DERİN ÖĞRENME

Hande Alemdar

hande.alemdar@imag.fr

08.09.2016

TANITIM

- Derin Öğrenme

üretici	ayrıştırıcı

- Neden?

zor yapay zeka problemlerinde başarı

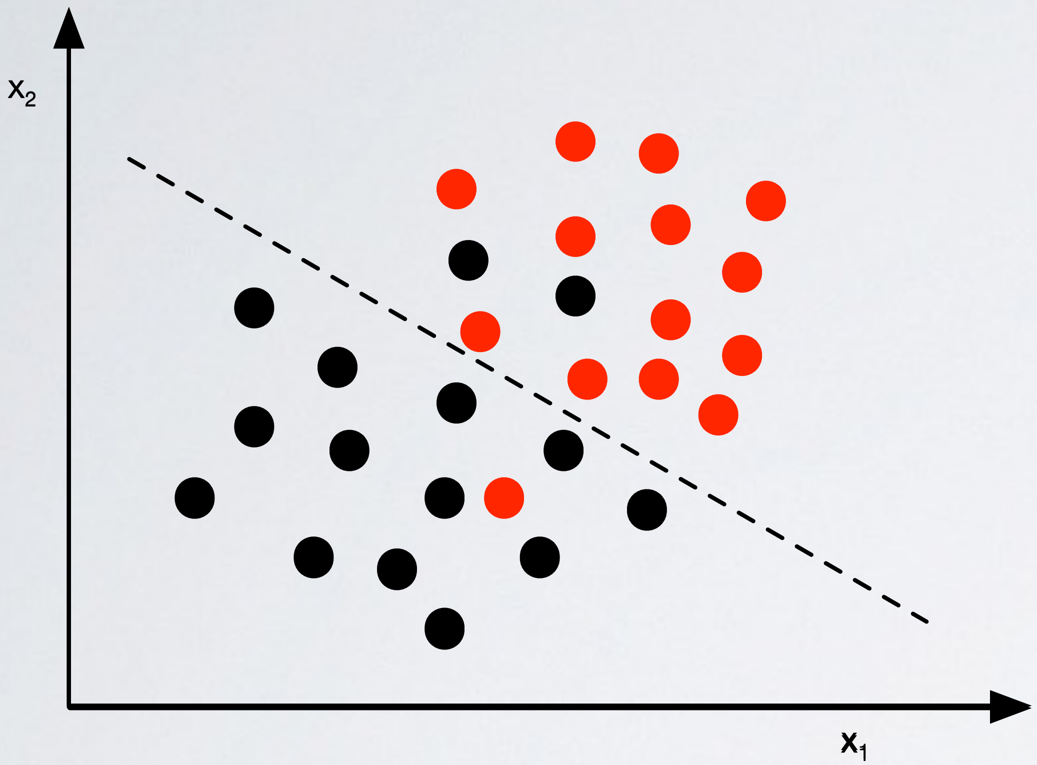
  - imge tanıma, konuşma tanıma, doğal dil işleme

- Neden şimdi?

yeterli veri ve işlem yapabilme kapasitesi

- Nasıl?
- Temel yapıtaşlarından kompleks yapılara

DOĞRUSAL SINIFLANDIRICI



Model $f(x) = Wx + b$

Eğitim

Parametreler $W = \begin{bmatrix} w_1 & w_2 \end{bmatrix}$ $b \in \mathcal{R}$

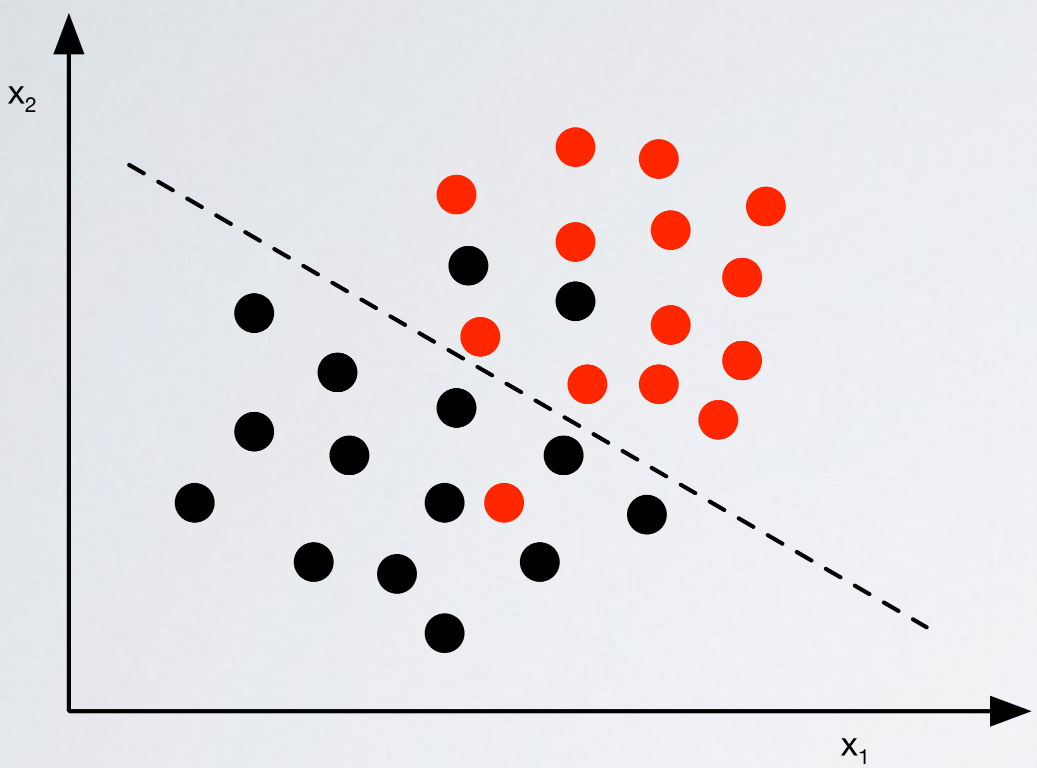
Girdi $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Çıkarsama

Sınıflandırıcı $h(x) = \begin{cases} 1, & \text{if } f(x) > 0 \\ 0, & \text{otherwise} \end{cases}$

Bağlanımcı $f(x)$

DOĞRUSAL PERSEPTRON

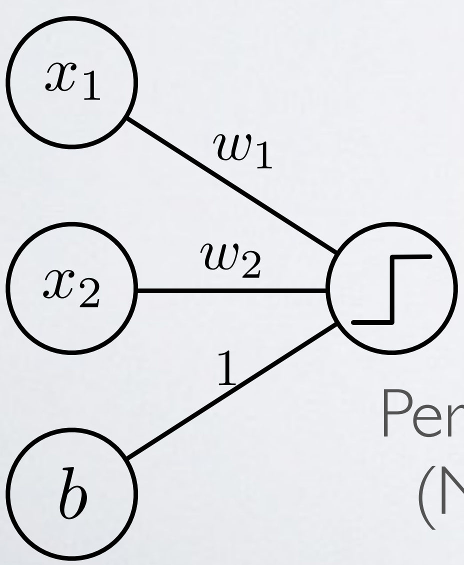


Transfer Fonksiyonu $$f(x) = Wx + b$$

Parametreler $$W = \begin{bmatrix} w_1 & w_2 \end{bmatrix}$$ Ağırlık

$$b \in \mathcal{R}$$ Sapkı

Girdi $$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



Perseptron (Nöron)

Etkinleştirme Fonksiyonu $$h(x) = \begin{cases} 1, & \text{if } f(x) > 0 \\ 0, & \text{otherwise} \end{cases}$$

PERSEPTRON EĞİTİMİ

$$W\mathbf{x} + b = \mathbf{y}$$

$$W'\mathbf{x}' = \mathbf{y} \qquad \mathbf{x}' = \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \qquad W' = \begin{bmatrix} W & b \end{bmatrix}$$

$$W'\mathbf{x}'\mathbf{x}'^{-1} = \mathbf{y}\mathbf{x}'^{-1}$$

$$W' = \mathbf{y}\mathbf{x}'^{-1}$$

- Doğrusal cebir ?
- Tekillik ?
- Matris evirme ! O(n^3)

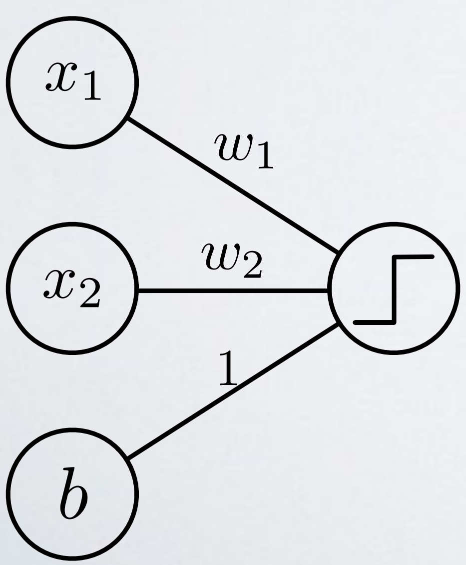
PERSEPTRON EĞİTİMİ

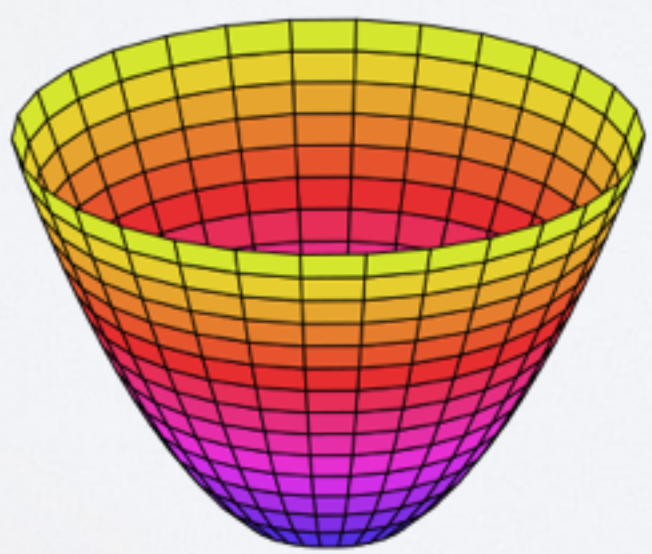
$$W\mathbf{x} + b = \mathbf{y}$$

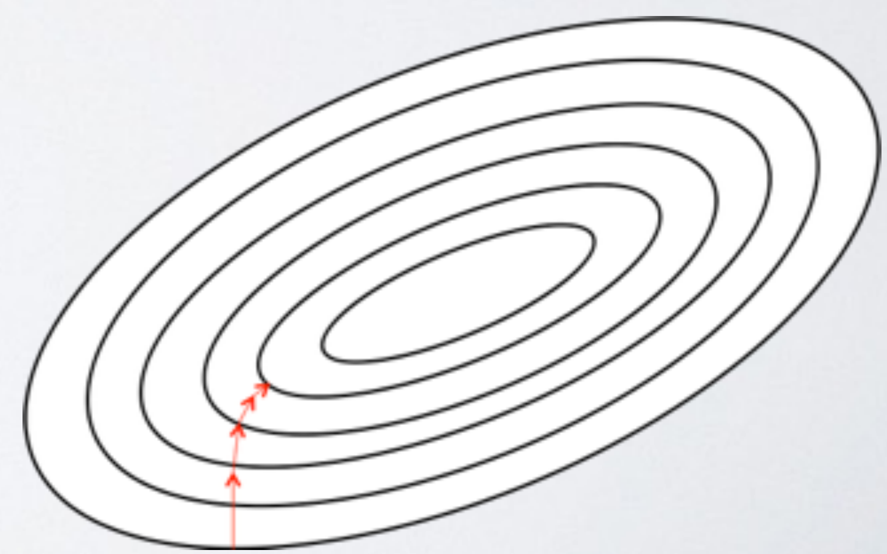
$$W^{(0)}\mathbf{x} + b^{(0)} = \mathbf{t} \qquad min(|\mathbf{t} - \mathbf{y}|)$$

