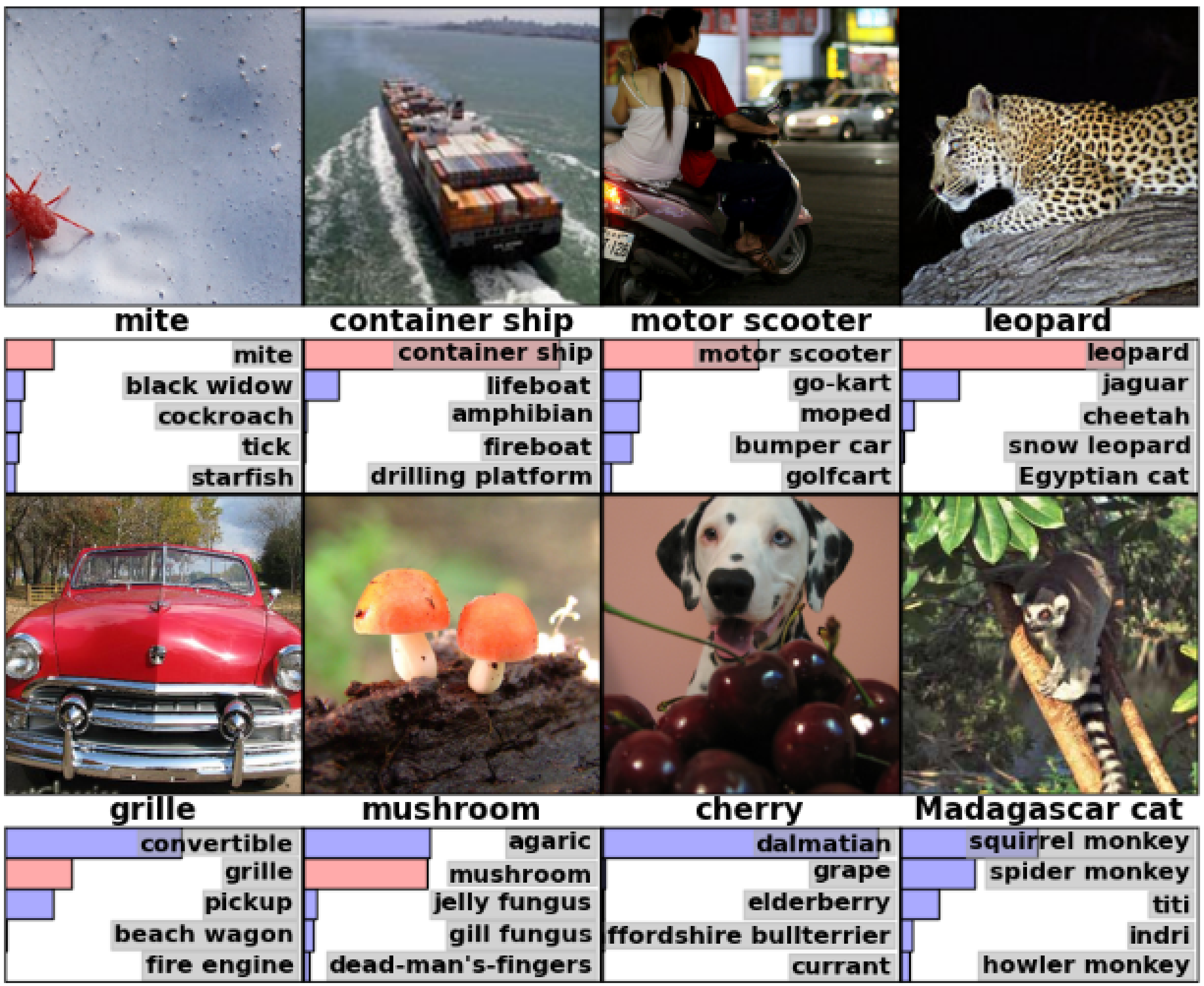
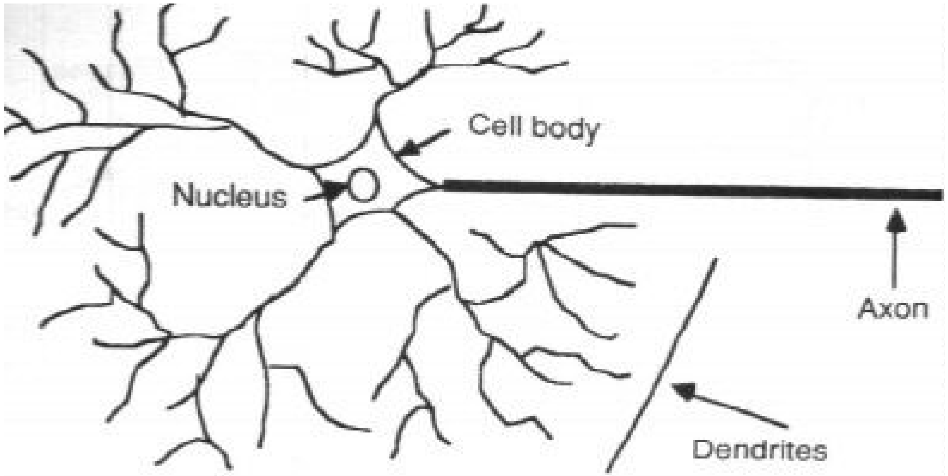
Neural Networks

