

INTRODUCTION TO DEEP LEARNING

Seminar @ UPC TelecomBCN Barcelona (3rd edition). 22-28 January 2020.



Instructors



Xavier
Giro-i-Nieto



Marta R.
Costa-Jussà



Noé
Casas



Verónica
Vilaplana



Ramon
Morros



Javier
Ruiz

Organizers



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



Supporters



Google Cloud

GitHub Education

+ info: <http://bit.ly/idl2020>

<http://bit.ly/idl2020>

Day 1 Lecture 4

Softmax Regression



Xavier Giro-i-Nieto



@DocXavi



xavier.giro@upc.edu

Associate Professor
Universitat Politècnica de Catalunya



Acknowledgements



Santiago Pascual



Kevin McGuinness

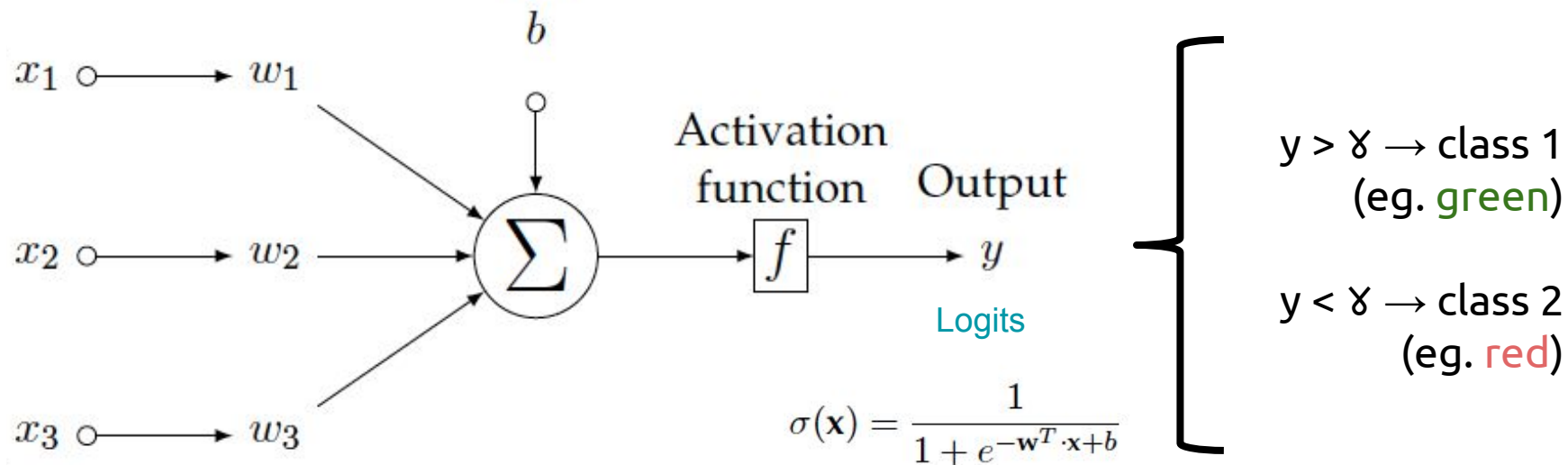
kevin.mcguinness@dcu.ie

Research Fellow

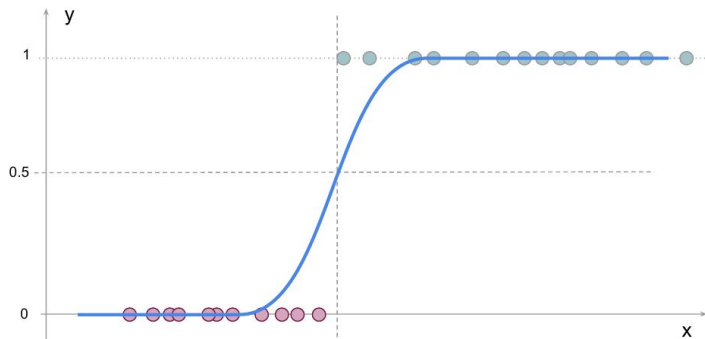
Insight Centre for Data Analytics
Dublin City University



