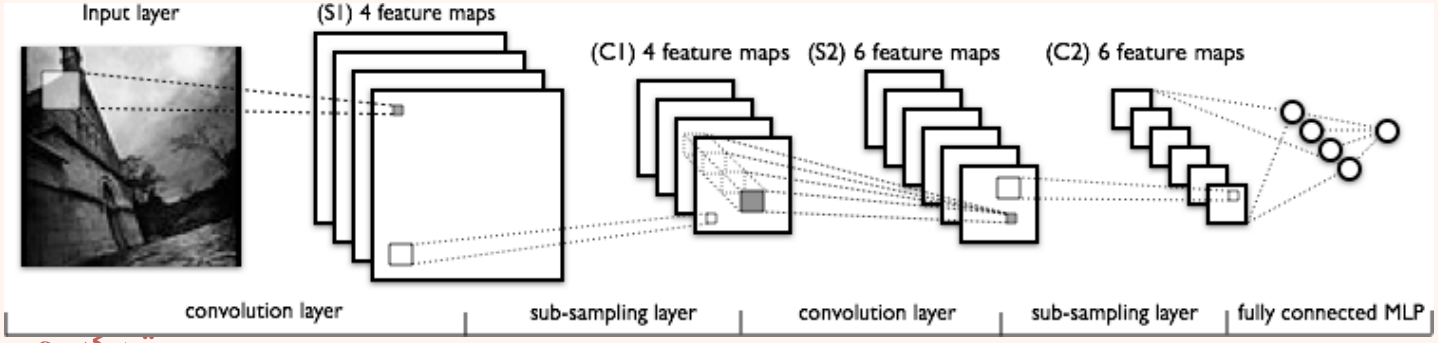
– از سوی دیگر این کار باعث می‌شود پارامترهای آزاد شبکه به شدت کاهش یابد.

Shared Weights



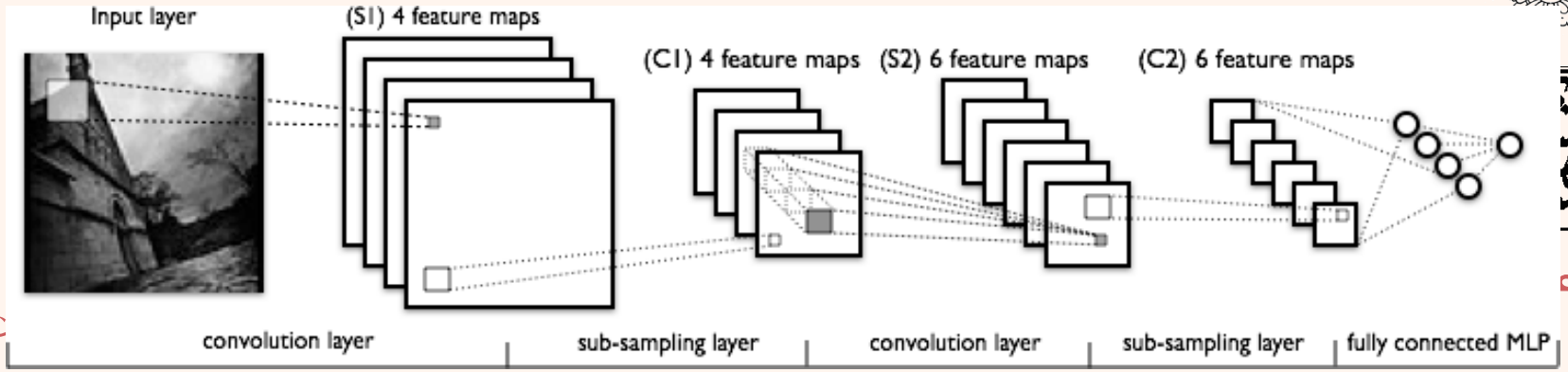
Convolutional Neural Networks

- در این حالت خروجی 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.