AlexNet 2012



The best Machine

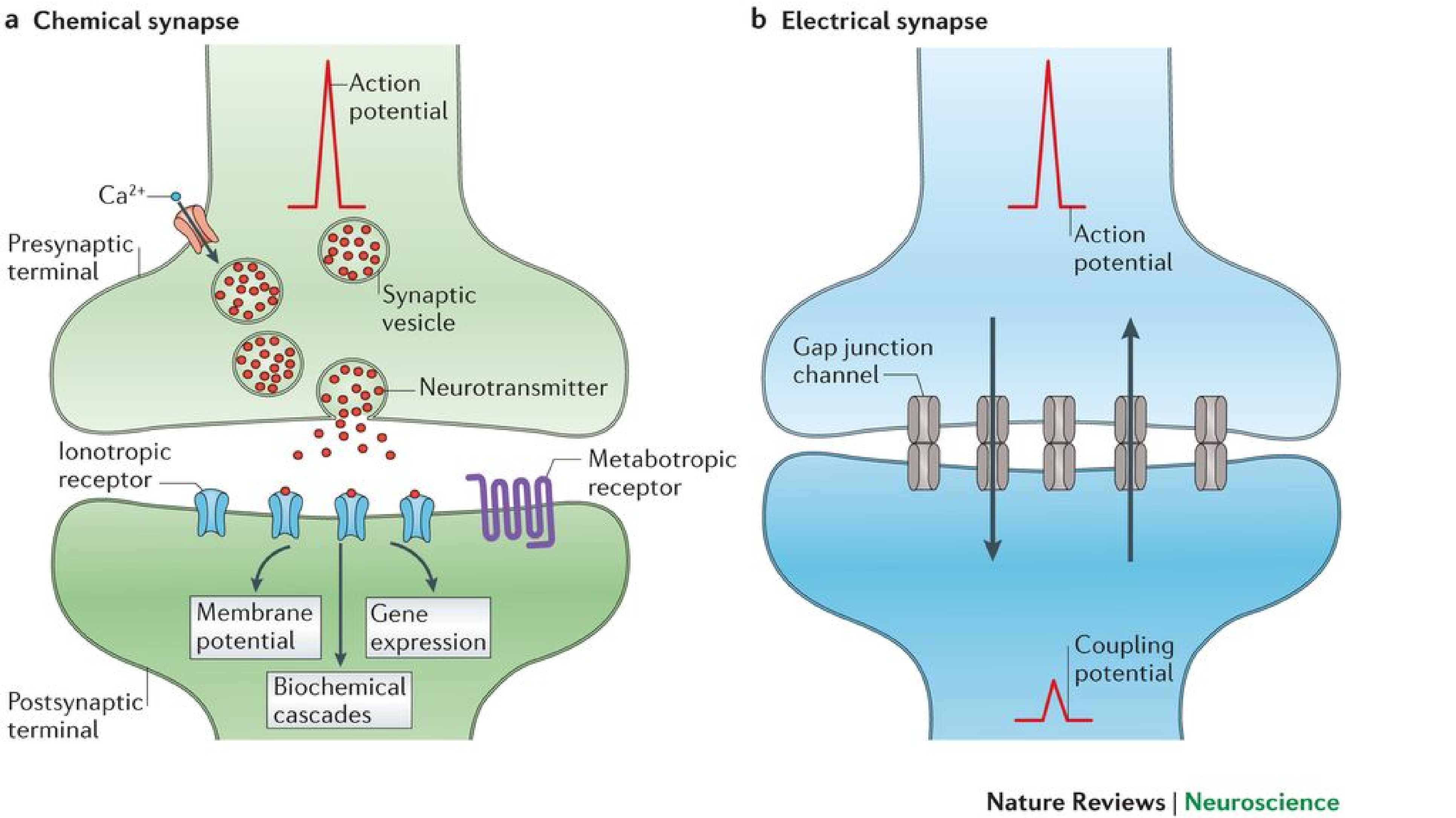
Inside of your brain...