Ceza (Maliyet) Fonksiyonu

$$E = \frac{1}{2}(\mathbf{t} - \mathbf{y})^2$$



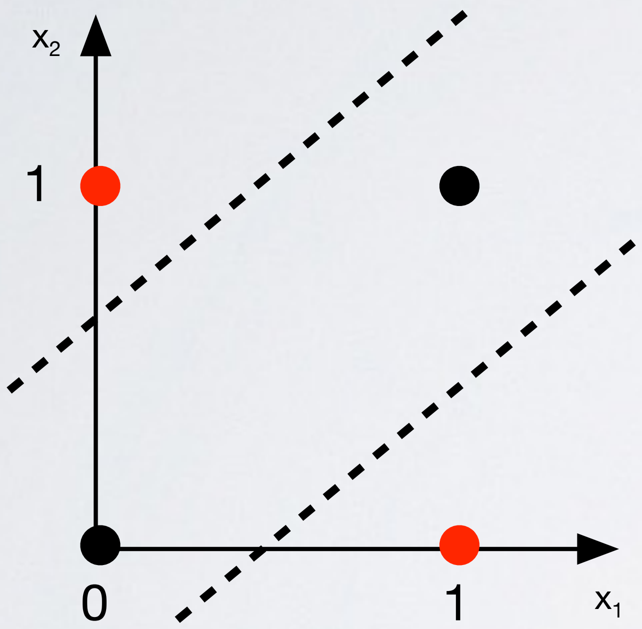




Gradyan İnişi

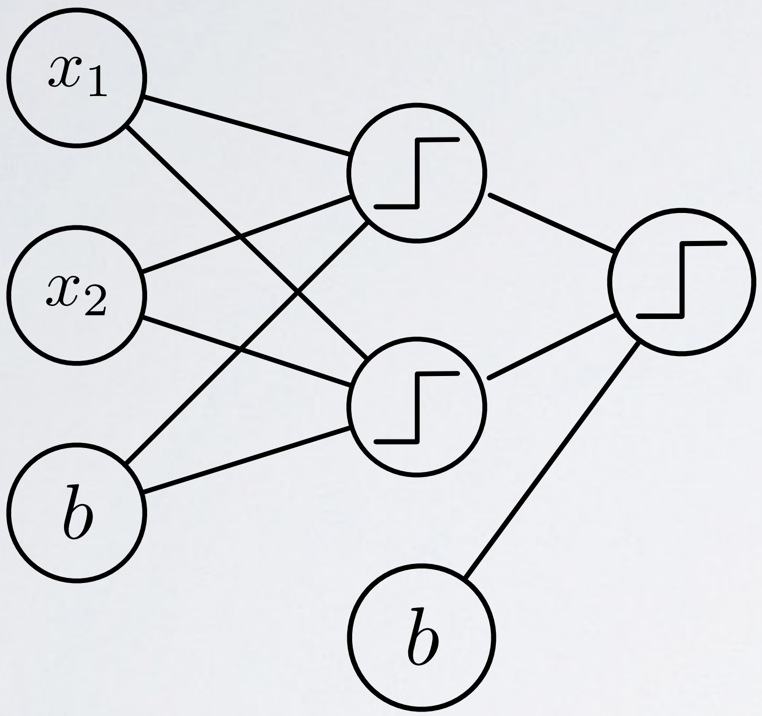
TEK PERSEPTRON KISITI

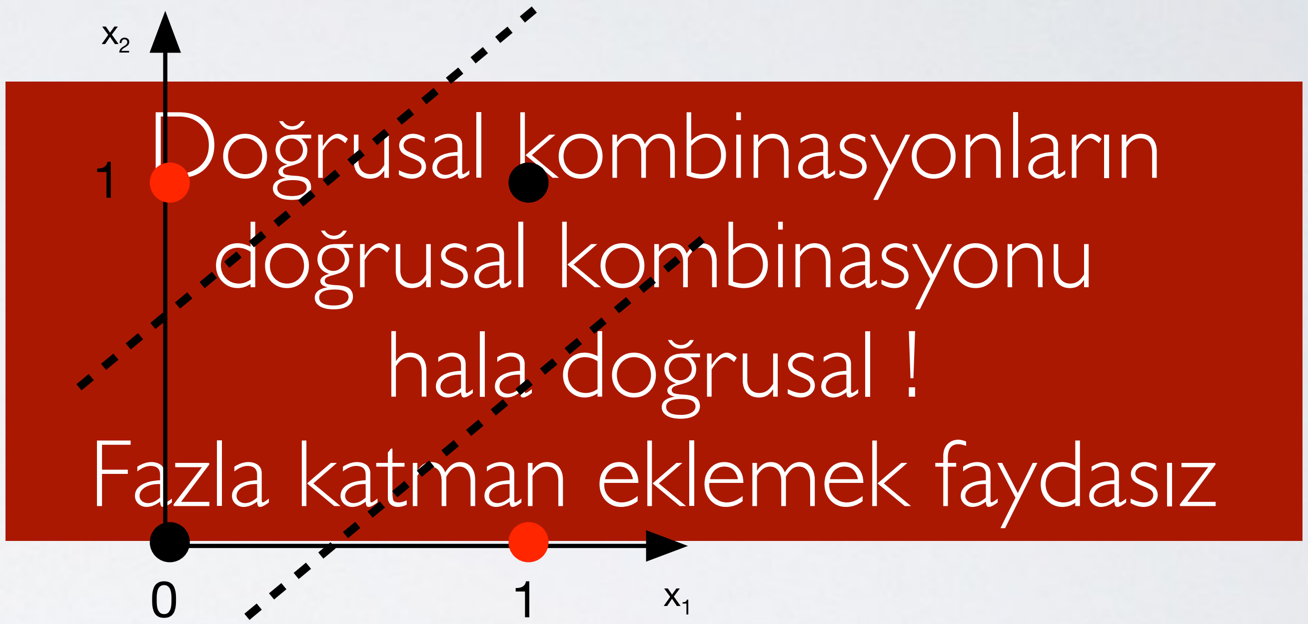
XOR Problemi



- Sadece doğrusal olarak ayrıştırılabilir fonksiyonlar
- Sadece iki sınıflı sınıflandırma

ÇOKLU PERSEPTRONLAR

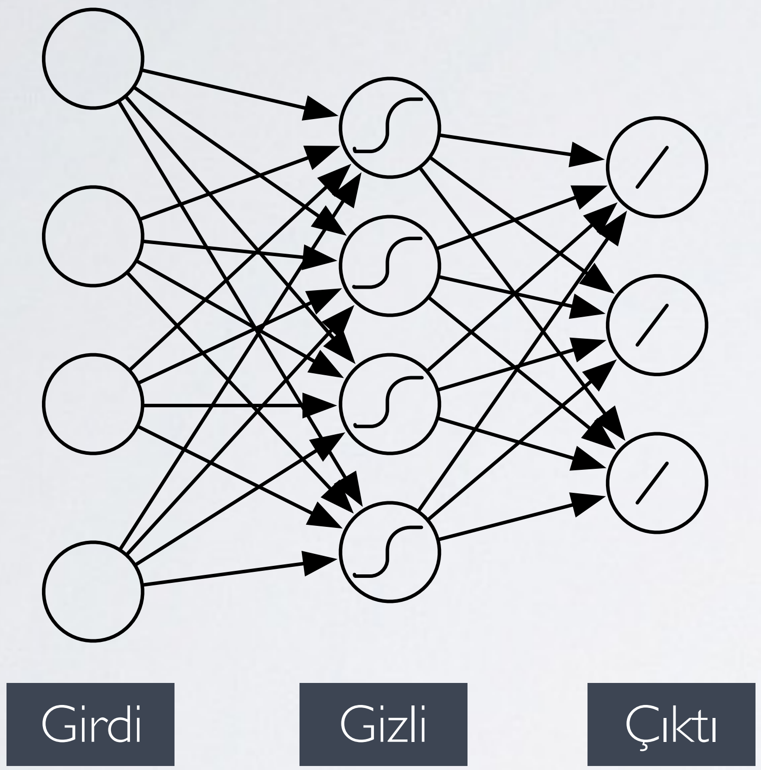


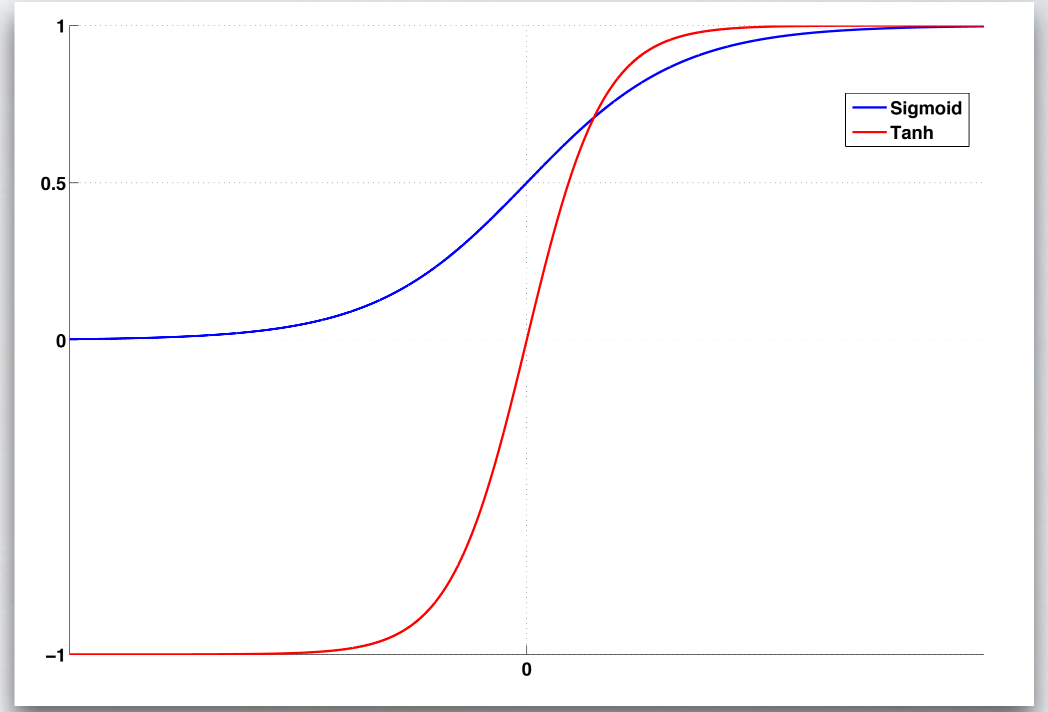


- XOR problemini çözüyor ancak hala doğrusal !
- Doğrusallıktan kurtulma
    - Girdi Katmanında - Karar Destek Makineleri (SVM)
    - Etkinleştirme Fonksiyonunda - Çok-katmanlı Yapay Sinir Ağları

ÇOK-KATMANLI PERSEPTRON

Çok-Katmanlı Yapay Sinir Ağları
İleri-Besleme Yapay Sinir Ağları

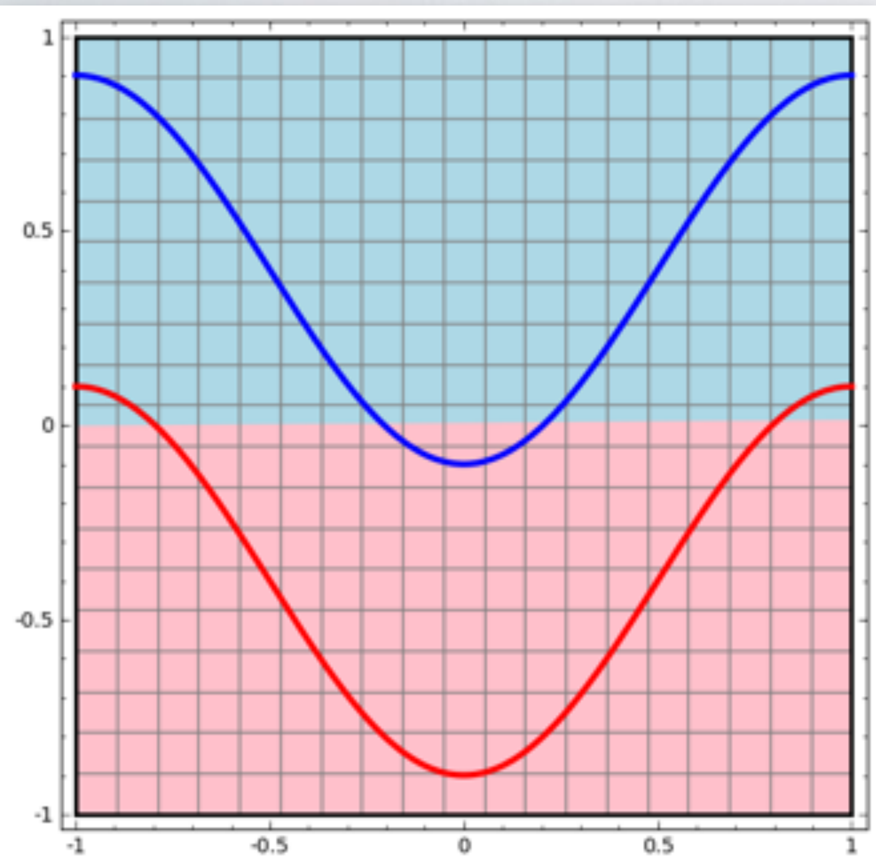




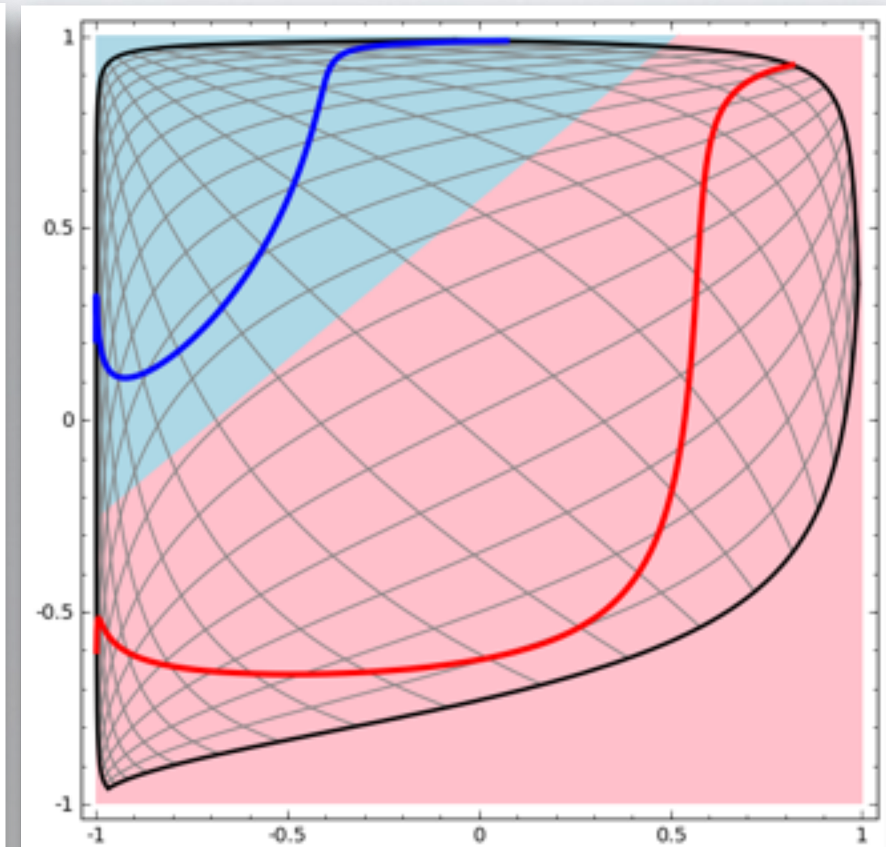
- 1 Girdi Katmanı
- Doğrusal-olmayan gizli katmanlar (1 ya da daha fazla)
- 1 Çıkış Katmanı
- Doğrusal-olmayan Etkinleştirme Fonksiyonu