Previously... Logistic Regression

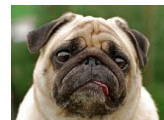
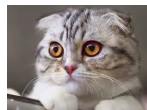


Threshold (γ)



Softmax regression: Multiclass (N classes)

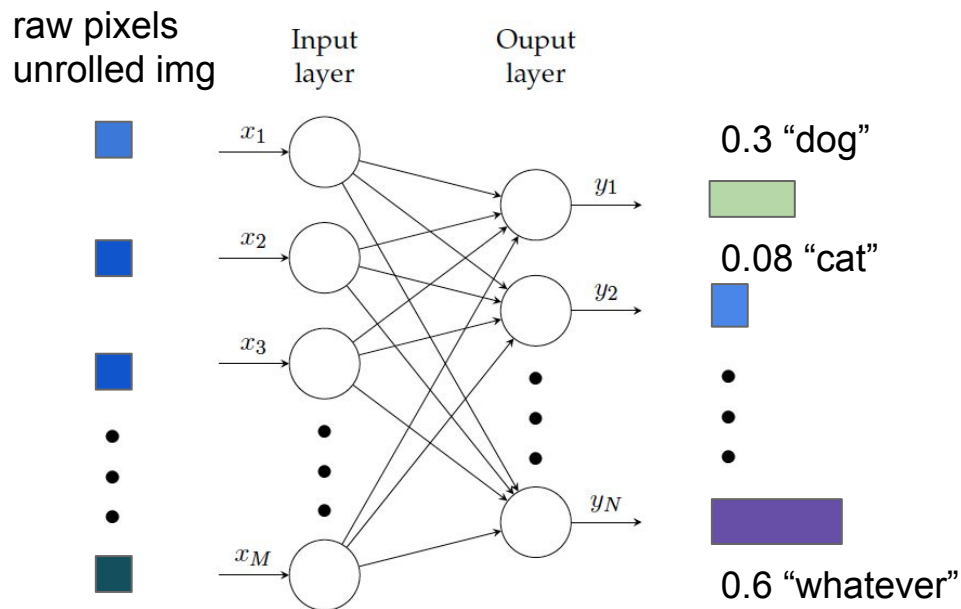
Question: How could the binary classifier with logistic regression to a problem with more than 2 classes (N) ?



Softmax regression: Multiclass (N classes)

Question: How could the binary classifier with logistic regression to a problem with more than 2 classes (N) ?

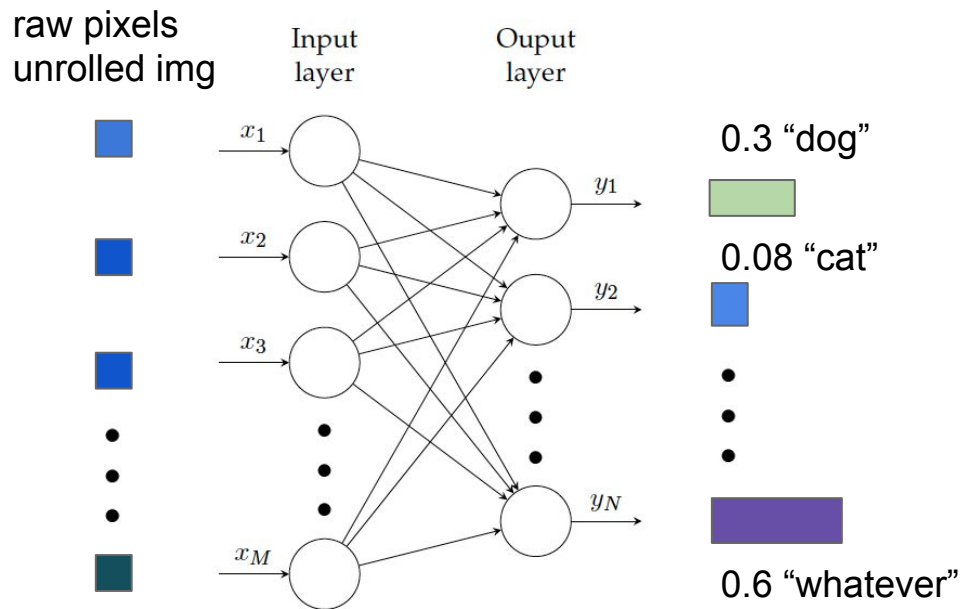
A multiclass classification problem can be solved by assigning a perceptron for each class and choosing the **maximum** logit...