Sub-Sampling (Pooling)

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



LeNet-5(60000/340000)

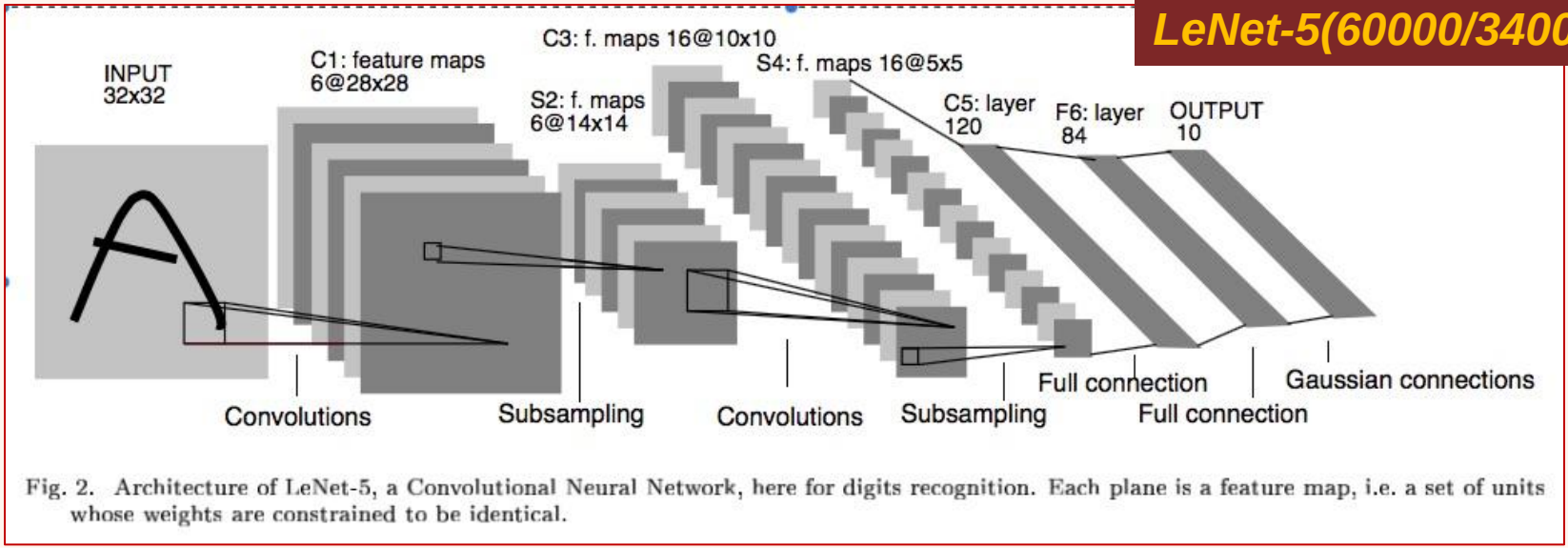


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

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

آموزش(پس‌انتشار خطا)

در این حالت قانون پس‌انتشار خطا به صورت زیر اصلاح می‌شود:

$$To\ constrain: \quad w_1 = w_2$$
$$we\ need: \quad \Delta w_1 = \Delta w_2$$

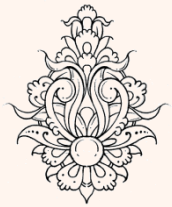
$$compute: \quad \frac{\partial E}{\partial w_1} \quad and \quad \frac{\partial E}{\partial w_2}$$