Gizli Katmanlar çoğaldıkça Derin YSA

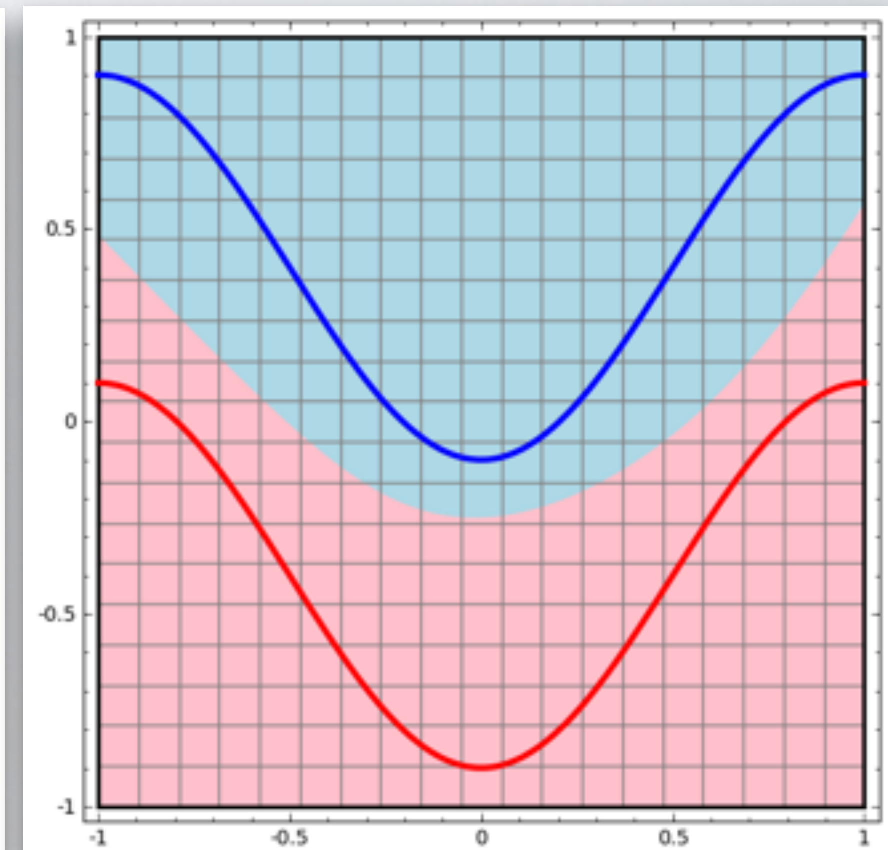
NASIL ÇALIŞIYOR



Doğrusal Sınıflandırma
başarısız



Gizli Katmanlar
Girdinin doğrusal olmayan
temsillerini oluşturuyor

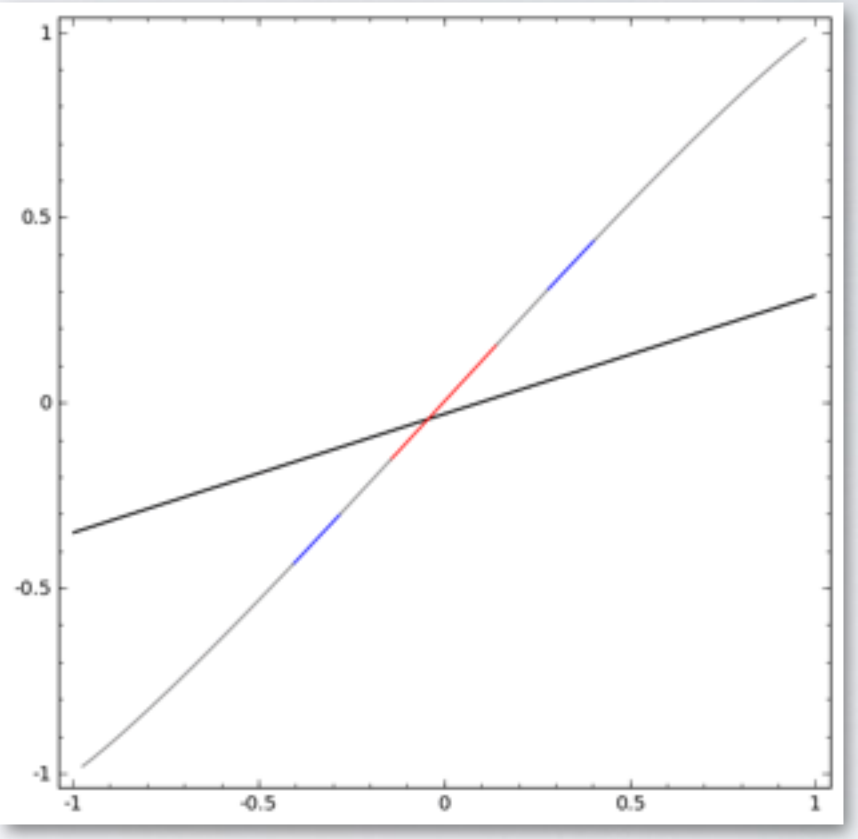


Doğrusal sınıflandırma
girdinin bu yeni temsilinde
başarılı

Derin Öğrenme
Temsilleri Otomatik Buluyor

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

NASIL ÇALIŞIYOR



Sınıflandırması kolay olmayan
tek boyutlu veri

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

NASIL EĞİTİYORUZ

$$W\mathbf{x} + b = \mathbf{y}$$

$$W^{(0)}\mathbf{x} + b^{(0)} = \mathbf{t} \qquad min(|\mathbf{t} - \mathbf{y}|)$$

Maliyeti geriye doğru ağırlıklara yansıtarak yayıyoruz

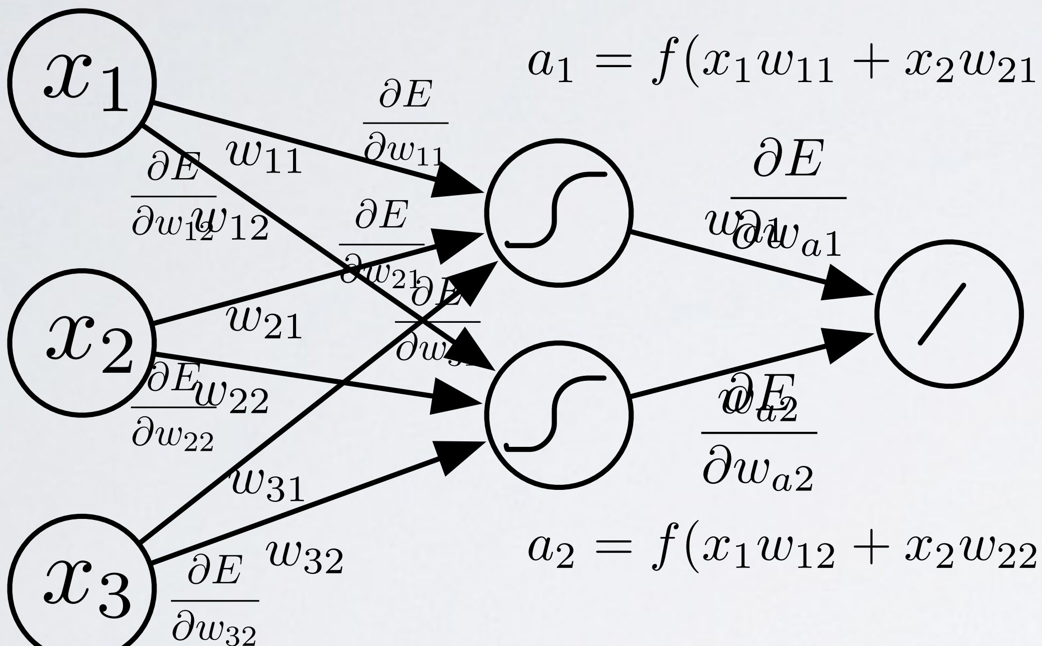
Ceza (Maliyet) Fonksiyonu

$$E = \frac{1}{2}(\mathbf{t} - \mathbf{y})^2$$

$$w_{new} = w_{current} + \eta \frac{\partial E}{\partial w}$$

- Tek nöron ile hemen hemen aynı şekilde ama biraz daha karışık matematik kullanarak
- Gradyan-inişli Geriye Yayılım Yöntemi

GERİYE YAYILIM YÖNTEMİ



$$a_1 = f(x_1 w_{11} + x_2 w_{21} + x_3 w_{31})$$

$$E = e(y,t) = \frac{1}{2}(y-t)^2$$

$$t = a_1 w_{a1} + a_2 w_{a2}$$

$$a_2 = f(x_1 w_{12} + x_2 w_{22} + x_3 w_{32})$$

$$w_{new} = w_{current} + \eta \frac{\partial E}{\partial w}$$

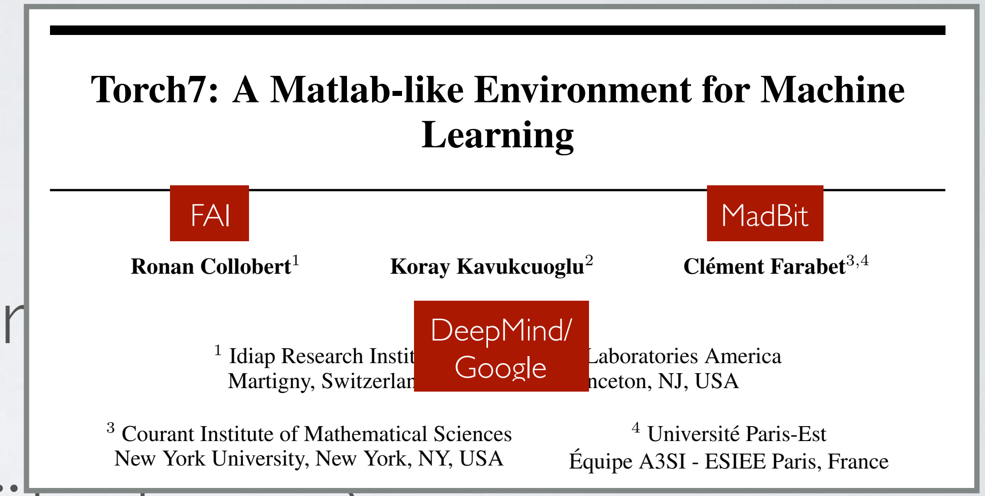
öğrenme oranı

TORCH 7

- Yapay öğrenme için MATLAB benzeri bir ortam
- Lua temelli nümerik işlem çerçevesi ve yapay öğr
- N-boyutlu Tensör object, (tek başına C Tensör kütüphanesi)
- OpenMP desteği
- CUDA desteği

Ek geliştirme gerektirmeden hazır

- 8 hazır kütüphane paketi
    - torch, nn, plot, qt, image, optim, unsup



Torch7: A Matlab-like Environment for Machine Learning

Ronan Collobert[1] Koray Kavukcuoglu[2] Clément Farabet[3,4]