100 billion neurons

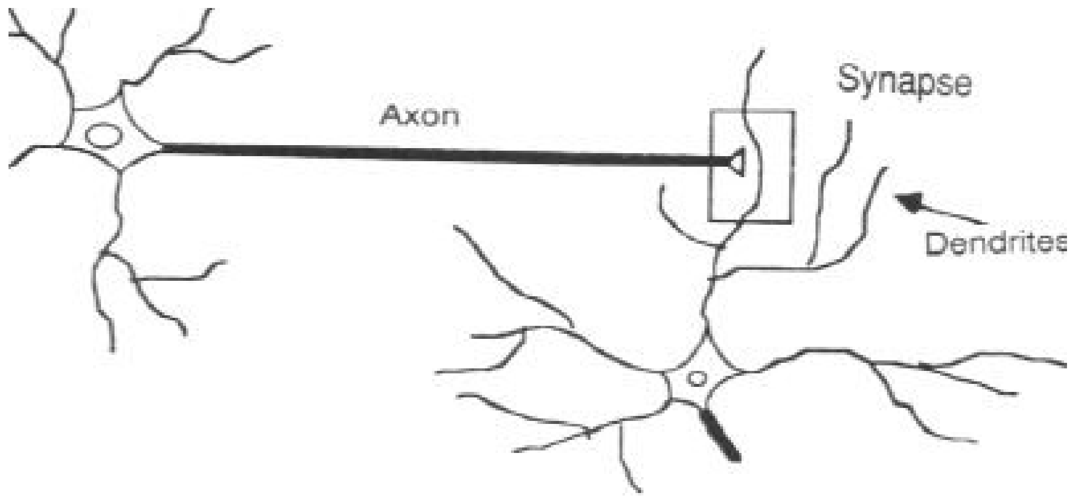
100 trillion synapses

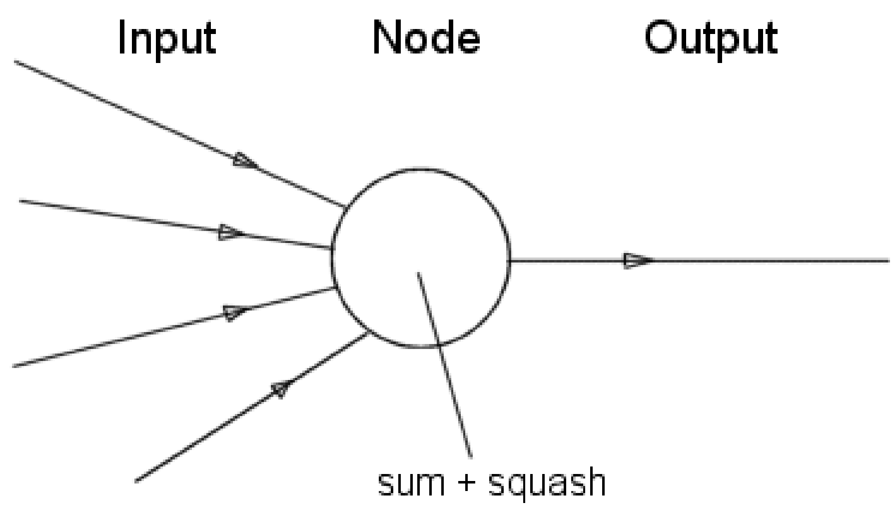
Synapses



a Chemical synapse
Action potential
Ca2+
Presynaptic terminal
Synaptic vesicle
Neurotransmitter
Ionotropic receptor
Metabotropic receptor
Membrane potential
Gene expression
Biochemical cascades
Postsynaptic terminal
b Electrical synapse
Action potential
Gap junction channel
Coupling potential
Nature Reviews | Neuroscience

