Non-linear Classification

Lecture 10 - DAMLF | ML1

Neural Image Captioning

Show and Tell: A Neural Image Caption Generator

Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan

(Submitted on 17 Nov 2014 (v1), last revised 20 Apr 2015 (this version, v2))

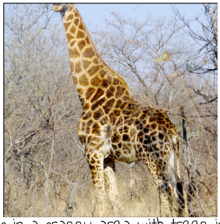
Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. In this paper, we present a generative model based on a deep recurrent architecture that combines recent advances in computer vision and machine translation and that can be used to generate natural sentences describing an image. The model is trained to maximize the likelihood of the target description sentence given the training image. Experiments on several datasets show the accuracy of the model and the fluency of the language it learns solely from image descriptions. Our model is often quite accurate, which we verify both qualitatively and quantitatively. For instance, while the current state-of-the-art BLEU-1 score (the higher the better) on the Pascal dataset is 25, our approach yields 59, to be compared to human performance around 69. We also show BLEU-1 score improvements on Flickr30k, from 56 to 66, and on SBU, from 19 to 28. Lastly, on the newly released COCO dataset, we achieve a BLEU-4 of 27.7, which is the current state-of-the-art.

Subjects: **Computer Vision and Pattern Recognition (cs.CV)**
Cite as: **arXiv:1411.4555 [cs.CV]**
(or **arXiv:1411.4555v2 [cs.CV]** for this version)

'a close up of a cake on a plate'

Source: Indico Data: http://indico.io

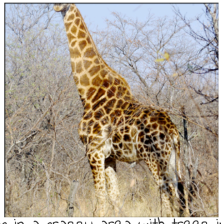
'a close up of a cake on a plate'



'a giraffe standing in a grassy area with trees in the background'

Source: Indico Data: http://indico.io

'a close up of a cake on a plate'

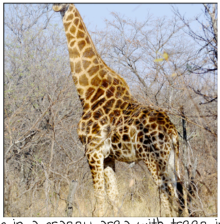


'a giraffe standing in a grassy area with trees in the background'

'a man riding a wave on top of a surfboard'

Source: Indico Data: http://indico.io

'a close up of a cake on a plate'



'a giraffe standing in a grassy area with trees in the background'

'a man riding a wave on top of a surfboard'

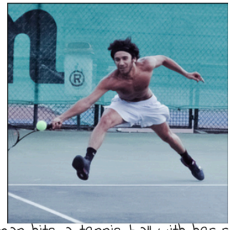
'the round clock is attached to the wing'

Source: Indico Data: http://indico.io

'a car with a lot of luggage on the back of it'

Source: Indico Data: http://indico.io

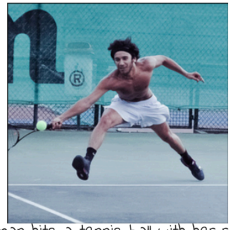
'a car with a lot of luggage on the back of it'



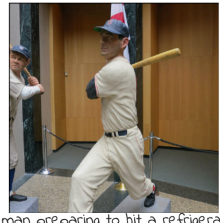
'a woman hits a tennis ball with her racquet'

Source: Indico Data: http://indico.io

'a car with a lot of luggage on the back of it'



'a woman hits a tennis ball with her racquet'

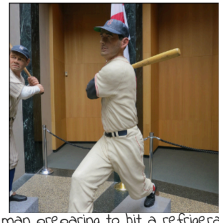


'a man preparing to hit a refrigerator'

Source: Indico Data: http://indico.io

'a car with a lot of luggage on the back of it'

'a woman hits a tennis ball with her racquet'

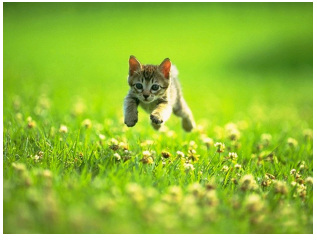


'a man preparing to hit a refrigerator'

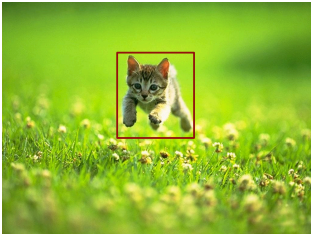
'a man standing next to a yellow motorcycle'