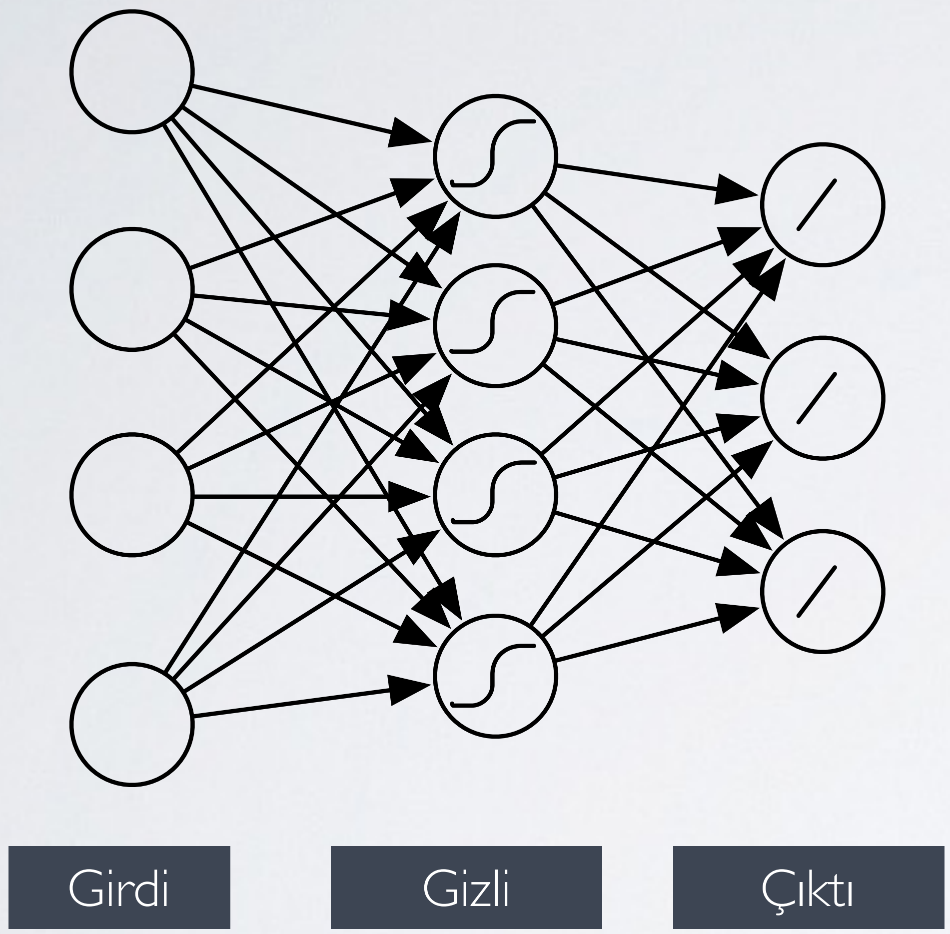
[1] Idiap Research Instit
Martigny, Switzerlan

Laboratories America
nceton, NJ, USA

[3] Courant Institute of Mathematical Sciences
New York University, New York, NY, USA

[4] Université Paris-Est
Équipe A3SI - ESIEE Paris, France

NN PAKETİ



Ağı Oluşturma

```
mlp = nn.Sequential()
mlp:add( nn.Linear(4, 4) ) -- 4 girdi, 4 gizli nöron
mlp:add( nn.Tanh() )       -- etkinleştirme fonksiyonu
mlp:add( nn.Linear(4, 3) ) -- 3 çıktı nöronu
```

Eğitme $E = \frac{1}{2}(\mathbf{t} - \mathbf{y})^2$

```
criterion = nn.MSECriterion() -- Ortalama Hata Karesi
trainer = nn.StochasticGradient(mlp, criterion)
trainer:train(dataset)      -- veri seti üzerinde eğit
```

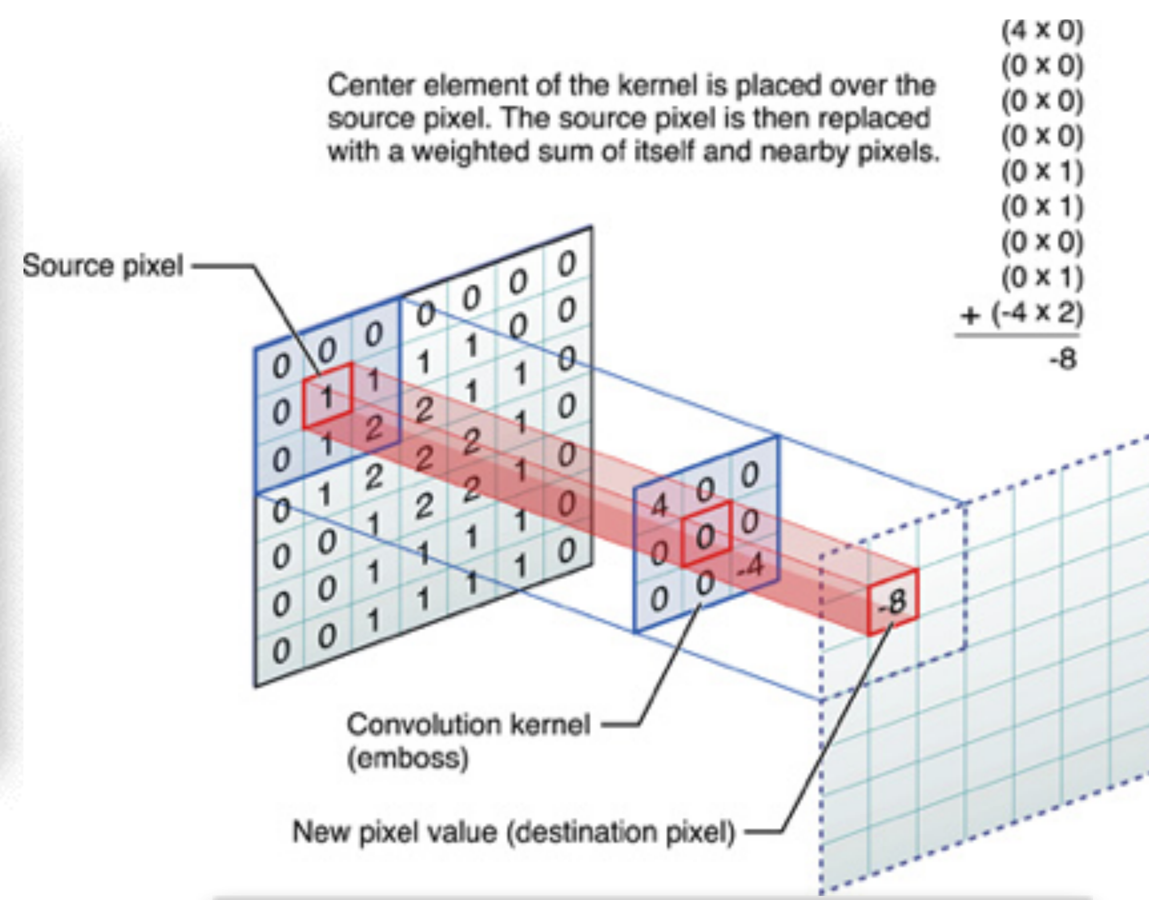
http://nn.readthedocs.org/en/latest/

EVRİŞİMSEL SİNİR AĞLARI

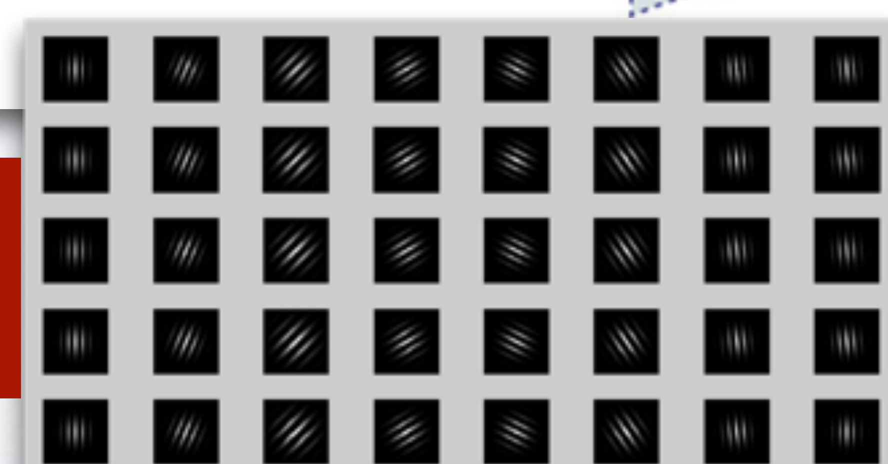
Filtre kullanarak Evrişim

Original

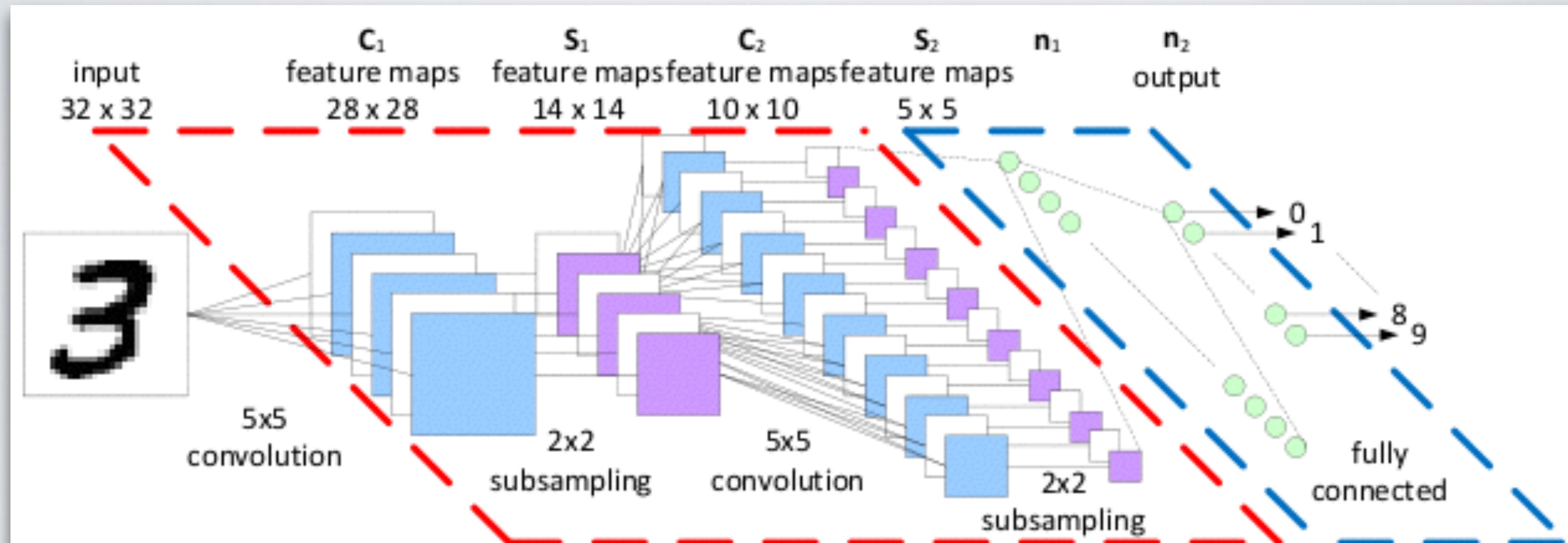
Emboss



Gabor Filtre Bankası kenar çıkarımı için



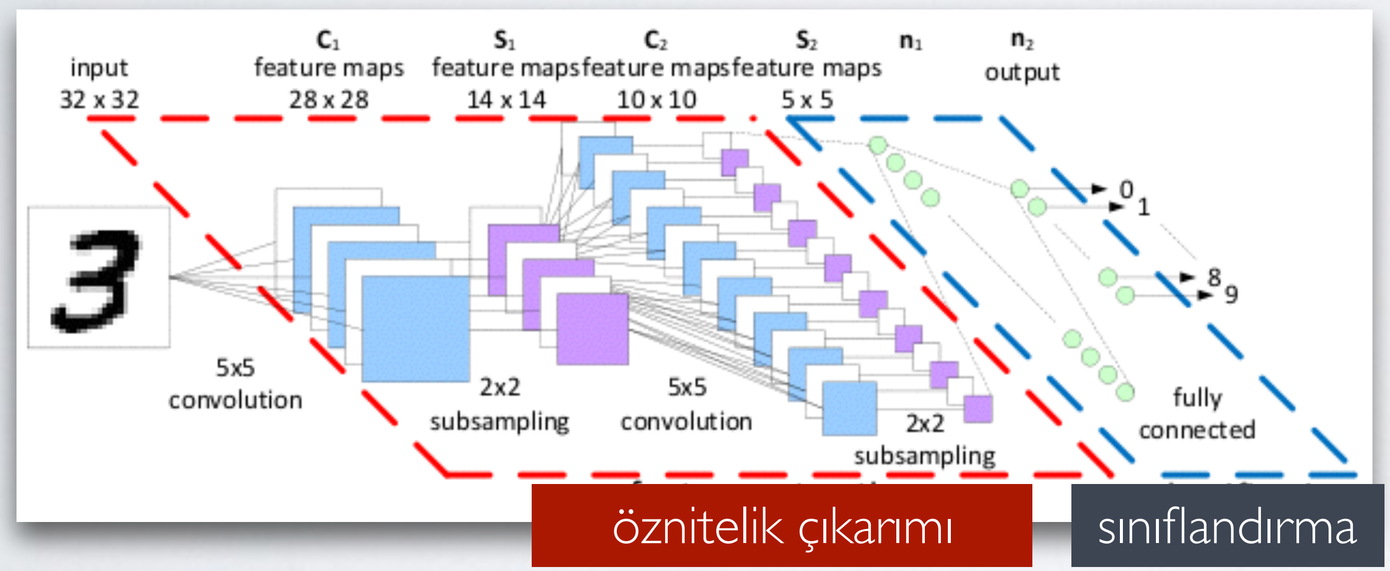
EVRİŞİMSEL SİNİR AĞLARI (ESA)