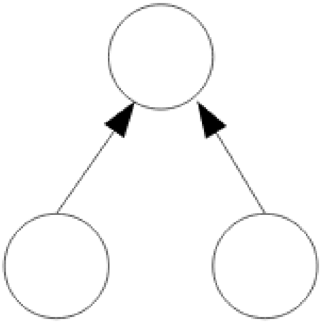
- Synapse vs. weight

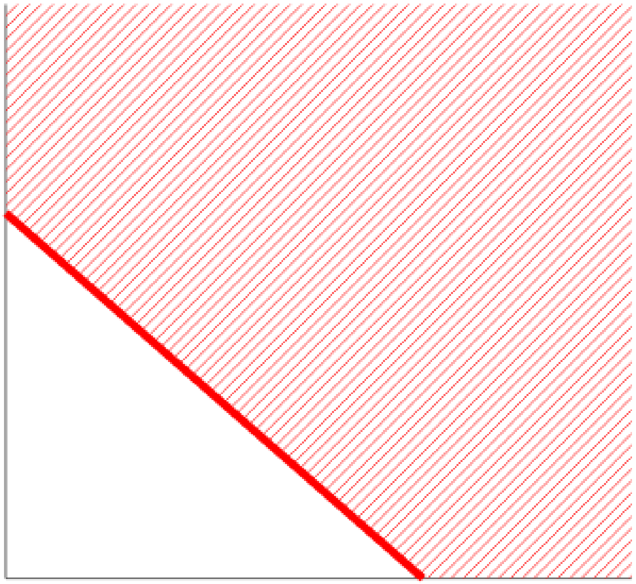




Adding Layer’s

1 layer of trainable weights

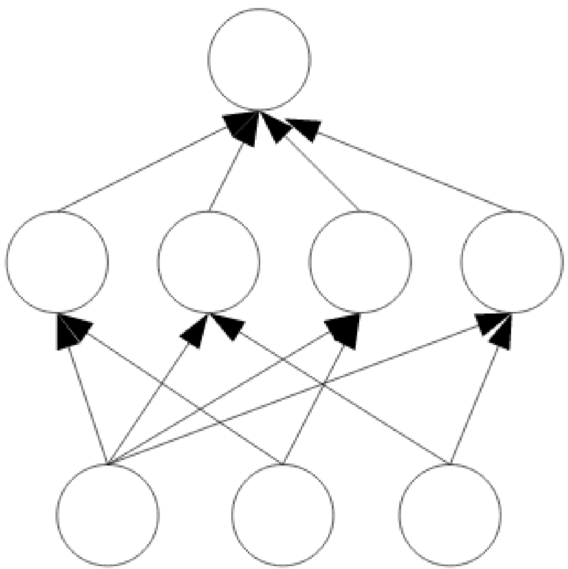


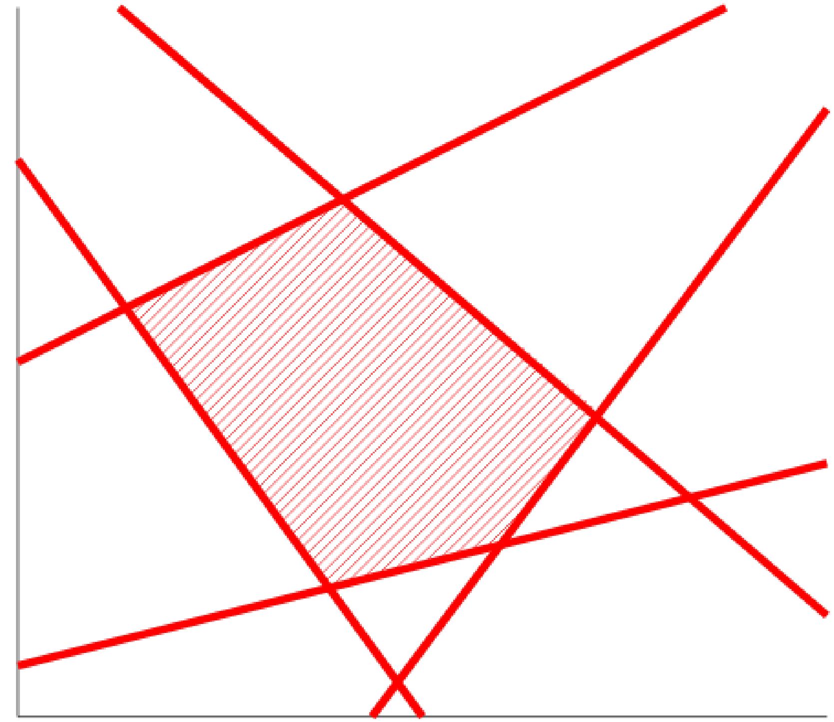


separating hyperplane

Adding Layer's

2 layers of trainable weights

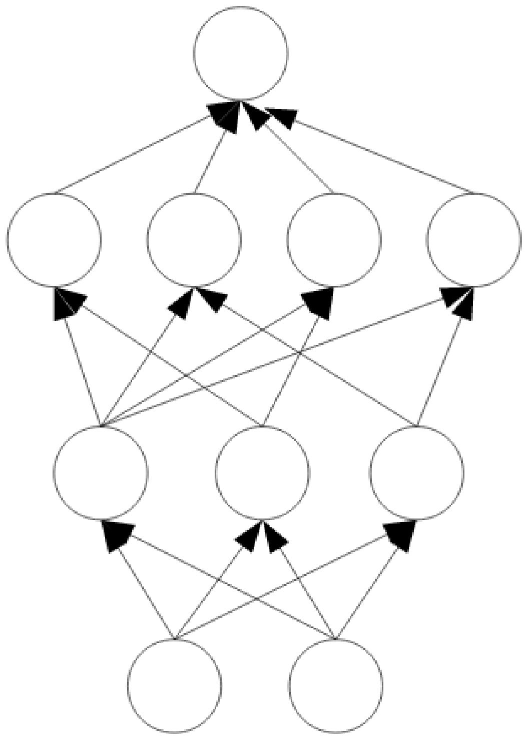


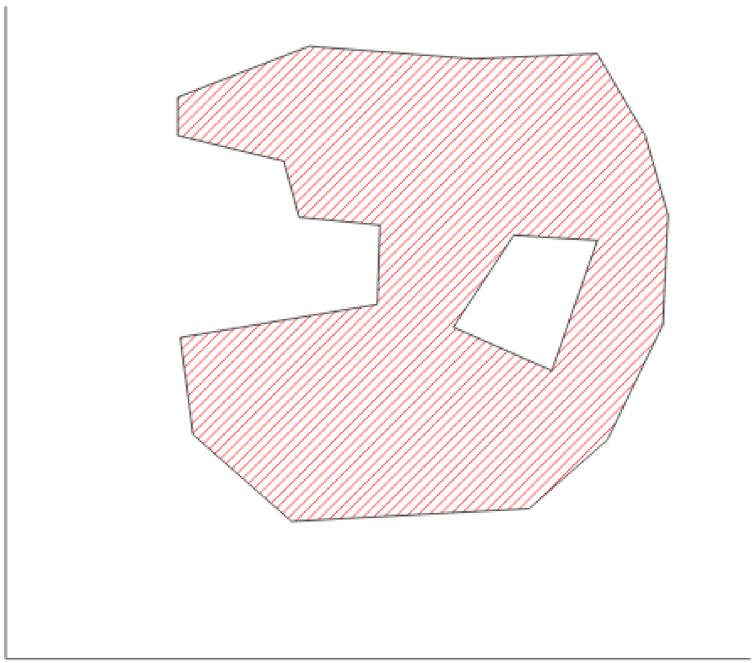


convex polygon region

Adding Layer’s

3 layers of trainable weights

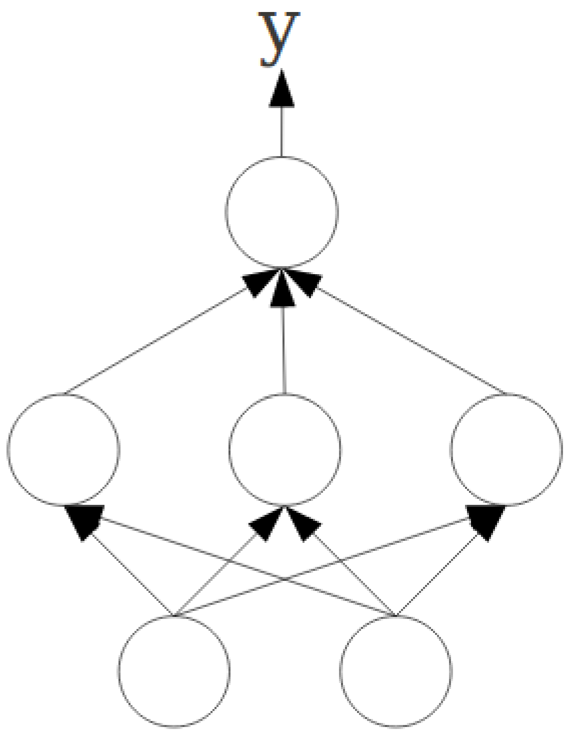




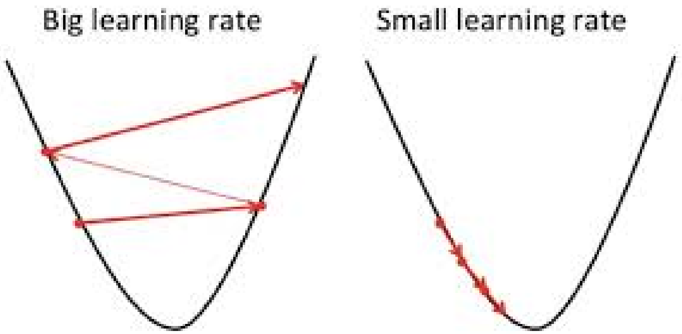
composition of polygons: convex regions

Cost Function

$$E = \frac{1}{2}\sum_{p} (d^p - y^p)^2$$



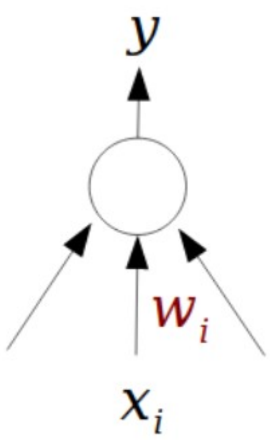
Gradient Descendent



$$y = \sum_i w_i x_i$$

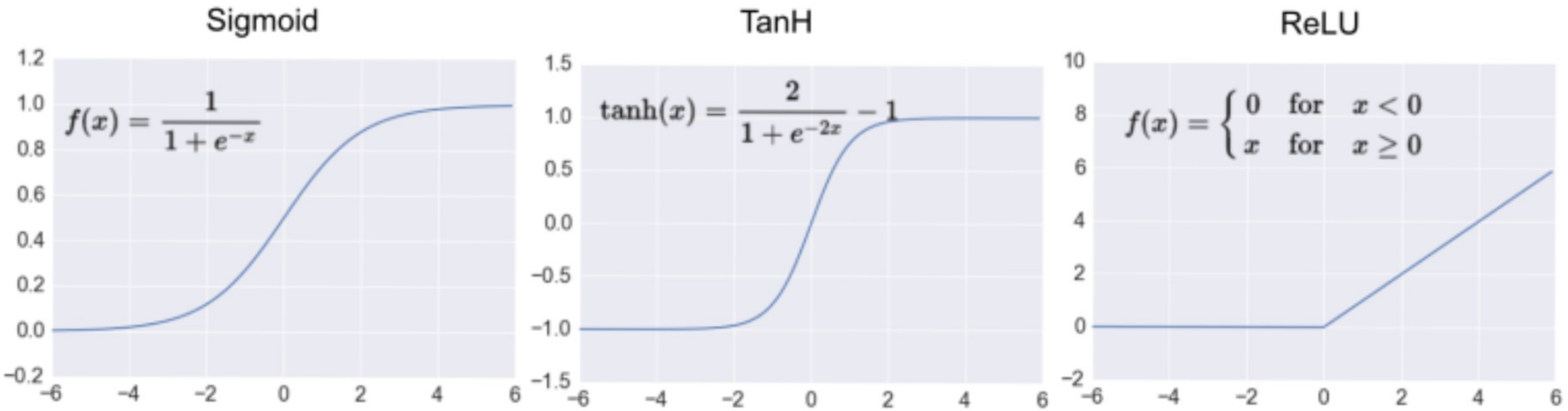
$$E = \frac{1}{2}\sum_p (d^p - y^p)^2$$

$$\frac{dE}{dy} = y - d$$



$$\frac{\partial E}{\partial w_i} = \frac{dE}{dy} \cdot \frac{\partial y}{\partial w_i} = (y-d)x_i$$

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} = -\eta (y-d)x_i$$


Sigmoid
$f(x) = \frac{1}{1+e^{-x}}$
TanH
$\tanh(x) = \frac{2}{1+e^{-2x}} - 1$
ReLU
$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$

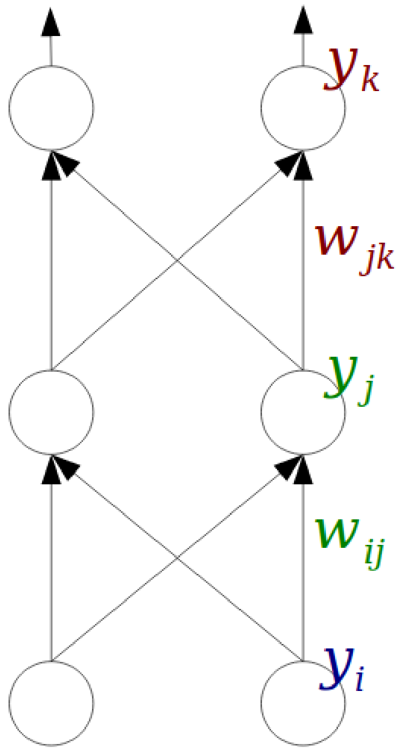
Non-linear activation functions

$$y = g(net) = \tanh\left(\sum_i w_i x_i\right)$$

$$\frac{dE}{dy} = (y - d), \qquad \frac{dy}{d\,net} = 1/\cosh^2(net), \qquad \frac{\partial net}{\partial w_i} = x_i$$

$$\begin{aligned} \frac{\partial E}{\partial w_i} &= \frac{dE}{dy} \cdot \frac{dy}{d\,net} \cdot \frac{\partial net}{\partial w_i} \\ &= (y - d)/\cosh^2\left(\sum_i w_i x_i\right) \cdot x_i \end{aligned}$$