Softmax regression: Multiclass (N classes)

Question: How could the binary classifier with logistic regression to a problem with more than 2 classes (N) ?

A multiclass classification problem can be solved by assigning a perceptron for each class and choosing the **maximum** logit...
but the max function is **non-differentiable**.



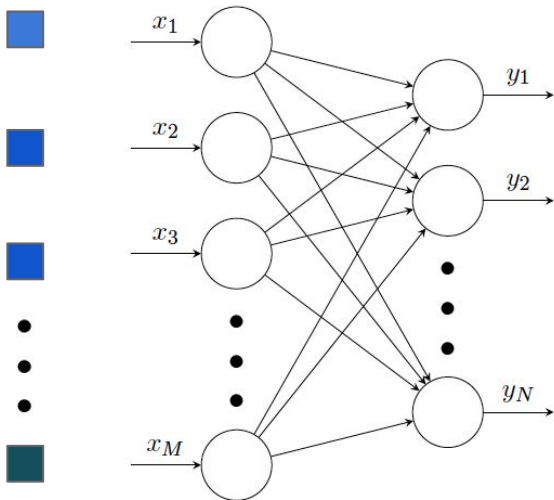
Softmax regression: Multiclass (N classes)

The output logits are normalized with the [softmax function](#), which is **differentiable**:

raw pixels
unrolled img

Input
layer

Ouput
layer



0.3 “dog”



0.08 “cat”



⋮

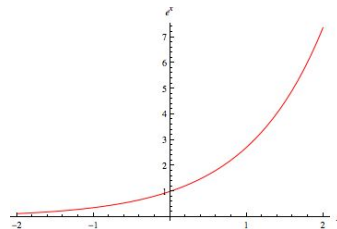


0.6 “whatever”

Softmax
regression

$$P(y = k | \mathbf{x}) = \frac{\exp \mathbf{x}^T \mathbf{w}_k}{\sum_{n=1}^N \exp \mathbf{x}^T \mathbf{w}_n}$$

Exponential $\exp(\cdot)$
boosts higher
logits (max effect).



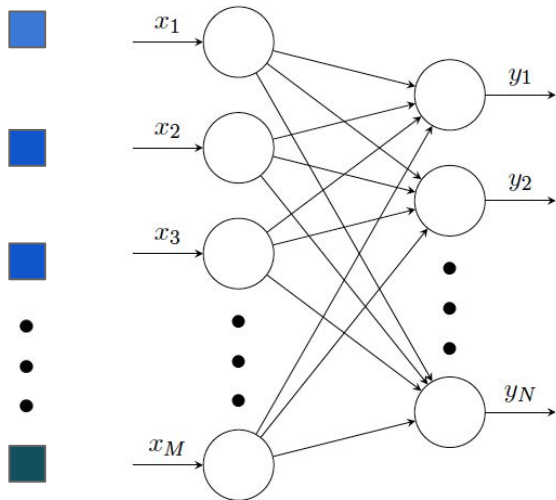
Softmax regression: Multiclass (N classes)

The output logits are normalized with the [softmax function](#), which is **differentiable**:

raw pixels
unrolled img

Input
layer

Ouput
layer



0.3 “dog”



0.08 “cat”