öznitelik çıkarımı

sınıflandırma

TORCH İLE ESA




```
model = nn.Sequential()

-- aşama 1 : 6 adet 5x5 filtre

model:add(nn.SpatialConvolution(1, 6, 5, 5))  -- çıktı 28x28

model:add(nn.Tanh())

model:add(nn.SpatialMaxPooling(2, 2, 2, 2))    --çıktı 14x14

-- aşama 2 : 16 adet 5x5 filtre

model:add(nn.SpatialConvolution(6, 16, 5, 5))  --çıktı 10x10

model:add(nn.Tanh())

model:add(nn.SpatialMaxPooling(2, 2, 2, 2))    --çıktı 5x5

-- aşama 3 : standart 2 katmanlı YSA

model:add(nn.Reshape(16*5*5))

model:add(nn.Linear(16*5*5, 120))

model:add(nn.Tanh())

model:add(nn.Linear(120, #classes))
```

IMAGENET ÜZERİNDE ESA

- Derin: 7 gizli katman
  - 5 evrişimsel
  - 2 tam-bağlantılı
- 2 GPU üzerinde 1 hafta eğitim
- 650,000 nöron
- 60,000,000 parametre
- 630,000,000 bağlantı

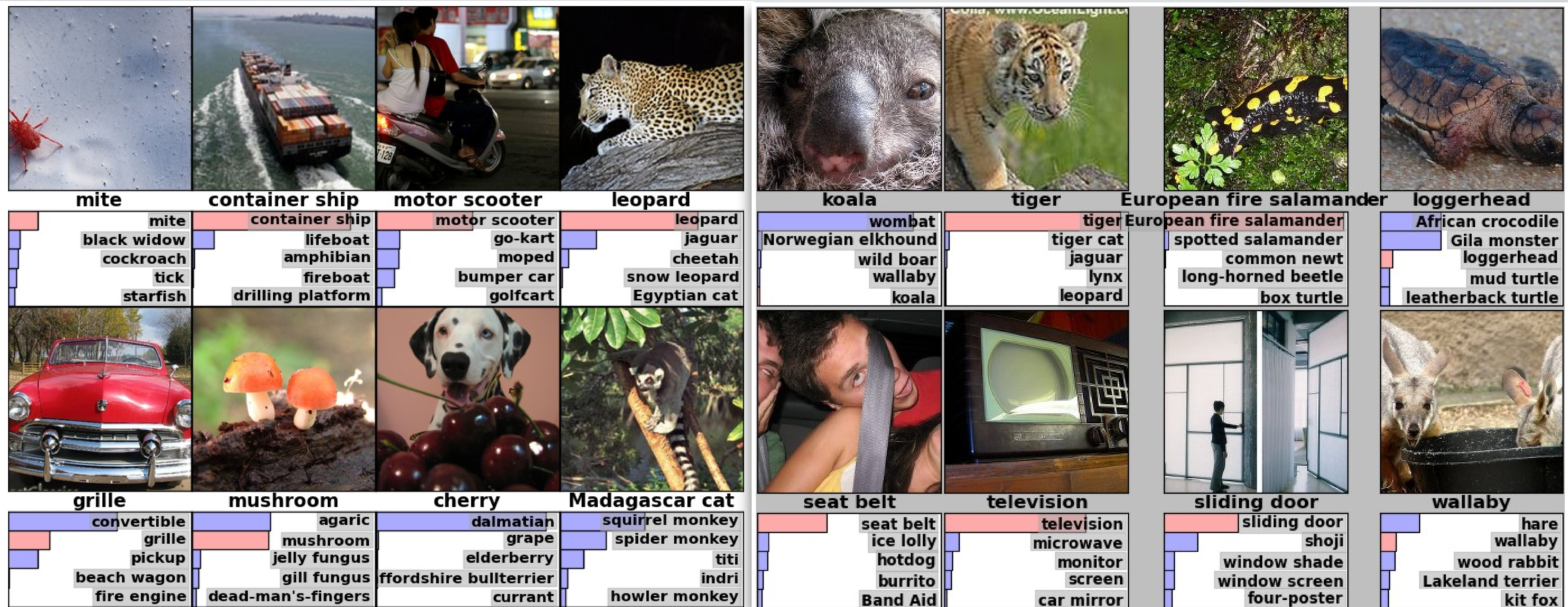
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton

NIPS 2012

Öğrenilen Filtreler

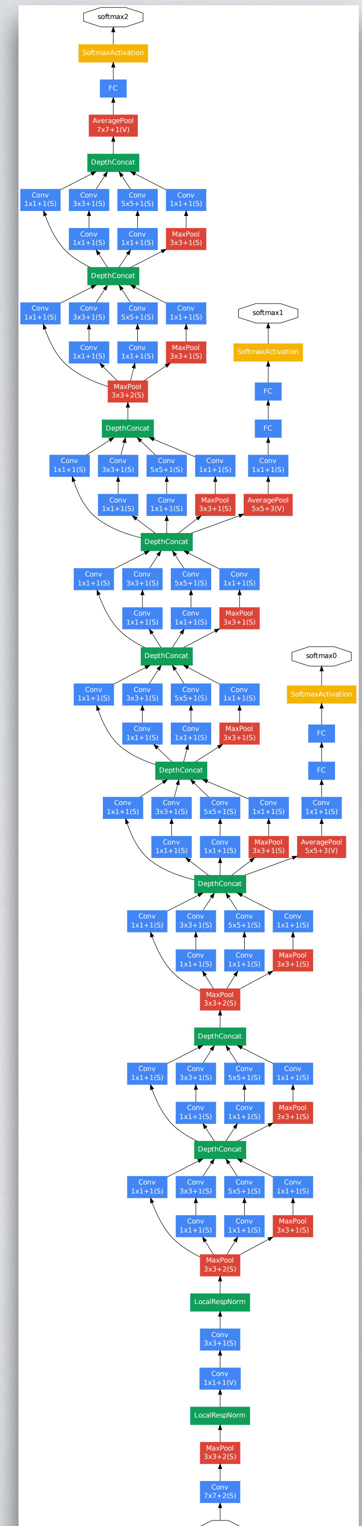
ÖRNEKLER



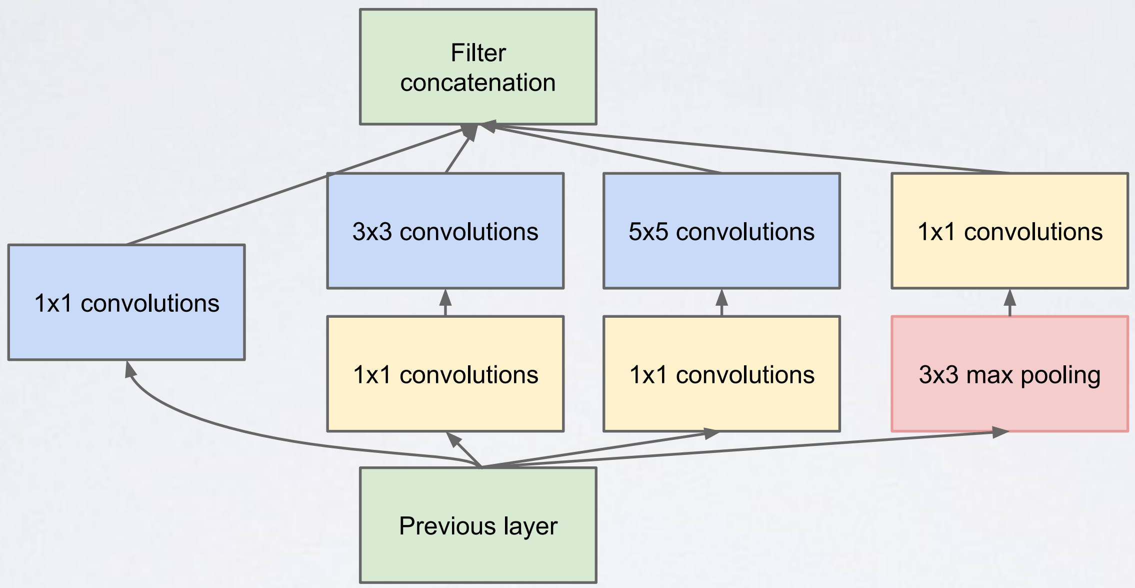
ImageNet Large Scale Visual Recognition Challenge (ILSVRC) - 2012 Kazananı