Source: Indico Data: http://indico.io

Classification, Localization, Detection

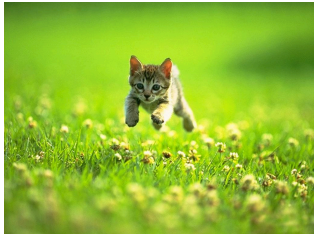


Cat Classification

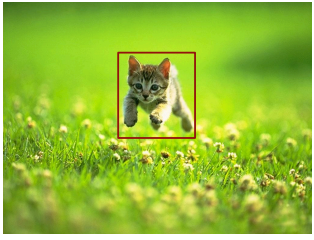


Classification & Localization

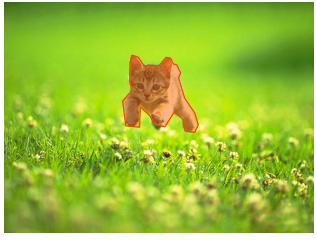
Classification, Localization, Detection



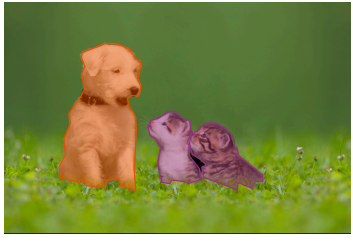
Cat Classification



Classification & Localization

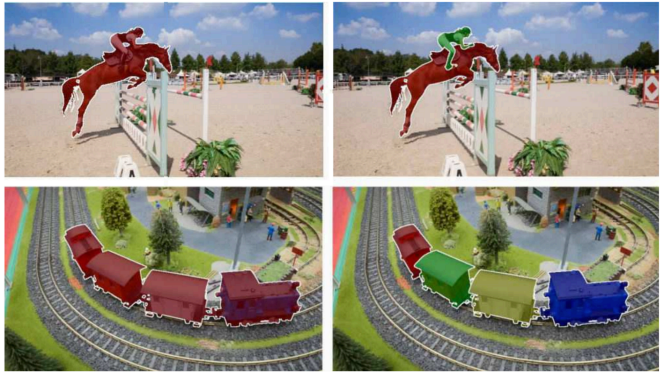


Object Detection



Instance Segmentation

Object Segmentation DAVIS Data Set



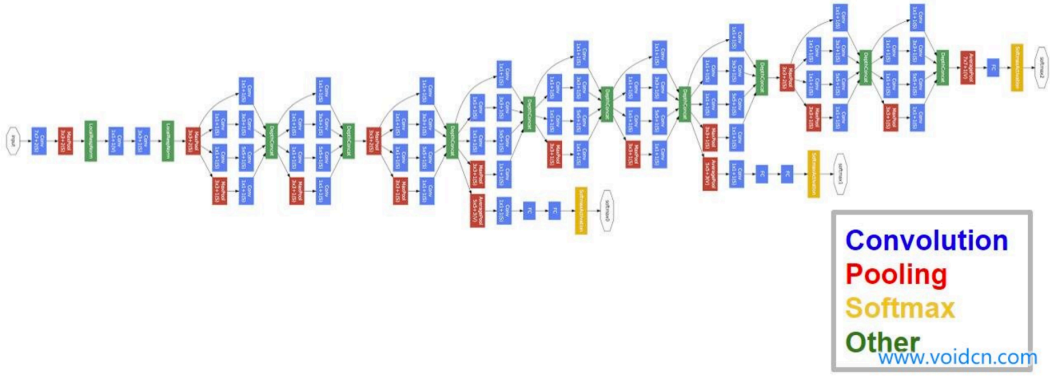
The main difference between annotations of DAVIS-2016 (left) and DAVIS-2017 (right): multi-instance segmentation