⋮



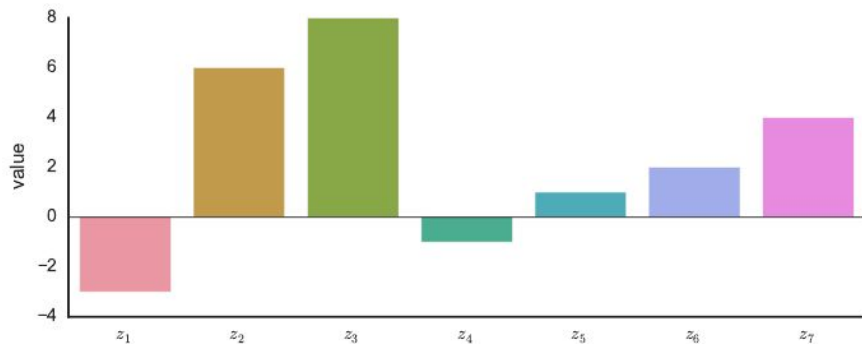
0.6 “whatever”

Softmax
regression

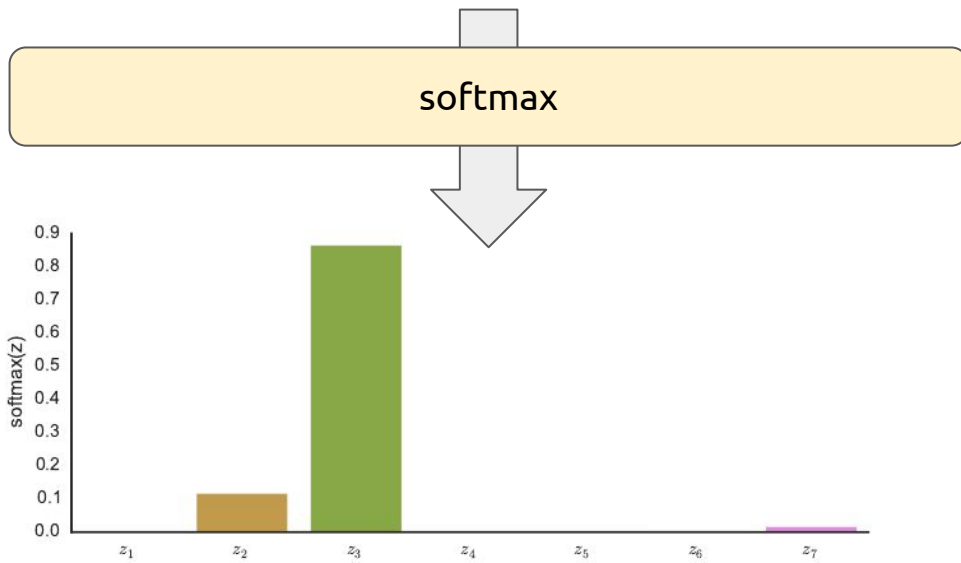
$$P(y = k|\mathbf{x}) = \frac{\exp \mathbf{x}^T \mathbf{w}_k}{\sum_{n=1}^N \exp \mathbf{x}^T \mathbf{w}_n}$$

Normalization factor so that the sum of probabilities sum up to 1.