16.4 % top-5 hata 1000 sınıf, 1.2 milyon fotoğraf

ILSVRC 2014



GoogLeNet 6.67% top-5 hata

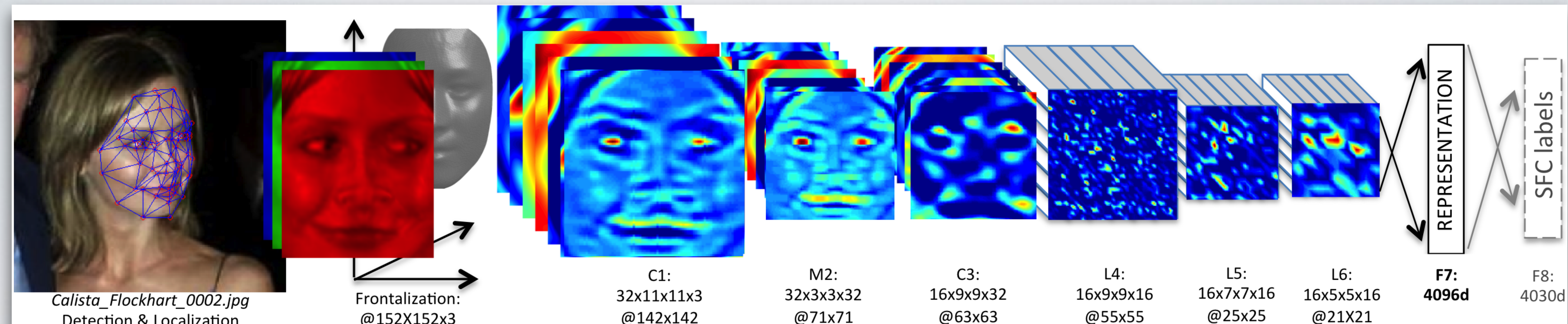


inception modulleri

Siberya Haskisi

Eskimo Köpeği

FACEBOOK DERİN YÜZ

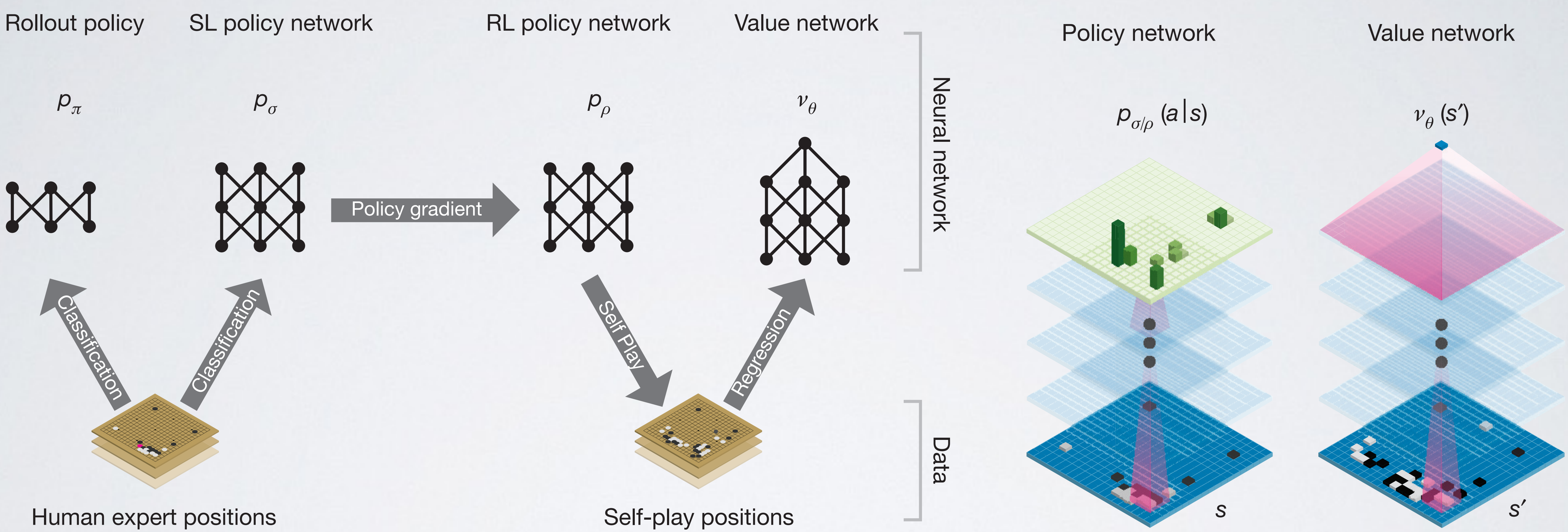


97.35% doğruluk oranı. Labeled Faces in the Wild (LFW) veri seti

İnsan performansı 97.5 %

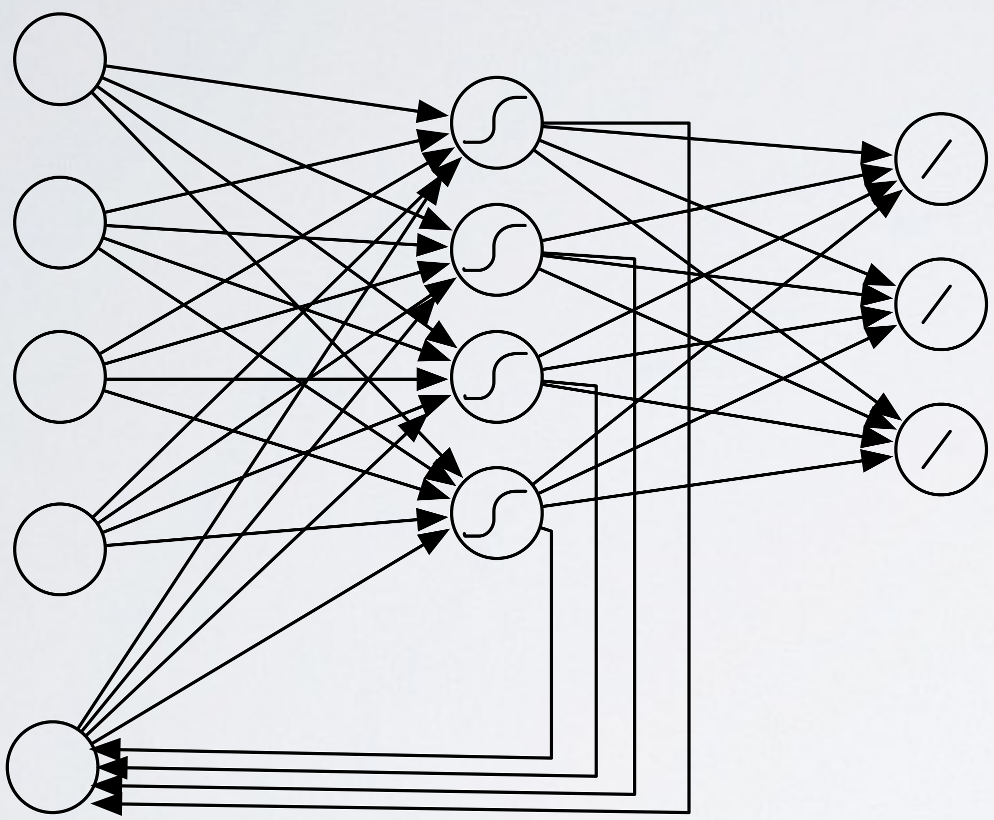
DeepFace: Closing the Gap to Human-Level Performance in Face Verification
Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf
Conference on Computer Vision and Pattern Recognition (CVPR 2014)

ALPHA GO



Rollout policy
SL policy network
RL policy network
Value network
p_π
p_σ
p_ρ
v_θ
Policy gradient
Neural network
Classification
Classification
Self Play
Regression
Data
Human expert positions
Self-play positions
Policy network
Value network
$p_{\sigma/\rho}(a|s)$
$v_\theta(s')$
s
s'

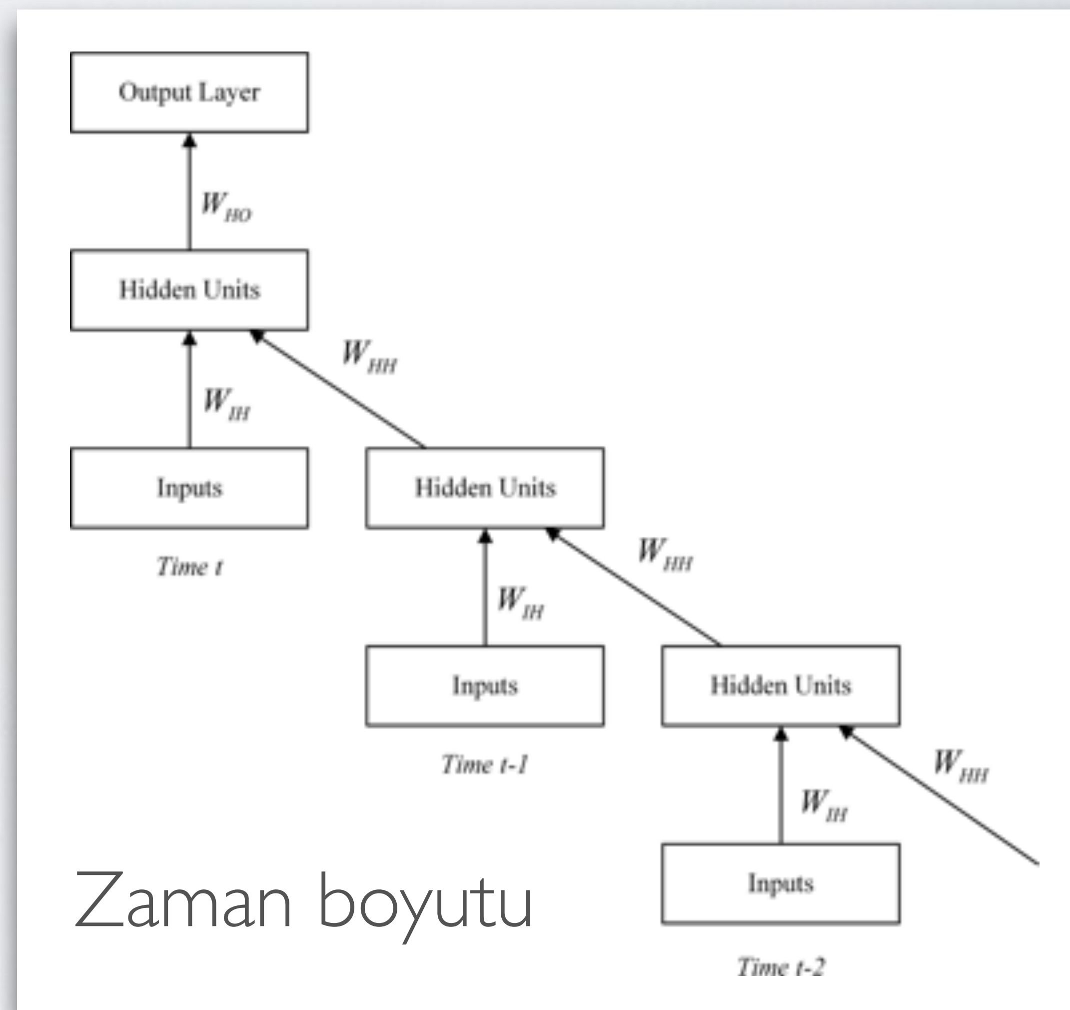
ÖZYİNELEMELİ YAPAY SİNİR AĞLARI



bağlam birimleri

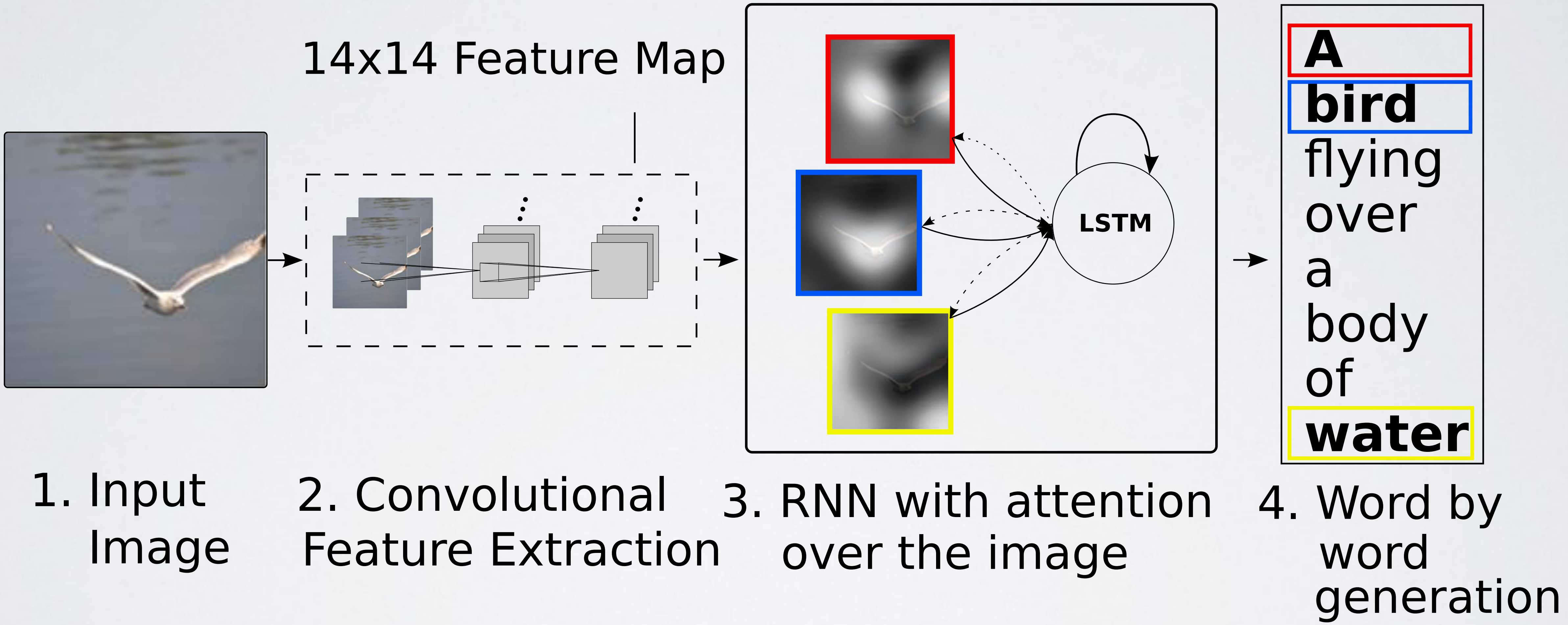
DDİ, metin madenciliği, otomatik çeviri

video, imge işleme



Zamanda Geriye Doğru Yayılım Yöntemi

ÖRNEK : FOTOĞRAF BAŞLIĞI OLUŞTURMA



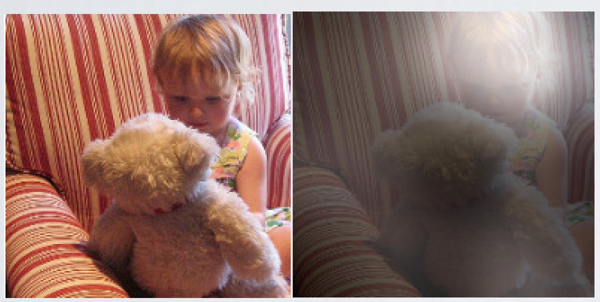
Show, Attend and Tell: Neural Image Caption Generation with Visual Attention (2015)
K. Xu , J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio

ÖRNEKLER

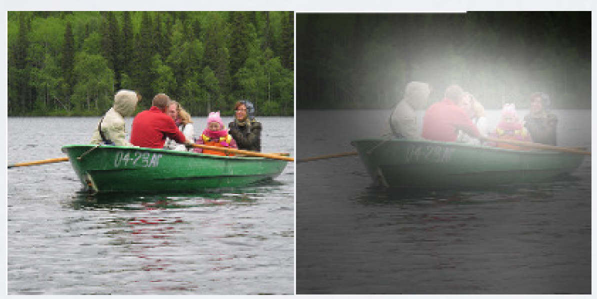
A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

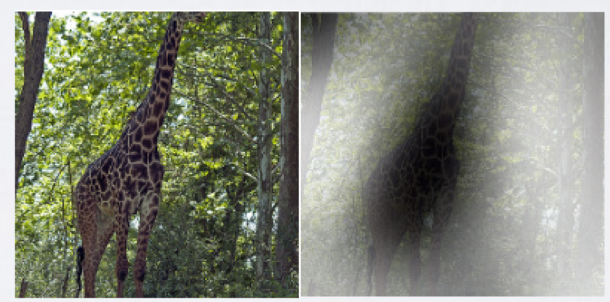
A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.

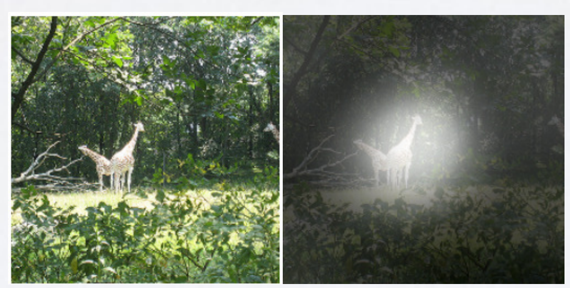


A group of people sitting on a boat in the water.