Apply the chain Rule...



$$\frac{\partial E}{\partial y_k} = (y_k - d_k)$$

$$\delta_k = \frac{\partial E}{\partial net_k} = (y_k - d_k) \cdot g'(net_k)$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial w_{jk}} = \frac{\partial E}{\partial net_k} \cdot y_j$$

$$\frac{\partial E}{\partial y_j} = \sum_k \left(\frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial y_j} \right)$$

$$\delta_j = \frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial y_j} \cdot g'(net_j)$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \cdot y_i$$

Weight's Update

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial w_{jk}} = \delta_k \cdot y_j$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ij}} = \delta_j \cdot y_i$$

$$\Delta w_{jk} = -\eta \cdot \frac{\partial E}{\partial w_{jk}} \qquad \Delta w_{ij} = -\eta \cdot \frac{\partial E}{\partial w_{ij}}$$

Recurrent Neural Network’s

Exploding and vanishing gradients...

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{10} = \begin{pmatrix} 2^{10} & 0^{10} \\ 0^{10} & 2^{10} \end{pmatrix} = \begin{pmatrix} 1024 & 0 \\ 0 & 1024 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{20} = \begin{pmatrix} 2^{20} & 0^{20} \\ 0^{20} & 2^{20} \end{pmatrix} = \begin{pmatrix} 1048576 & 0 \\ 0 & 1048576 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{50} = \begin{pmatrix} 2^{50} & 0^{50} \\ 0^{50} & 2^{50} \end{pmatrix} = \begin{pmatrix} 1.12589907_{E}^{15} & 0 \\ 0 & 1.12589907_{E}^{15} \end{pmatrix}$$