Softmax regression

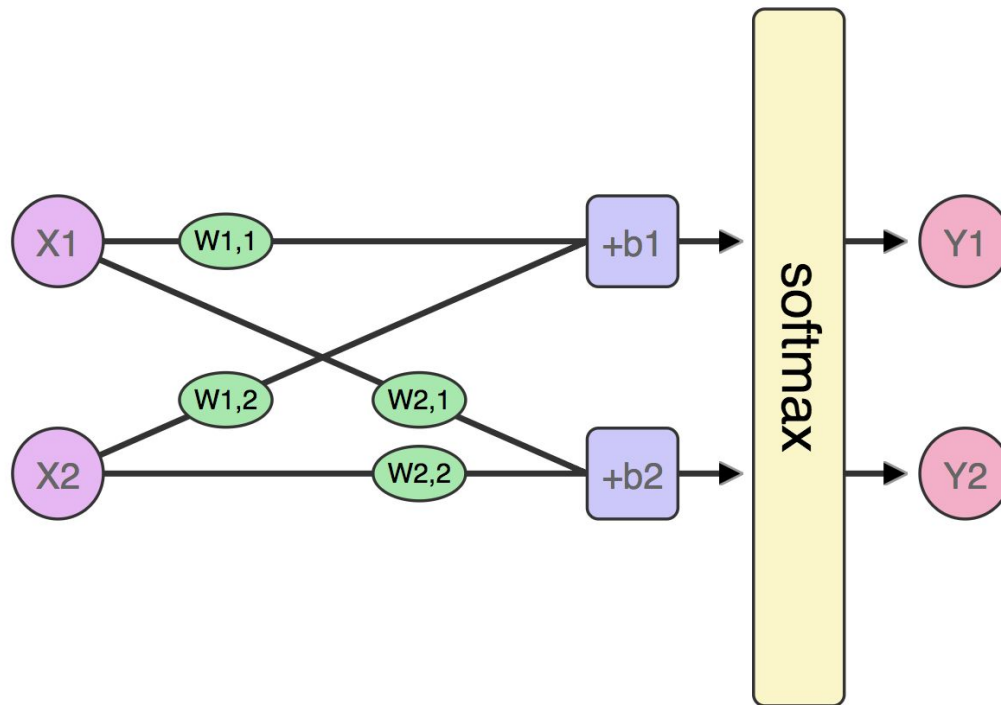


$$\text{softmax}(\mathbf{x}) = \frac{1}{\sum_{j=1}^K \exp(x_j)} \begin{bmatrix} \exp(x_1) \\ \exp(x_2) \\ \vdots \\ \exp(x_K) \end{bmatrix}$$