Eyes on the Road: How Autonomous Cars Understand What They're Seeing

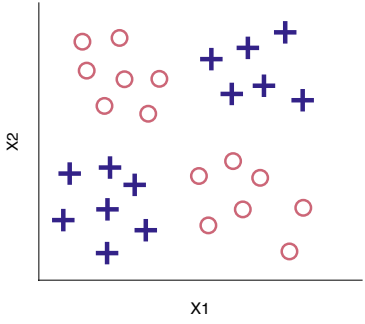
January 5, 2016 by DANNY SHAPIRO

Deep Neural Networks

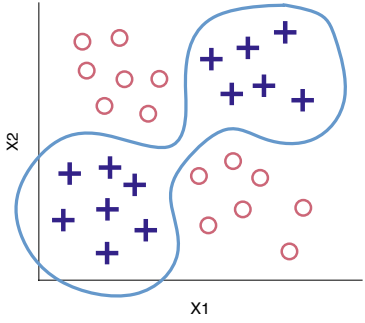
GoogLeNet Architecture



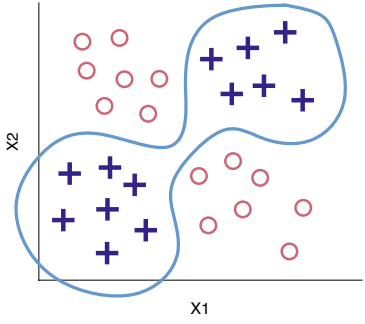
Simple Non-linear Classification

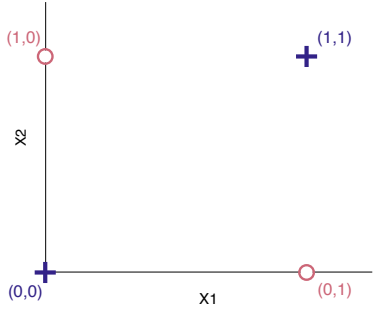


Simple Non-linear Classification



Simple Non-linear Classification





Simple Non-linear Classification

Exclusive or (XOR)

$y, x_1, x_2 \in \{0, 1\}$

$y = 1$ iff x_1 XOR x_2

iff $(x_1 = 1 \text{ and } x_2 = 0)$ OR $(x_1 = 0 \text{ and } x_2 = 1)$

(XNOR)

$y, x_1, x_2 \in \{0, 1\}$

$y = 1$ iff NOT $(x_1$ XOR $x_2)$

iff x_1 XNOR x_2

Simple Non-linear Classification

Exclusive or (XOR)

$y, x_1, x_2 \in \{0, 1\}$

$y = 1$ iff x_1 XOR x_2

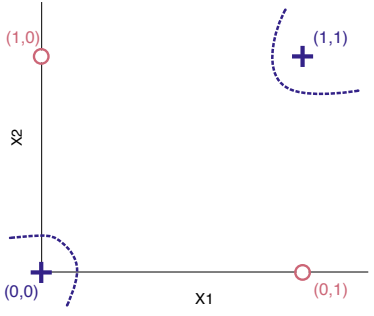
iff $(x_1 = 1 \text{ and } x_2 = 0)$ OR $(x_1 = 0 \text{ and } x_2 = 1)$

(XNOR)