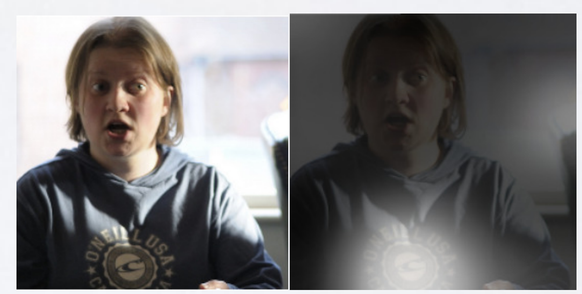


A giraffe standing in a forest with trees in the background.

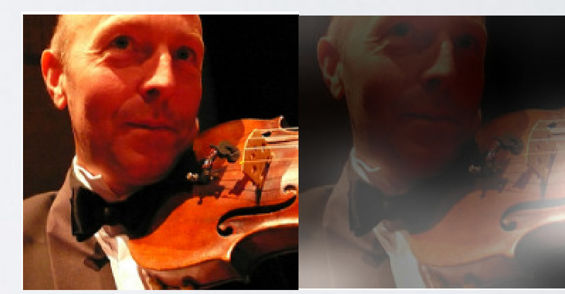
Başarılı



A large white bird standing in a forest.

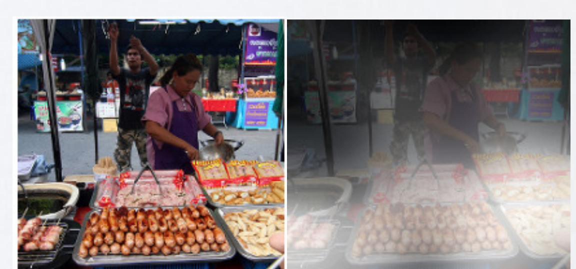


A woman holding a clock in her hand.



A man wearing a hat and a hat on a skateboard.

A person is standing on a beach with a surfboard.



A woman is sitting at a table with a large pizza.

A man is talking on his cell phone while another man watches.

Başarısız

ÖYSA İLE EĞLENCE

writing.html

TED-RNN—Machine generated TED-Talks

Ideas worth generating

Recently, I experimented with Recurrent Neural Networks as applied to political speech writing. The result, **Obama-RNN** generated a fair amount of buzz and fuelled a productive debate, including speculation that Sara Palin is a bot. For the next experiment I decided to apply the system to TED.

TED Ideas worth generating

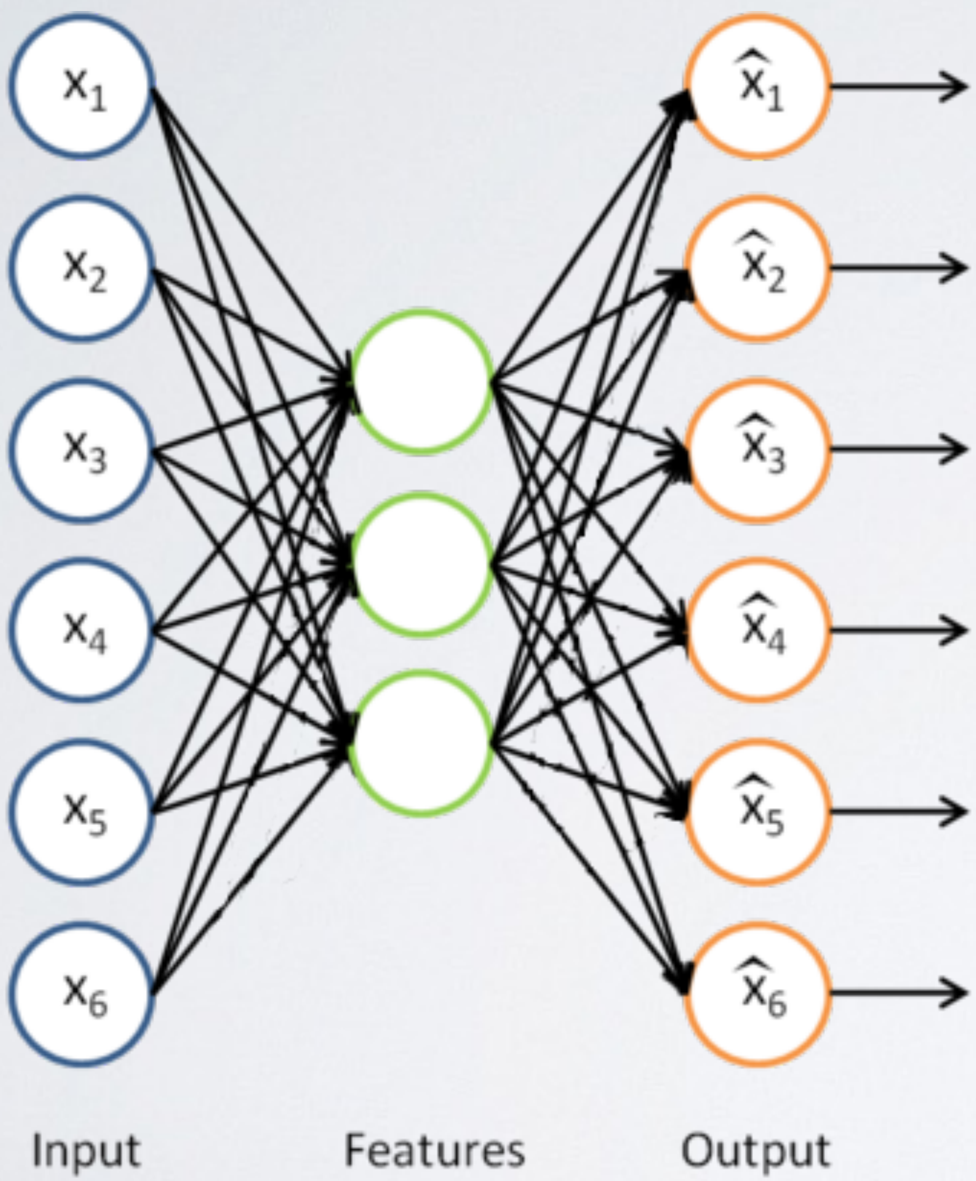
TED has become somewhat of an institution with the Silicon Valley intelligentsia. Over the years, they have accumulated **1904 speeches or 22.4MB of text which amounts to 4038409 words.** All transcripts are readable here and cover a wide range of topics, from scientific to cultural.

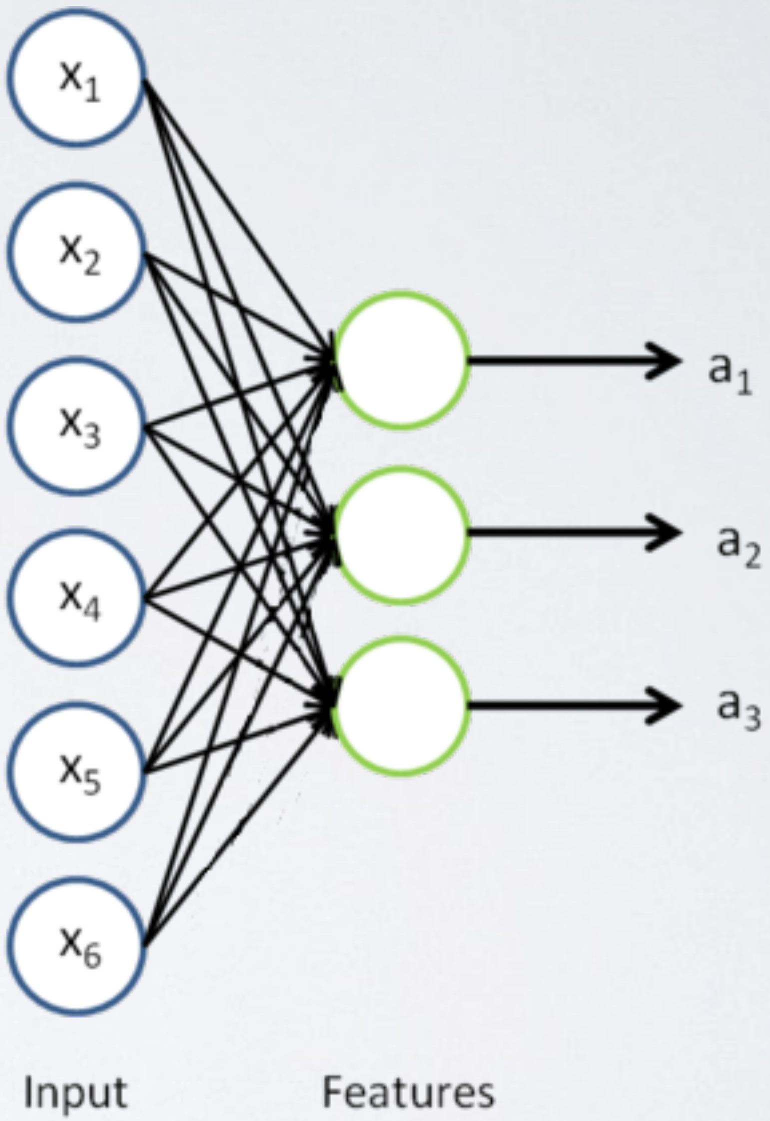
üretici model

denetimsiz öğrenme

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

ÖZKODLAYICI



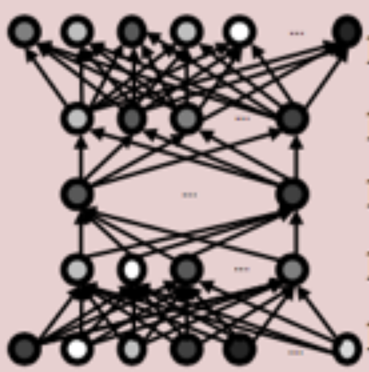
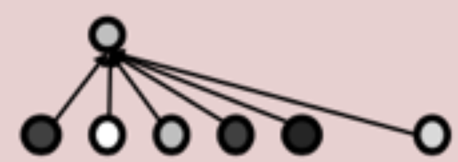
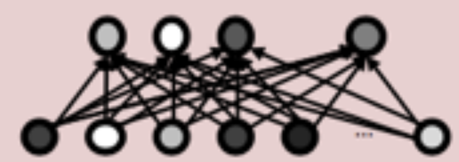
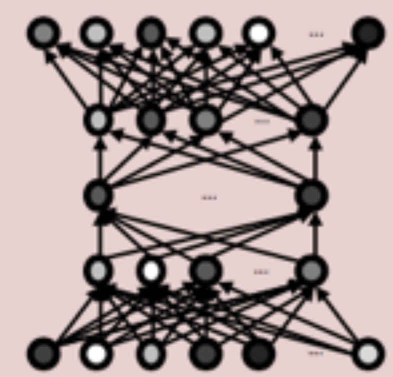


SONUÇLAR

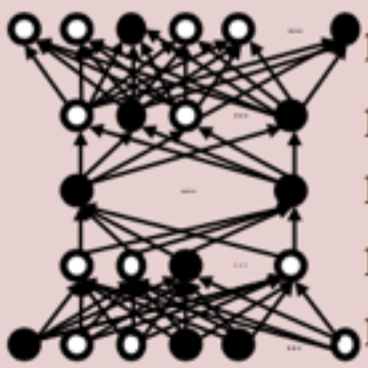
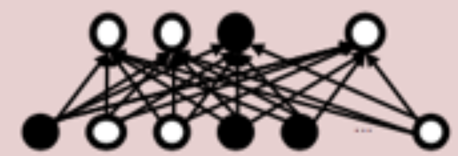
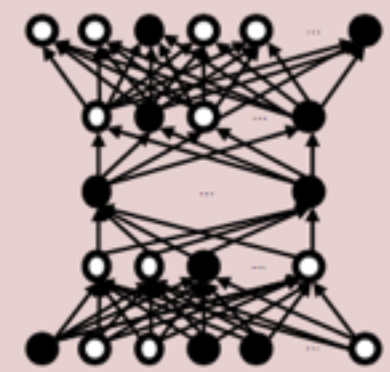
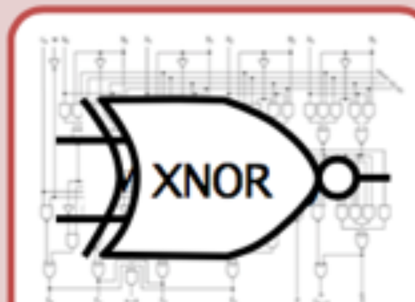
- Begin gibi işlem yapma 1940lardan beri var
- Büyük Veri + Yüksek Kapasiteli İşlem Gücü
- Çok etkili öznitelik çıkarma kapasitesi
- LEGO parçaları gibi
- Daha karmaşık ağlar için
    - daha çok veri, daha çok deneme, daha çok işlem gücü