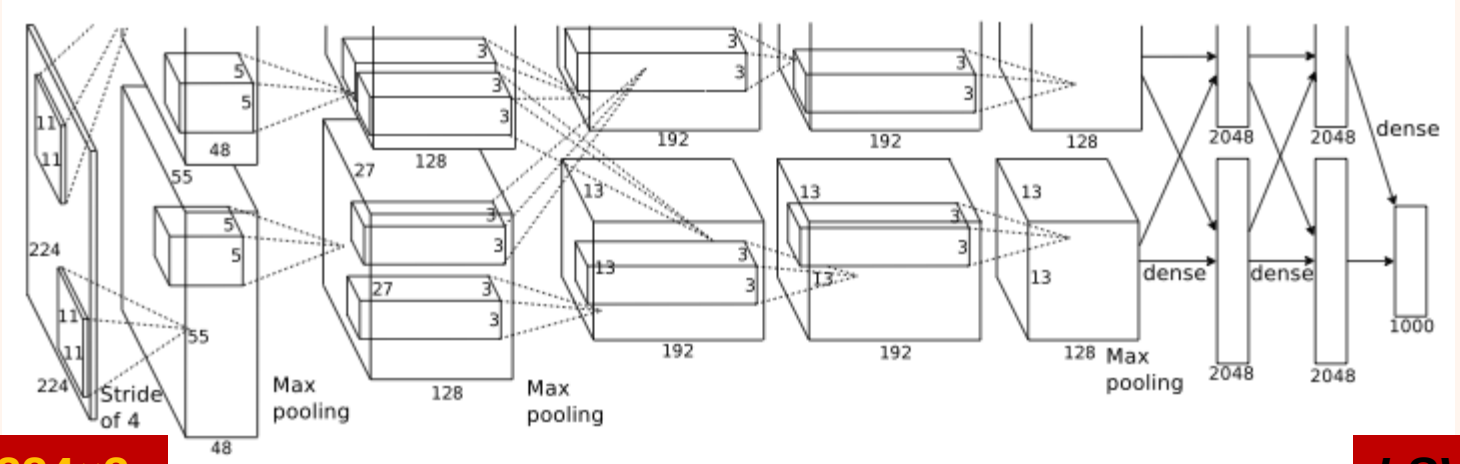
$$use \quad \frac{\partial E}{\partial w_1} + \frac{\partial E}{\partial w_2} \quad for\ w_1\ and\ w_2$$

تمام پارامترهای آزاد، با کمک الگوریتم پس‌انتشار خطا قابل آموزش می‌باشند.

آموزش(پس‌انتشار خطا)

- استفاده از پیش‌آموزش به صورت بی‌نظارت و سپس آموزش با الگوریتم پس‌انتشار خطا می‌تواند باعث افزایش کارایی 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.





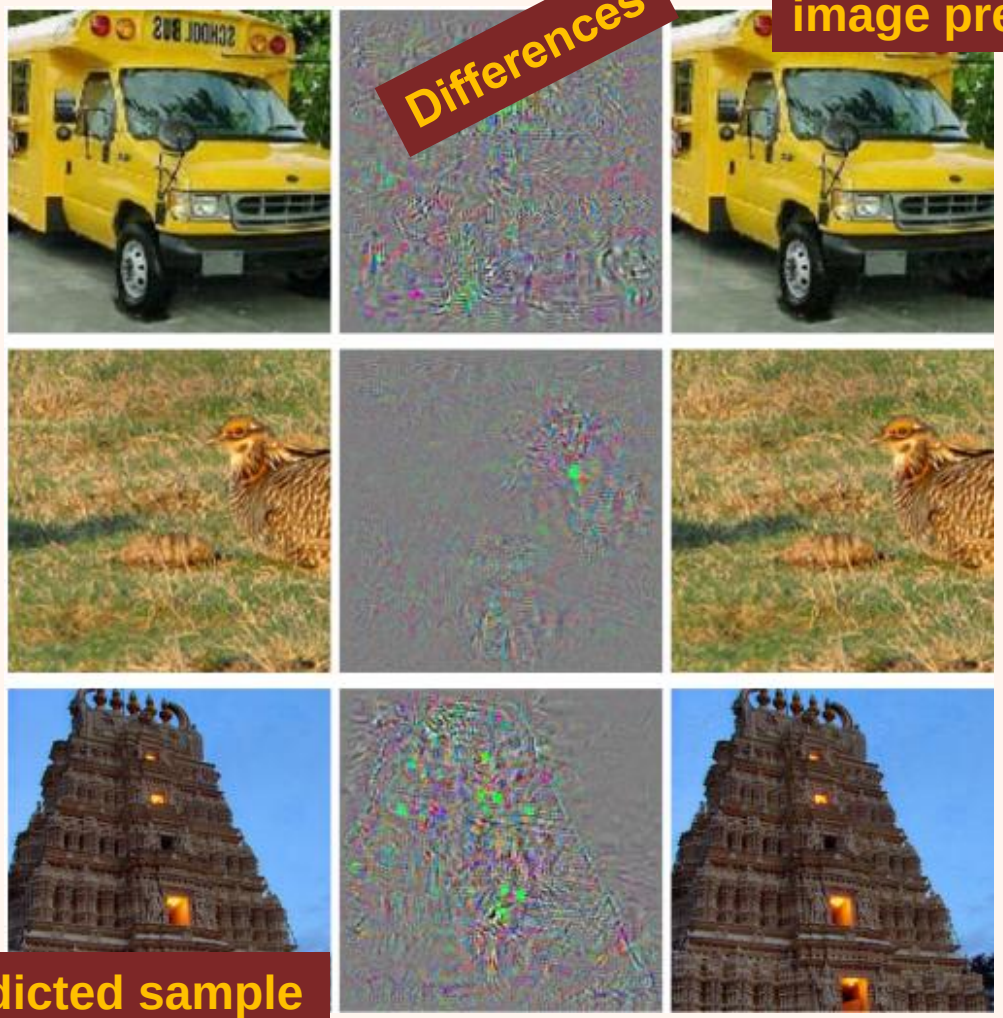
برای دسته‌بندی ۱,۲ میلیون تصویر
۱۰۰۰ دسته‌ی مختلف
۶۵۰۰۰۰ نورون و ۶۰ میلیون پارامتر آزاد
پنج لایه‌ی کانولوشن و سه لایه‌ی متصل
۳۷ و ۱۷ درصد خطای top5 و top1
از drop out استفاده می‌کند.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