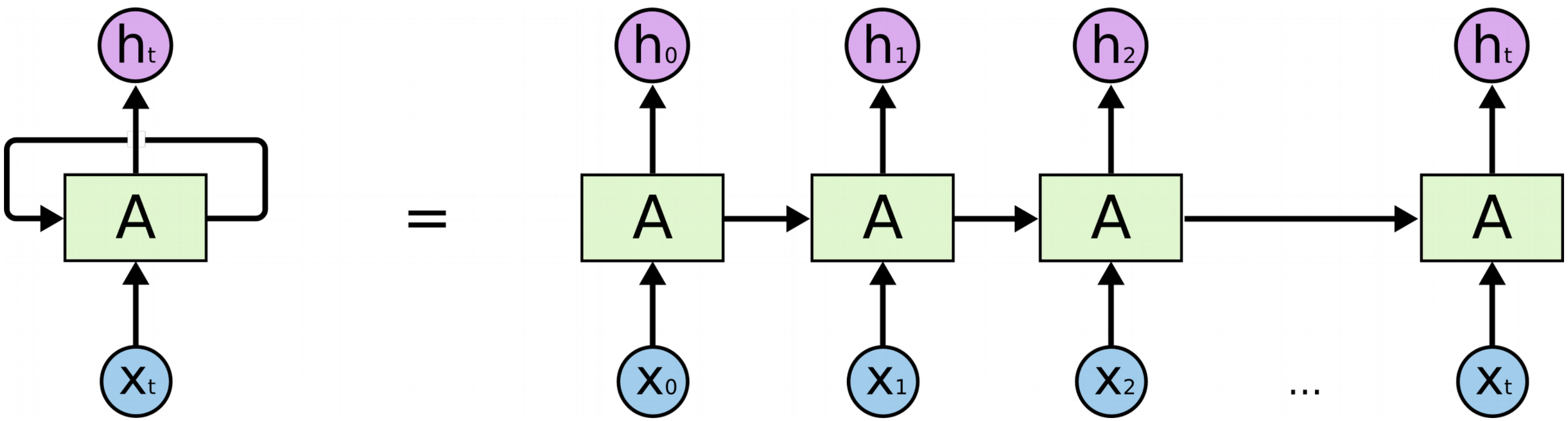
Solution

We clip the Gradient

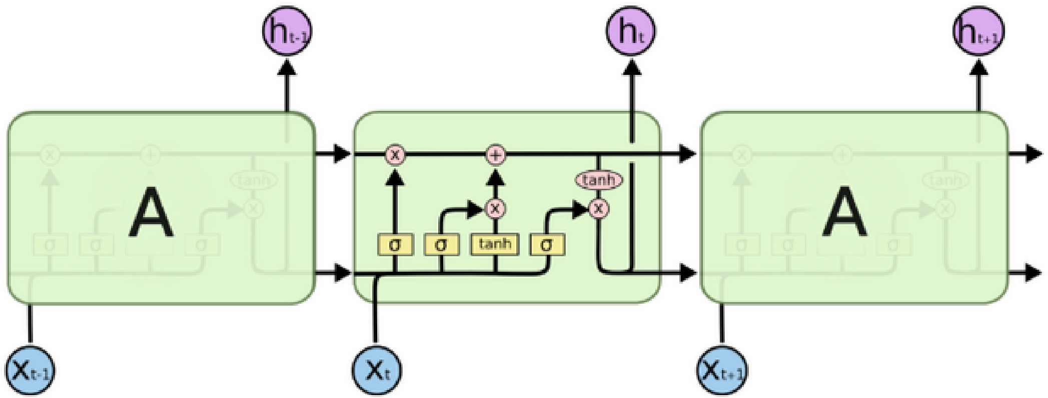
Algorithm 1 Pseudo-code for norm clipping

$\hat{\mathbf{g}} \leftarrow \frac{\partial \mathcal{E}}{\partial \theta}$
if $\|\hat{\mathbf{g}}\| \geq threshold$ **then**
  $\hat{\mathbf{g}} \leftarrow \frac{threshold}{\|\hat{\mathbf{g}}\|}\hat{\mathbf{g}}$
end if

LSTM



What are the main differences?