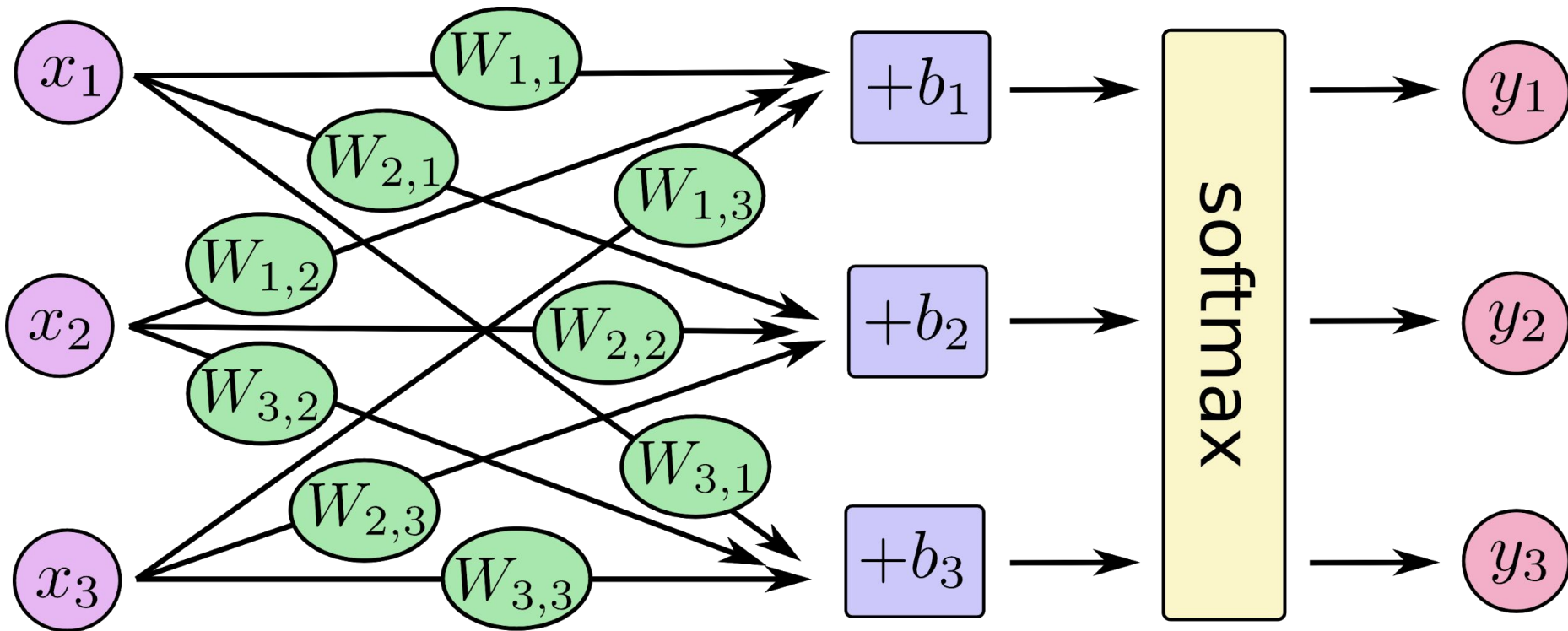


Softmax regression: Binary case

Example: Binary classification can also be solved with two perceptrons + softmax.

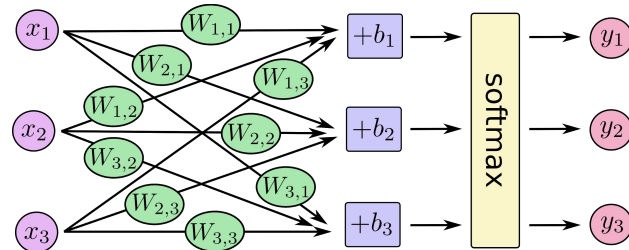


Softmax regression: Multiclass (3 classes)



Softmax regression: Multiclass (3 classes)

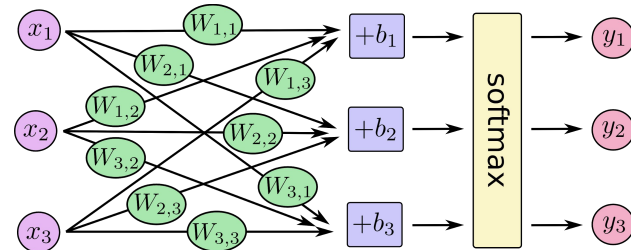
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \text{softmax} \left(\begin{bmatrix} W_{1,1}x_1 + W_{1,2}x_2 + W_{1,3}x_3 + b_1 \\ W_{2,1}x_1 + W_{2,2}x_2 + W_{2,3}x_3 + b_2 \\ W_{3,1}x_1 + W_{3,2}x_2 + W_{3,3}x_3 + b_3 \end{bmatrix} \right)$$



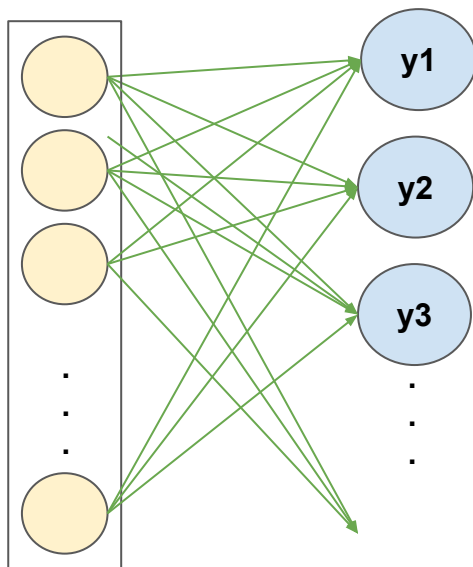
Softmax regression: Multiclass (3 classes)

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \text{softmax} \left(\begin{bmatrix} W_{1,1} & W_{1,2} & W_{1,3} \\ W_{2,1} & W_{2,2} & W_{2,3} \\ W_{3,1} & W_{3,2} & W_{3,3} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right)$$

$$y = \text{softmax}(Wx + b)$$



Software implementation



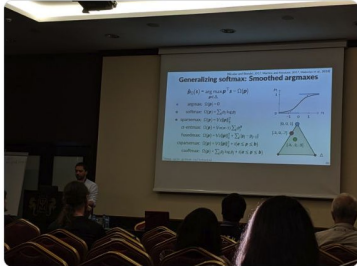
PYTORCH

```
smx = nn.Sequential(  
    nn.Linear(NUM_INPUTS, NUM_OUTPUTS),  
    nn.LogSoftmax(dim=1)  
)
```