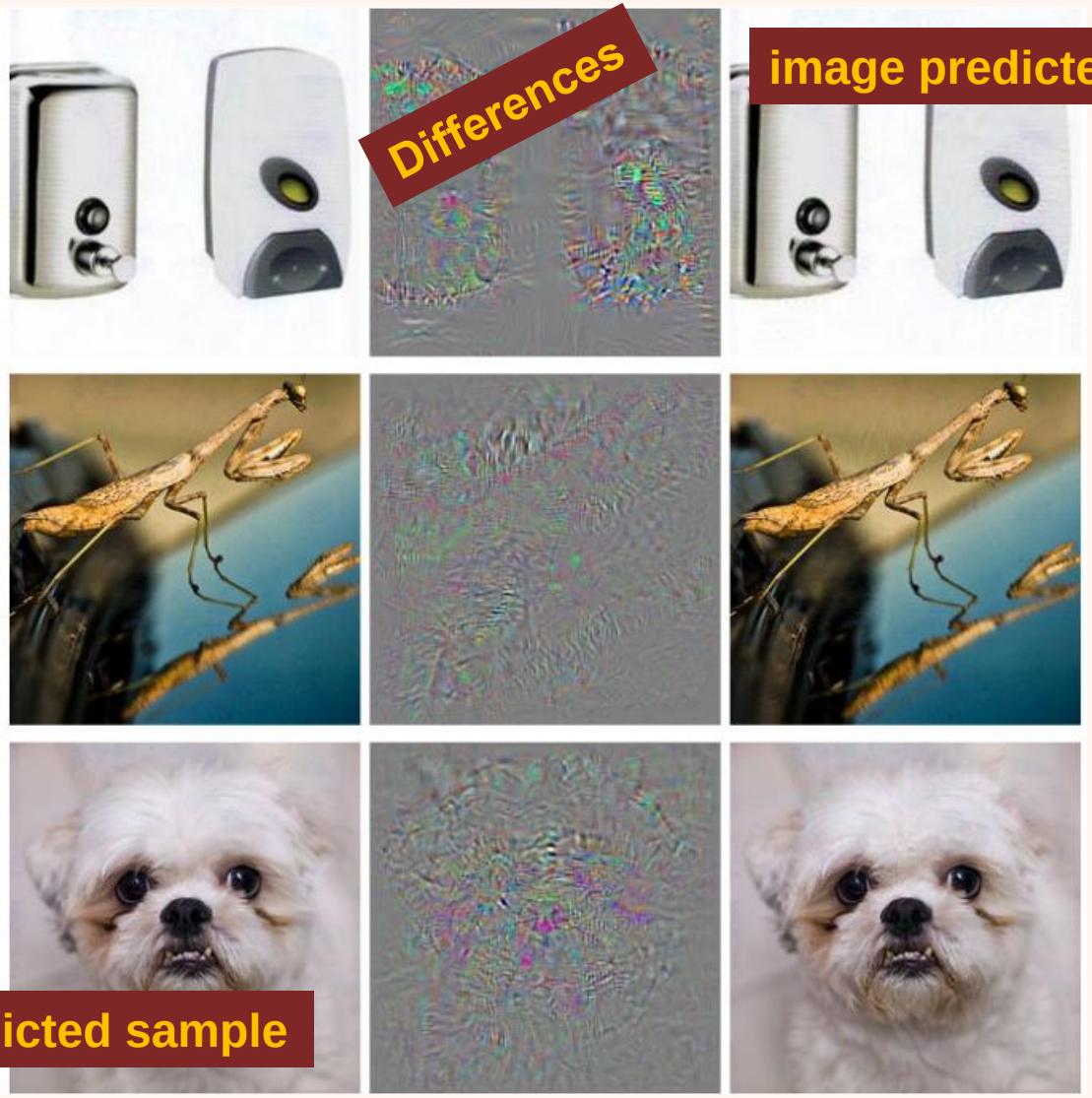
فریب شبکه‌های عمیق

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).