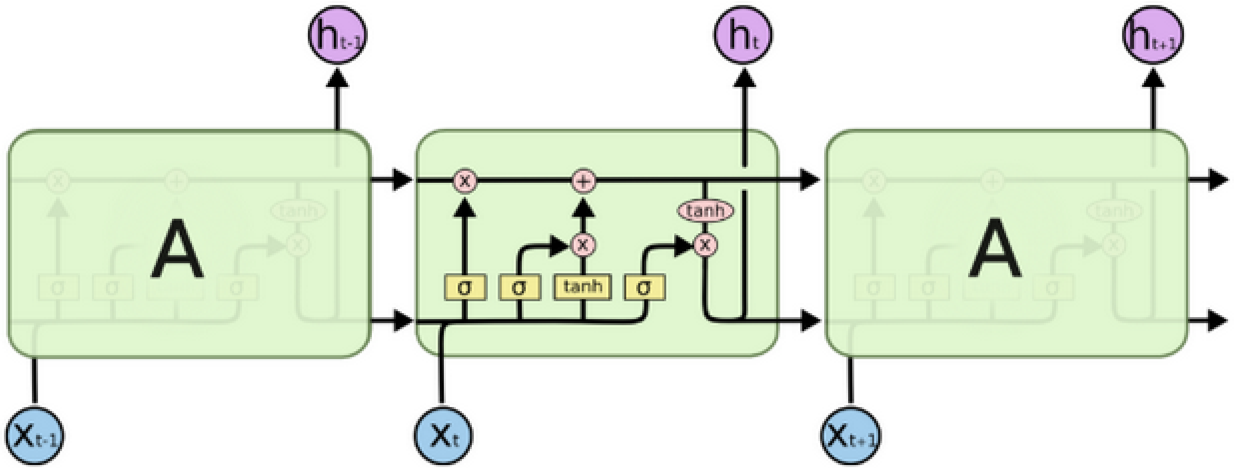


$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot g$$

$$h_t = o \odot \tanh(c_t)$$

How the Gradient flow's



$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot g$$

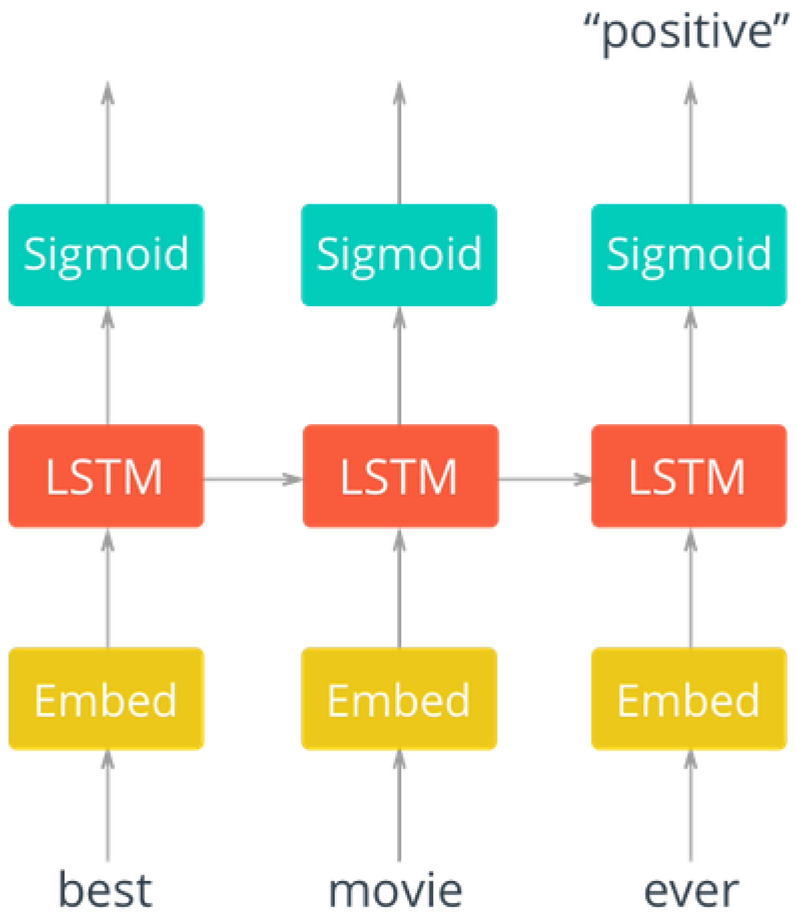
$$h_t = o \odot \tanh(c_t)$$

Backpropagation from c_t to c_{t-1} only requires an elementwise multiplication by f, no matrix multiplication.

The output of a forget gate (f) is a number between zero and one.

But have we actually solved the problem?

Sentiment analysis using RNNs(LSTM)



Applications?

Negative review

Dear #XYZ there is no network in my area and internet service is pathetic from the past one week. Kindly help me out

-Dated: 10/09/17

Mixed review

Although the value added services being provided are great but the prices are high
#VAS #XYZ

-Dated: 10/09/17

Positive review

Great work done #XYZ Problem resolved by customer care in just one day
#ThanksXYZ

-Dated: 5/06/17

Image Captioning

Can a machine generate a text description of a image?

Objects?

Actions?

Properties?

Objects interactions?

Image Captioning

> *A man and a girl sit on the ground and eat .*
> *A man and a little girl are sitting on a sidewalk near a blue bag eating .*
> *A man wearing a black shirt and a little girl wearing an orange dress share a treat .*

And this one?

Objects?

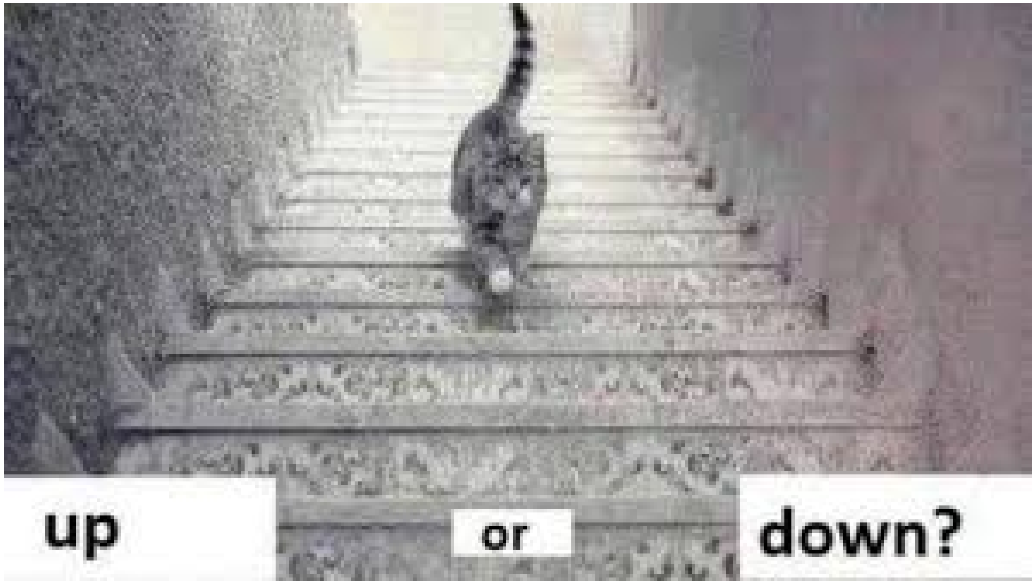
Properties?

Actions?

Objects interactions?