$y, x_1, x_2 \in \{0, 1\}$

$y = 1$ iff NOT $(x_1$ XOR $x_2)$

iff x_1 XNOR x_2

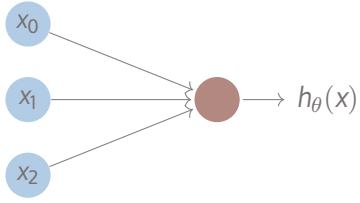


Logical Connectives with Neural Networks

Logical Connectives

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

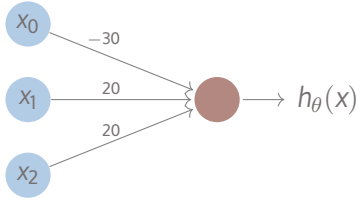
$y = 1$ iff $(x_1$ AND $x_2)$



Logical Connectives

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

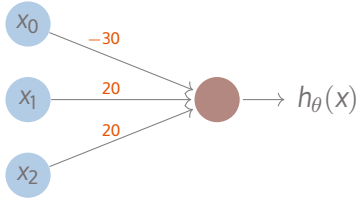
$y = 1$ iff $(x_1$ AND $x_2)$



Logical Connectives

$$x_1, x_2 \in \{0, 1\} \text{ and } x_0 = 1$$

$$y = 1 \text{ iff } (x_1 \text{ AND } x_2)$$

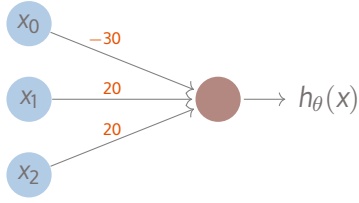


$$h_\Theta(x) = g\left(\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2\right)$$

Logical Connectives

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff $(x_1$ AND $x_2)$

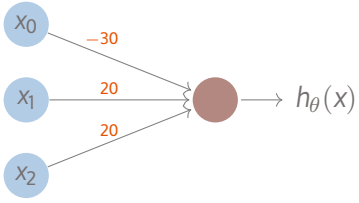


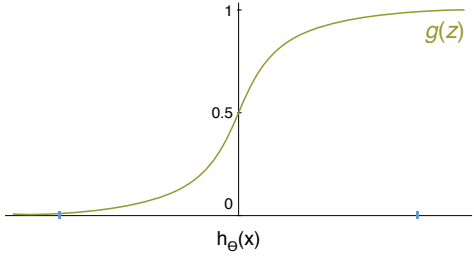
$$h_\Theta(x) = g(-30 + 20x_1 + 20x_2)$$

Logical Connectives: AND

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff (x_1 AND x_2)





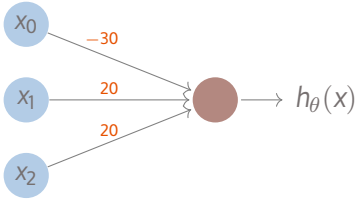
$h_\Theta(x) = g(-30 + 20 \cdot x_1 + 20 \cdot x_2)$

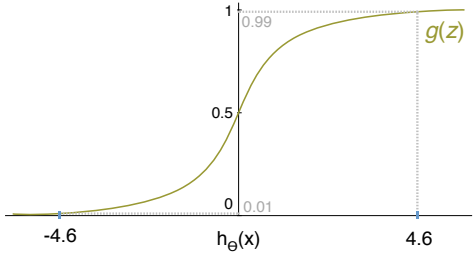
x_1	x_2	$h_\Theta(x)$	x_1 AND x_2
1	1		
1	0		
0	1		
0	0		

Logical Connectives: AND

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff (x_1 AND x_2)





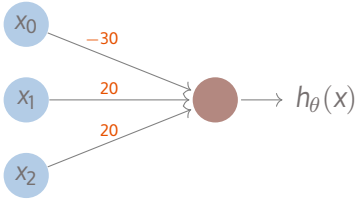
$h_\Theta(x) = g(-30 + 20 \cdot x_1 + 20 \cdot x_2)$

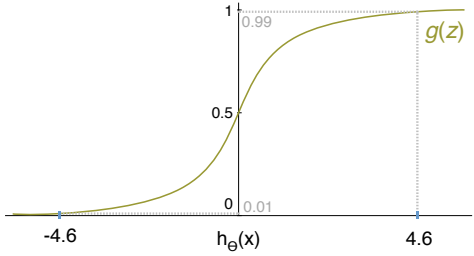
x_1	x_2	$h_\Theta(x)$	x_1 AND x_2
1	1		
1	0		
0	1		
0	0		

Logical Connectives: AND

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff (x_1 AND x_2)





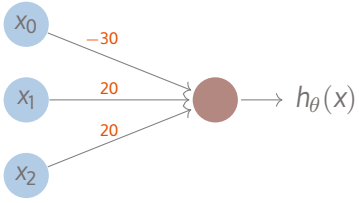
$h_\Theta(x) = g(-30 + 20 \cdot 1 + 20 \cdot 1)$