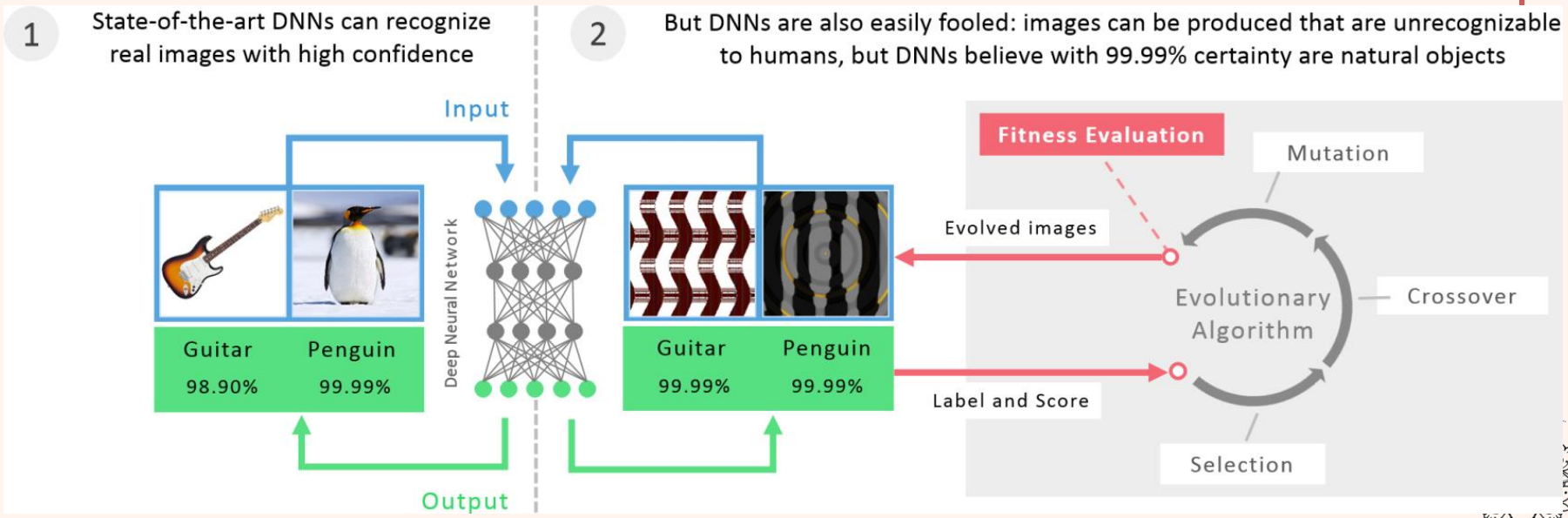
adversarial examples



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

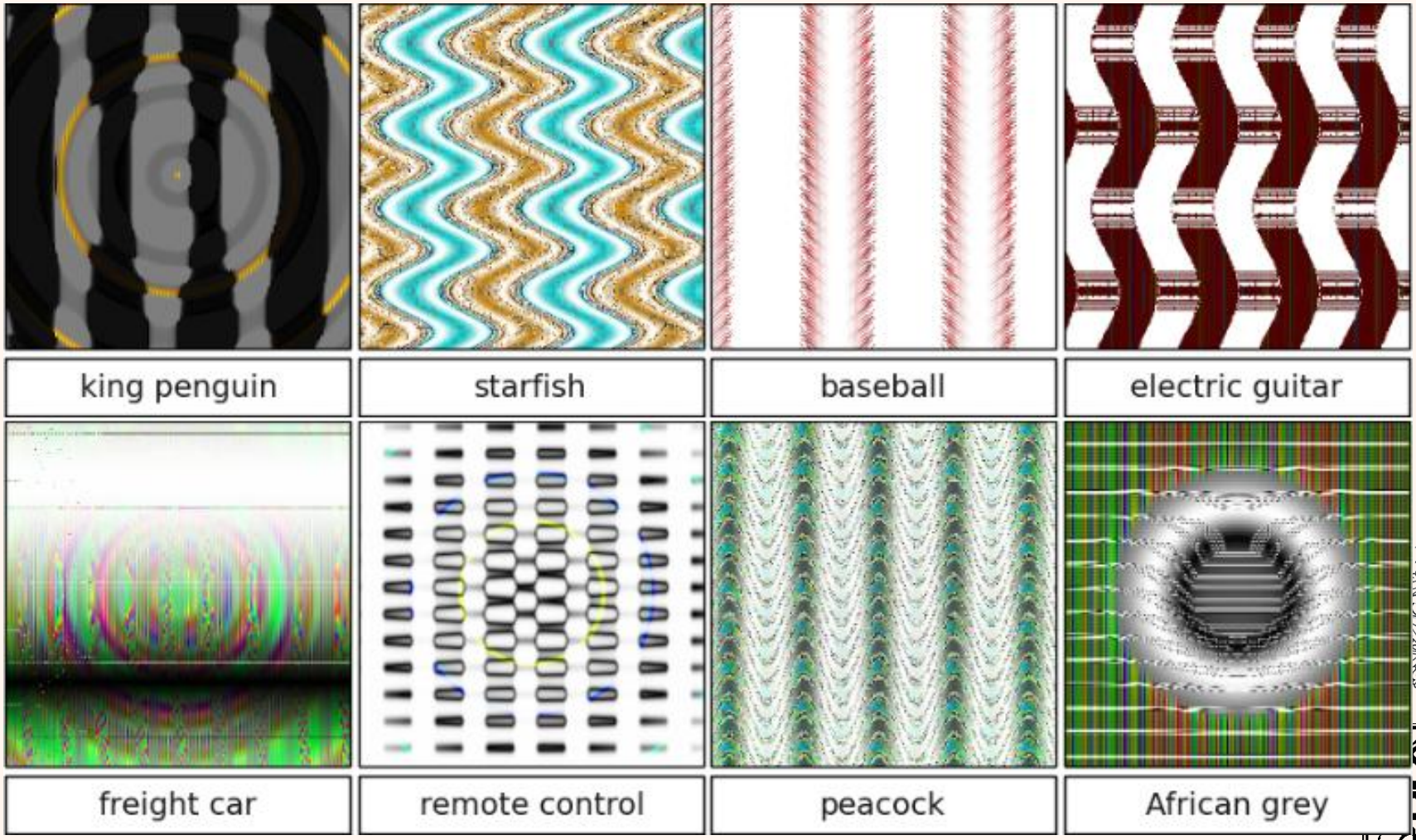

Differences
image predicted incorrectly
correctly predicted sample

فریب شبکه‌های عصبی عمیق



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



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