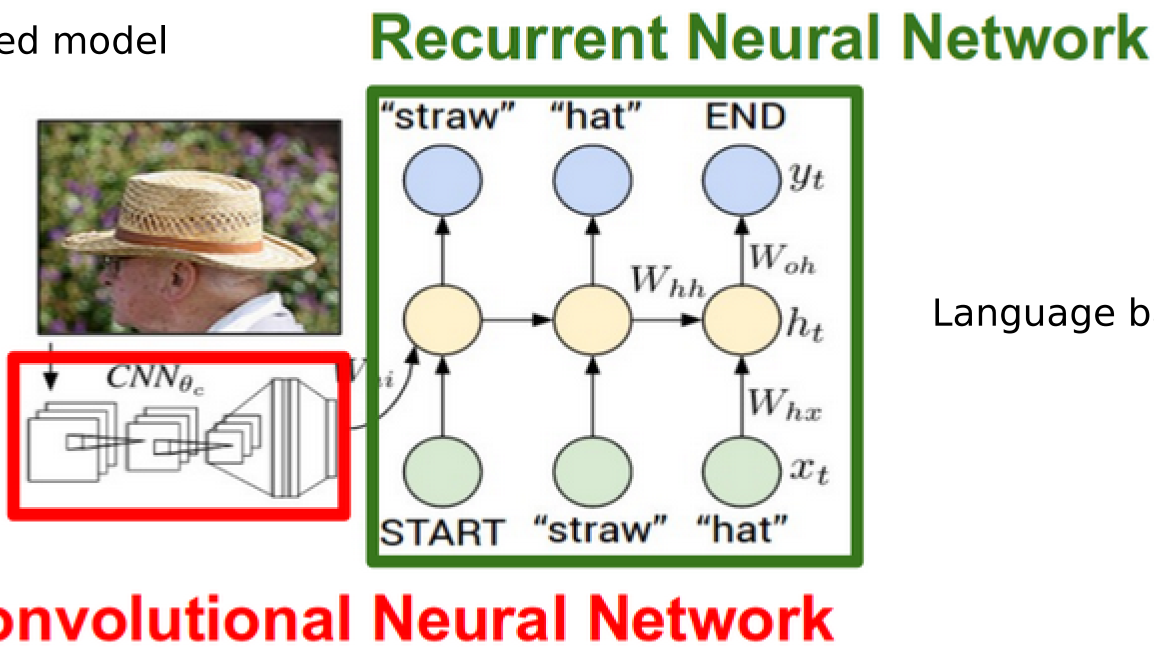
What about the cat ?

Guess this one…



CNN+LSTM

Image based model



Language based model

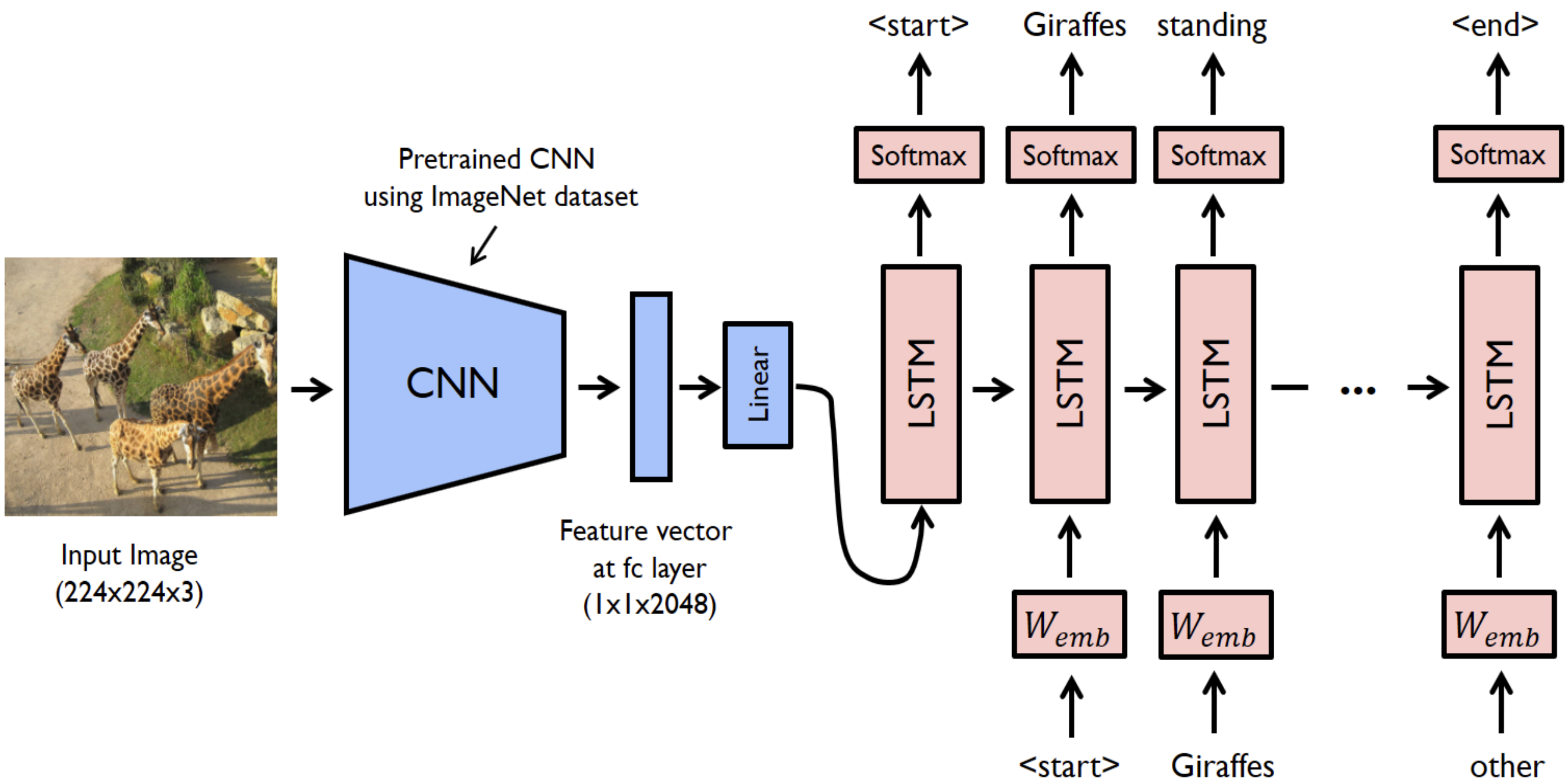
Encoding

Decoding

Training

Label – [<start>, A, man, and, a, girl, sit, on, the, ground, and, eat, .]
Target – [A, man, and, a, girl, sit, on, the, ground, and, eat, ., <end>]

Testing



<start> a group of giraffes standing in a grassy area . <end>

Some Positive Examples

A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A group of young people playing a game of frisbee.

Two hockey players are fighting over the puck.

Some Negative Examples

A man is taking a picture of a giraffe .

A group of brown cows are standing in the grass

Video Classification

https://www.youtube.com/watch?v=9qyD7vjVfLI

The next great challenge?

Thank you !

"It's a self-driving car."