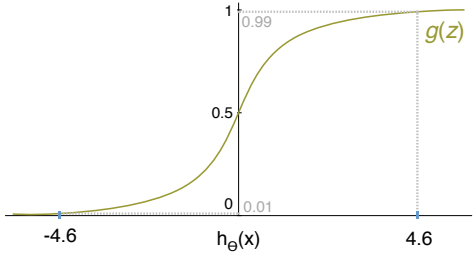
x_1	x_2	$h_\Theta(x)$	x_1 AND x_2
1	1	$g(10)$	≈ 1
1	0		
0	1		
0	0		

Logical Connectives: AND

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff (x_1 AND x_2)





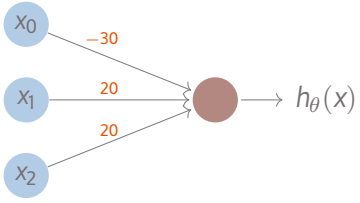
$h_\Theta(x) = g(-30 + 20 \cdot 1 + 20 \cdot 0)$

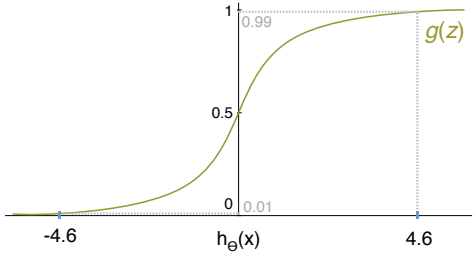
x_1	x_2	$h_\Theta(x)$	x_1 AND x_2
1	1	$g(10)$	≈ 1
1	0	$g(-10)$	≈ 0
0	1		
0	0		

Logical Connectives: AND

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff (x_1 AND x_2)





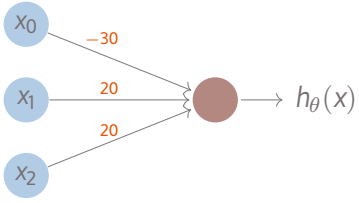
$h_\Theta(x) = g(-30 + 20 \cdot 0 + 20 \cdot 1)$

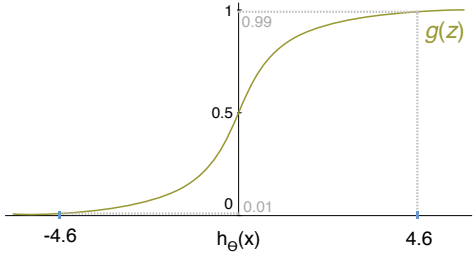
x_1	x_2	$h_\Theta(x)$	x_1 AND x_2
1	1	$g(10)$	≈ 1
1	0	$g(-10)$	≈ 0
0	1	$g(-10)$	≈ 0
0	0		

Logical Connectives: AND

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff (x_1 AND x_2)





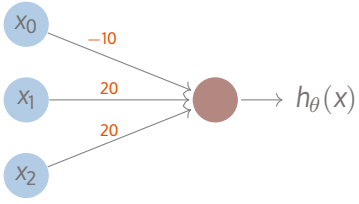
$h_\Theta(x) = g(-30 + 20 \cdot 0 + 20 \cdot 0)$

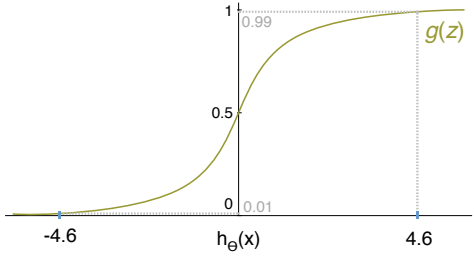
x_1	x_2	$h_\Theta(x)$	x_1 AND x_2
1	1	$g(10)$	≈ 1
1	0	$g(-10)$	≈ 0
0	1	$g(-10)$	≈ 0
0	0	$g(-30)$	≈ 0