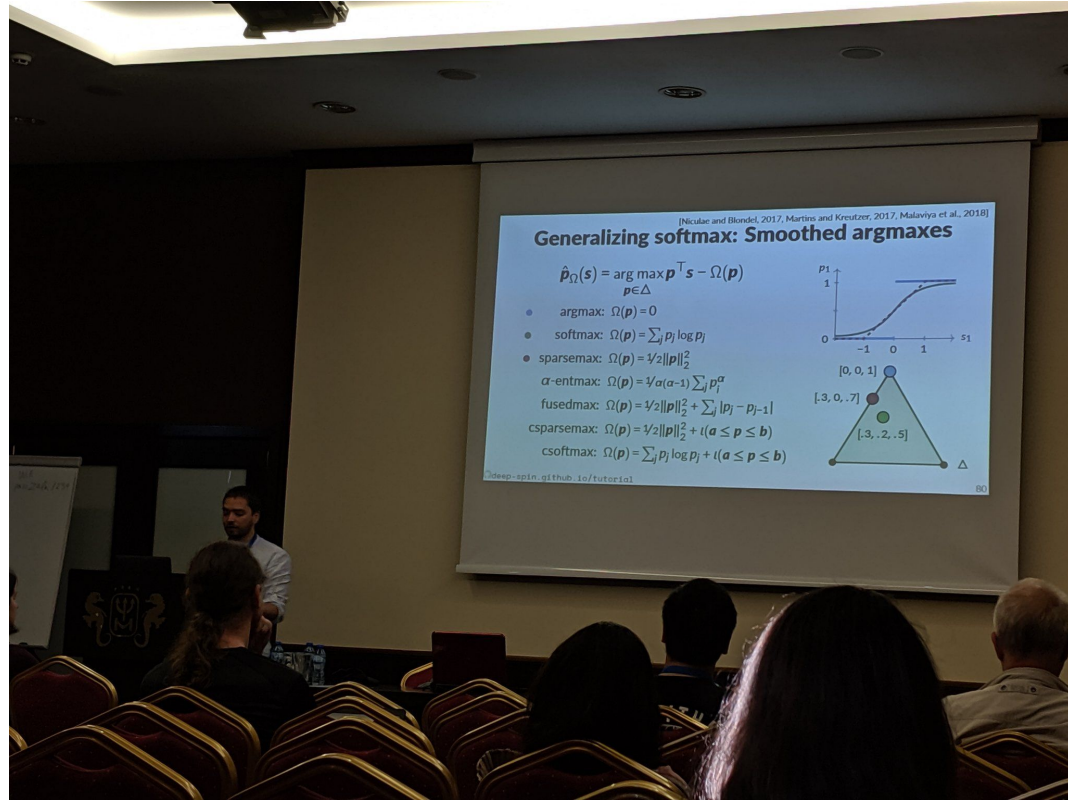
Learn more

Source:
Kyunghyun Cho (@kchonyc)

 Kyunghyun Cho
@kchonyc
a family of x-maxs



4:07 p. m. · 1 set, 2019 · Twitter for Android



End-to-end differentiable relaxations:
<https://deep-spin.github.io/tutorial/acl.pdf>

Undergradese

What undergrads ask vs. what they're REALLY asking

"Is it going to be an open book exam?"

Translation: "I don't have to actually memorize anything, do I?"

"Hmm, what do you mean by that?"

Translation: "What's the answer so we can all go home."

"Are you going to have office hours today?"

Translation: "Can I do my homework in your office?"

"Can i get an extension?"

Translation: "Can you re-arrange your life around mine?"

"Is this going to be on the test?"

Translation: "Tell us what's going to be on the test."

"Is grading going to be curved?"

Translation: "Can I do a mediocre job and still get an A?"