ENERJİ VERİMLİ DERİN ÖĞRENME

SORUN

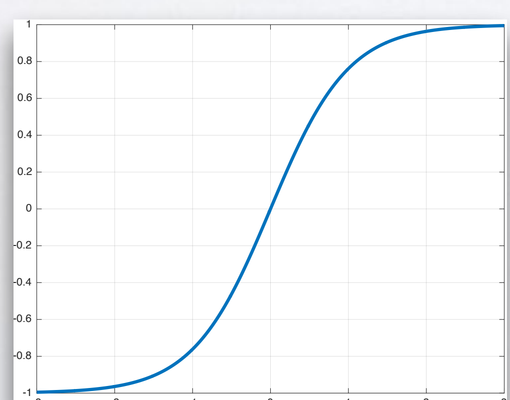
	Bağlantı	Nöron	Katman	Ağ
				
İşlem	Çarpma Toplama	1K 1K	2 Million FLOPS	8 Million FLOPS
Hafıza	32 bit	32 K bit	32 Million bit =4MB	16 MB

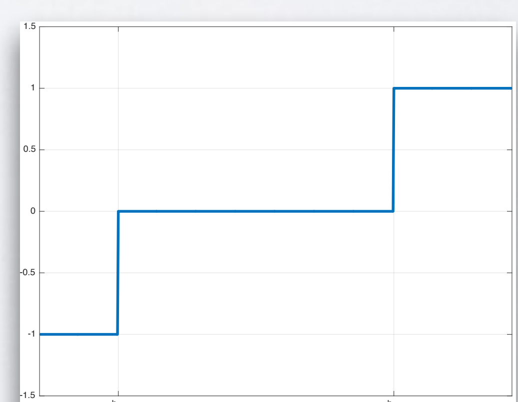
SORUN

	Bağlantı	Nöron	Katman	Ağ
 1,000 1,000 1,000 1,000 1,000				
İşlem		1K XNOR 1K Bit-Adds	2 Million Bit-Ops	8 Million Bit-Ops
Hafıza	1 bit	1K bit	0.125MB	0.5MB

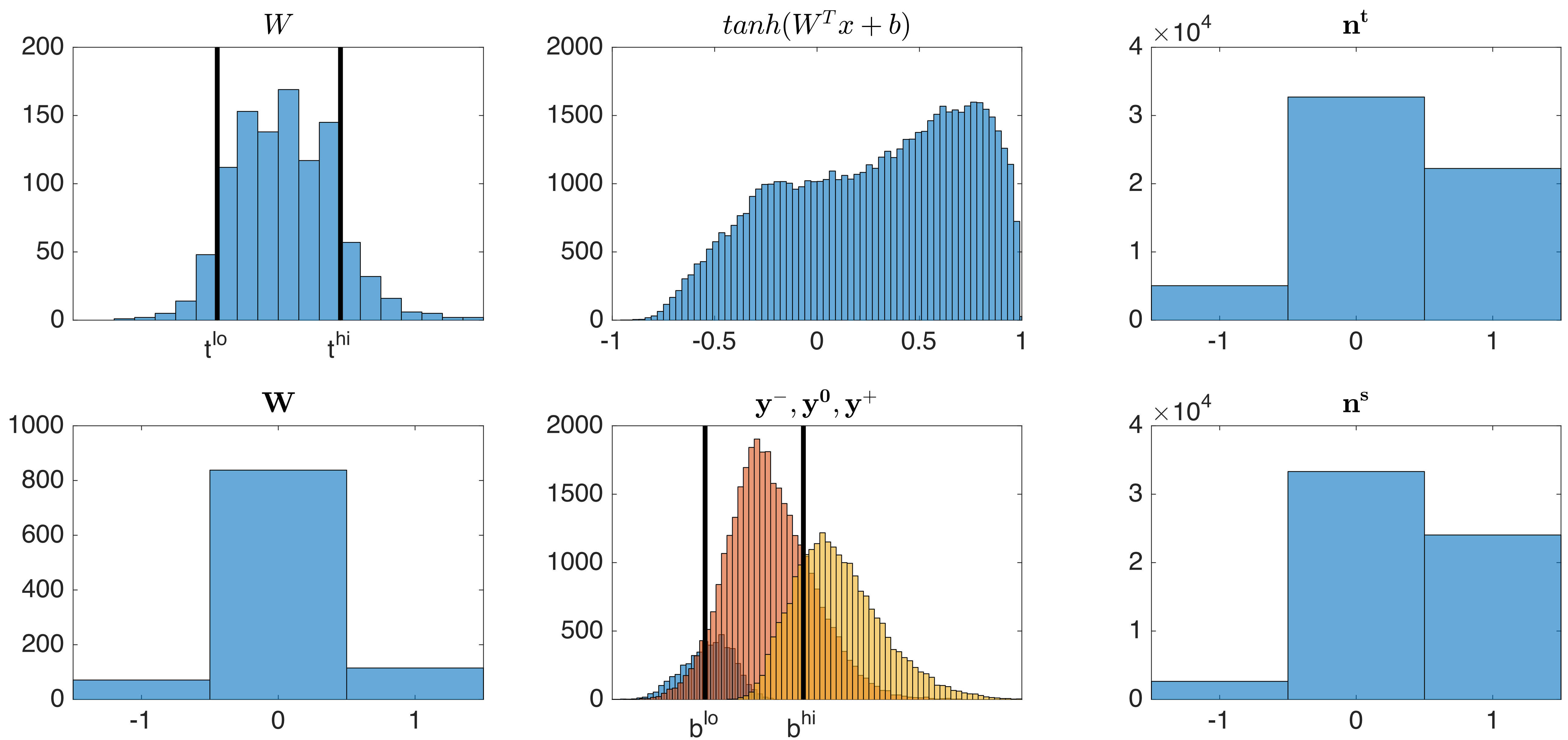
ÜÇLÜ YSA

	Öğretmen	Öğrenci
Parametreler	$W_i = [w_j], w_j \in \mathbb{R}$	$\mathbf{W_i} = [\mathbf{w_j}], \mathbf{w_j} \in \{-1, 0, 1\}$
	$b_i \in \mathbb{R}$	$\mathbf{b_i}^{lo} \in \mathbb{Z}$ $\mathbf{b_i}^{hi} \in \mathbb{Z}$
Transfer Fonksiyonu	$y_i = W_i^\intercal \mathbf{x} + b_i$	$\mathbf{y_i} = \mathbf{W_i}^\intercal \mathbf{x}$
Etkinleştirme Fonksiyonu	$\mathbf{n_i^t} = \begin{cases} -1 & \text{with prob. } -\rho \text{ if } \rho < 0 \\ 1 & \text{with prob. } \rho \text{ if } \rho > 0 \\ 0 & \text{otherwise} \end{cases}$ where $\rho = tanh(y_i), \rho \in (-1, 1)$	$\mathbf{n_i^s} = \begin{cases} -1 & \text{if } \mathbf{y_i} < \mathbf{b_i}^{lo} \\ 1 & \text{if } \mathbf{y_i} > \mathbf{b_i}^{hi} \\ 0 & \text{otherwise} \end{cases}$





ÜÇLÜ YSA



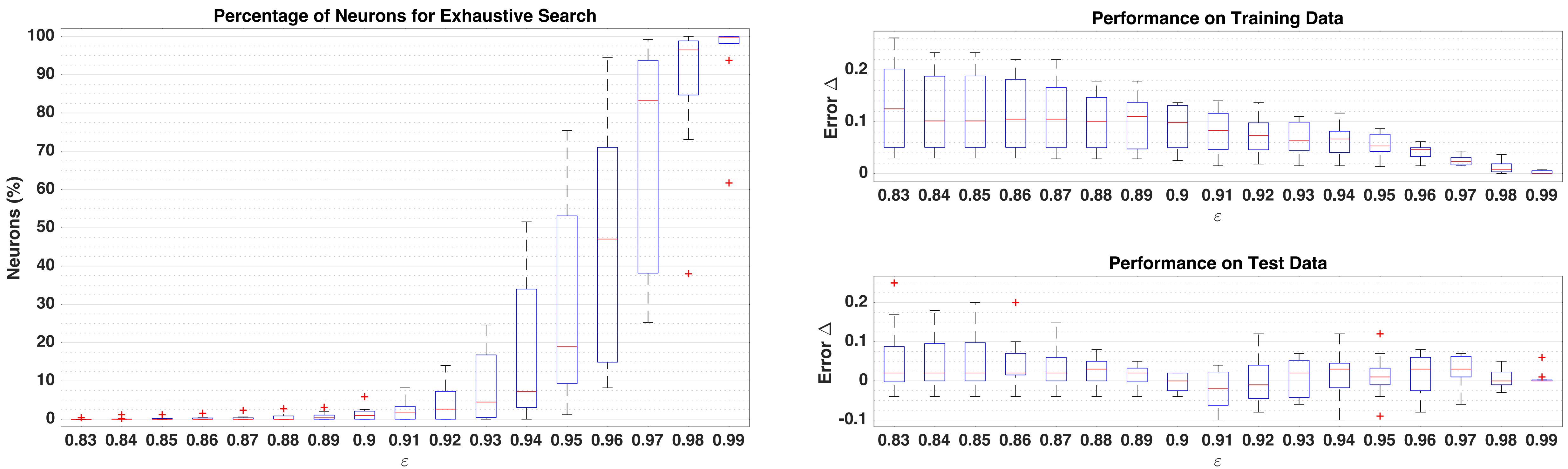
W
200
150
100
50
0
t^{lo}
t^{hi}
$tanh(W^T x + b)$
2000
1500
1000
500
0
-1
-0.5
0
0.5
1
$\mathbf{n^t}$
$\times 10^4$
4
3
2
1
0
-1
0
1
$\mathbf{W}$
1000
800
600
400
200
0
-1
0
1
$\mathbf{y^-}, \mathbf{y^0}, \mathbf{y^+}$
2000
1500
1000
500
0
b^{lo}
b^{hi}
$\mathbf{n^s}$
$\times 10^4$
4
3
2
1
0
-1
0
1

BAŞARIM

MNIST

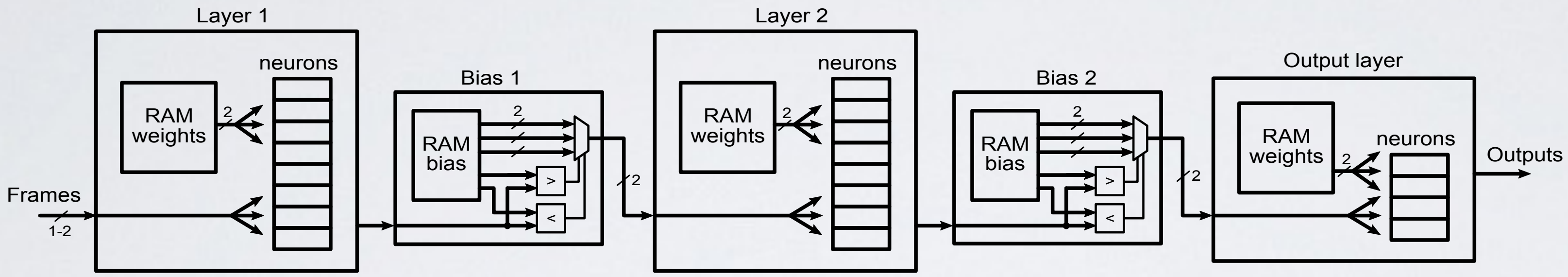
60K rakam imgesi

BAŞARIM



Percentage of Neurons for Exhaustive Search
Neurons (%)
100
90
80
70
60
50
40
30
20
10
0
0.83 0.84 0.85 0.86 0.87 0.88 0.89 0.9 0.91 0.92 0.93 0.94 0.95 0.96 0.97 0.98 0.99
ε
Performance on Training Data
Error Δ
0.2
0.1
0
0.83 0.84 0.85 0.86 0.87 0.88 0.89 0.9 0.91 0.92 0.93 0.94 0.95 0.96 0.97 0.98 0.99
ε
Performance on Test Data
Error Δ
0.2
0.1
0
-0.1
0.83 0.84 0.85 0.86 0.87 0.88 0.89 0.9 0.91 0.92 0.93 0.94 0.95 0.96 0.97 0.98 0.99
ε

BAŞARIM



nöron	hız(imge/s)	katman	enerji(uJ)	gecikme(us)	başarım (%)
250	255102	1	1.24	5.37	97.76
		2	2.44	6.73	98.13
		3	3.63	8.09	98.14
500	255102	1	2.44	6.63	97.75
		2	4.83	9.24	98.14
		3	7.22	11.9	98.29
750	255102	1	3.63	7.88	97.73
		2	7.22	11.8	98.10
		3	10.8	15.6	98.33
1000	198019	1	6.22	10.2	97.63
		2	12.4	15.3	98.09
		3	18.5	20.5	97.89

IBM TrueNorth
%95 başarım
4 uJ
1000 imge/s
1 ms gecikme

DAHA FAZLASI İÇİN

- http://deeplearning.net/tutorial/
- http://deeplearning.stanford.edu/tutorial/
- http://torch.ch
- http://colah.github.io
- Coursera'da Geoffrey Hinton dersi
- Oxford ML dersi

Teşekkürler

hande.alemdar@imag.fr