Logical Connectives: OR

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff $(x_1$ OR $x_2)$





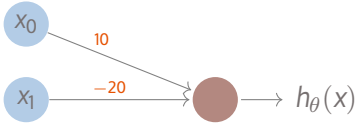
$h_\Theta(x) = g(-10 + 20x_1 + 20x_2)$

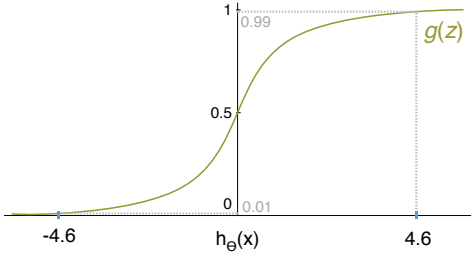
x_1	x_2	$h_\Theta(x)$	x_1 OR x_2
1	1	$g(30)$	≈ 1
1	0	$g(10)$	≈ 1
0	1	$g(10)$	≈ 1
0	0	$g(-10)$	≈ 0

Logical Connectives: NOT

$x_1, x_2 \in \{0, 1\}$ and $x_0 = 1$

$y = 1$ iff NOT x_1





$h_\Theta(x) = g(10 - 20x_1)$

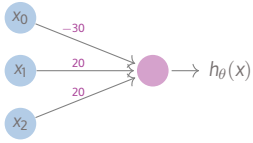
x_1	$h_\Theta(x)$	NOT x_1
1	$g(-10)$	≈ 0
0	$g(10)$	≈ 1

XNOR

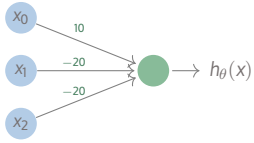
How do we represent XNOR?

XNOR

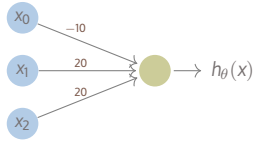
$(x_1 \text{ AND } x_2)$



(NOT x_1) AND (NOT x_2)

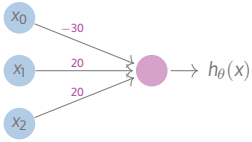


$(x_1 \text{ OR } x_2)$

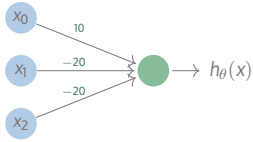


XNOR

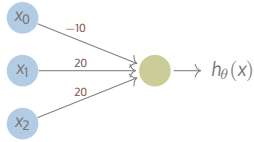
(x_1 AND x_2)

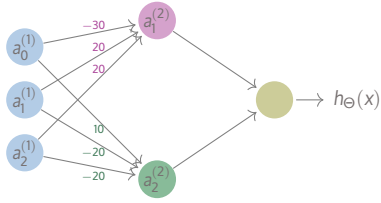


(NOT x_1) AND (NOT x_2)



(x_1 OR x_2)

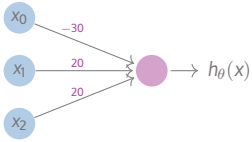




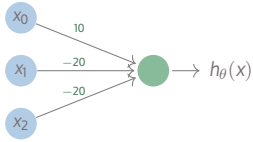
(x_1 XNOR x_2)

XNOR

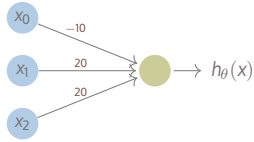
(x_1 AND x_2)

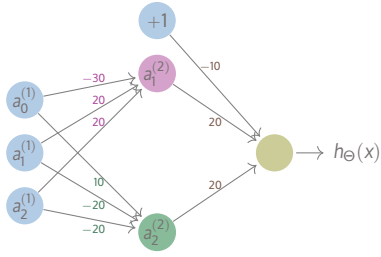


(NOT x_1) AND (NOT x_2)



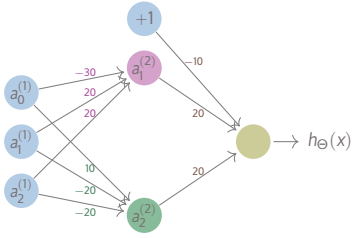
(x_1 OR x_2)





(x_1 XNOR x_2)

XNOR



(x_1 XNOR x_2)

x_1	x_2	$a_1^{(2)}$	$a_2^{(2)}$	$h_\Theta(x)$
1	1	1	0	1
1	0	0	0	0
0	1	0	0	0
0	0	0	1	1

References