

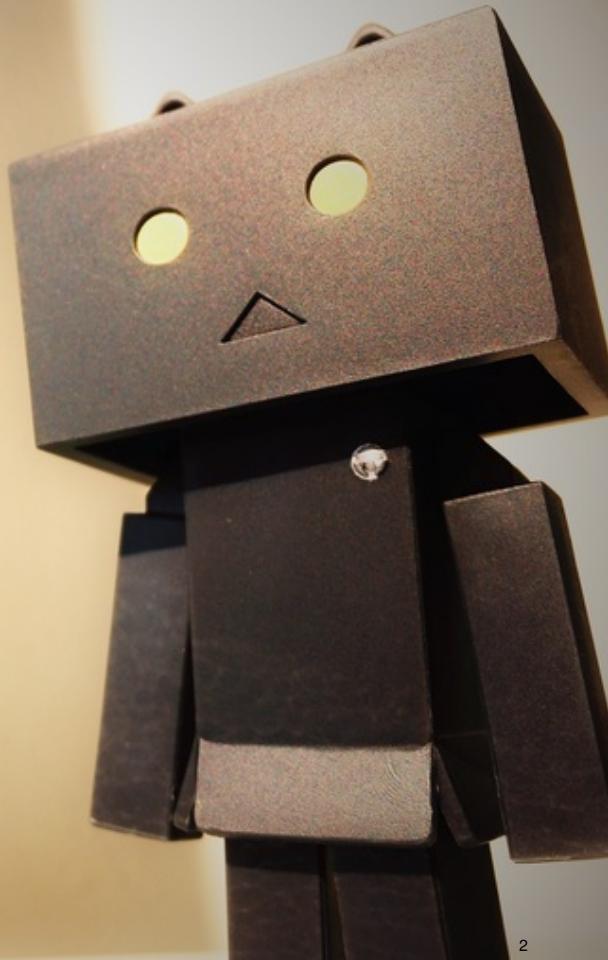
Je vois avec mes yeux

Jeff Abrahamson

20 avril 2017

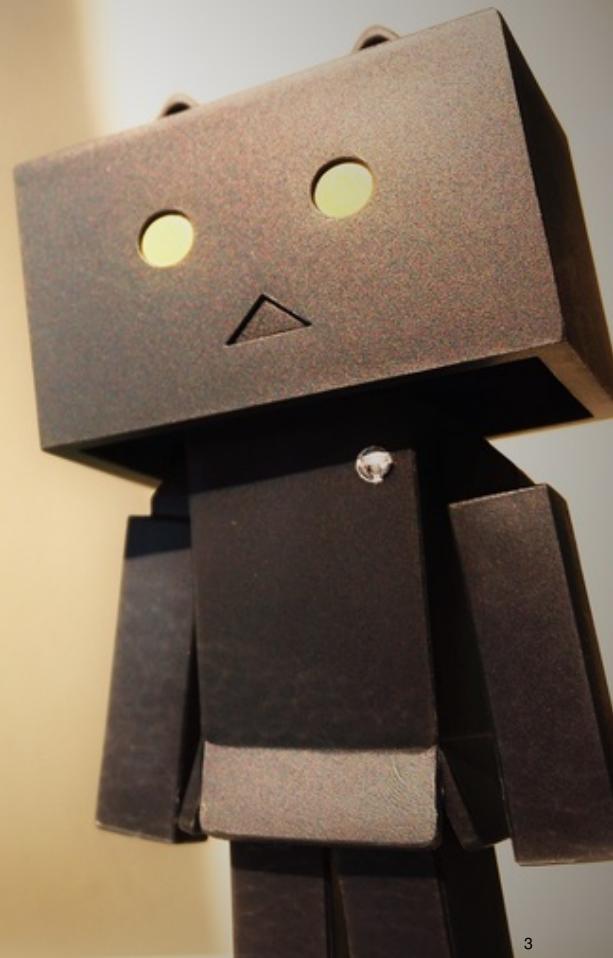
Le problème

Comprendre ce que je fais



Le problème

Prédire



Le problème

Seul : temps libre



Quelles données ?

SOFTPIA

www.softpia.com

Quelles données ?

Mon ordinateur

Quelles données ?

Titre de fenêtre

SOFTPIA
www.softpia.com



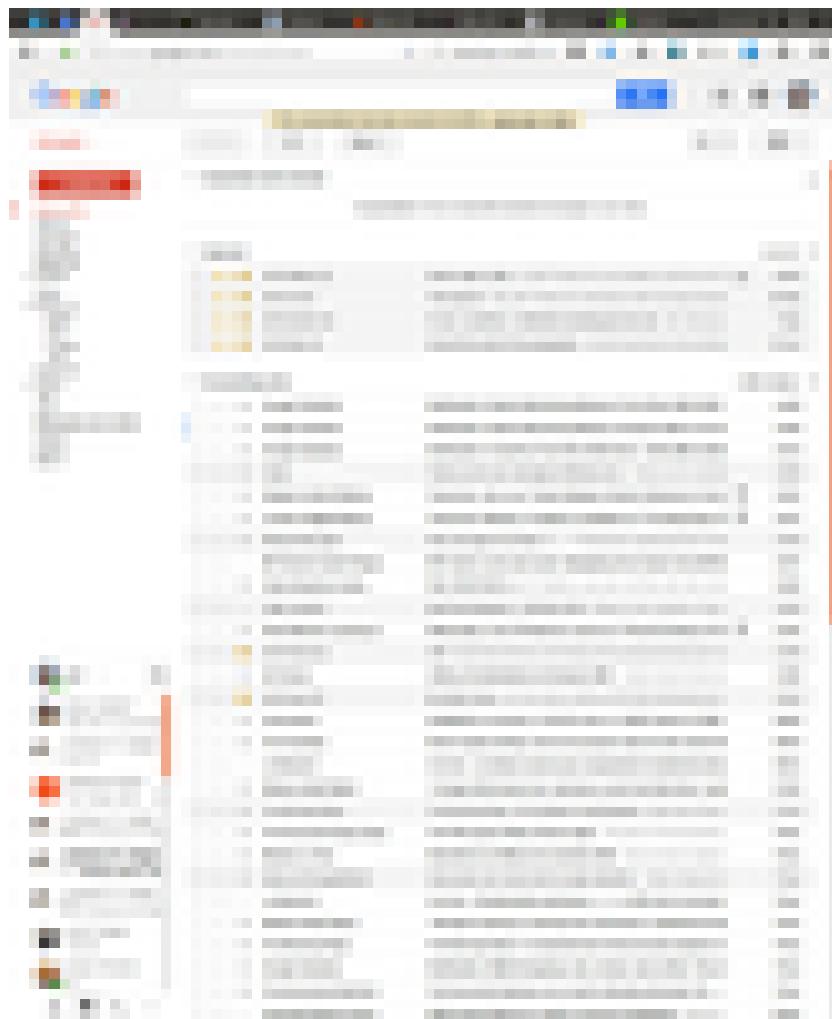
Quelles données ?

Miniature d'image de fenêtre

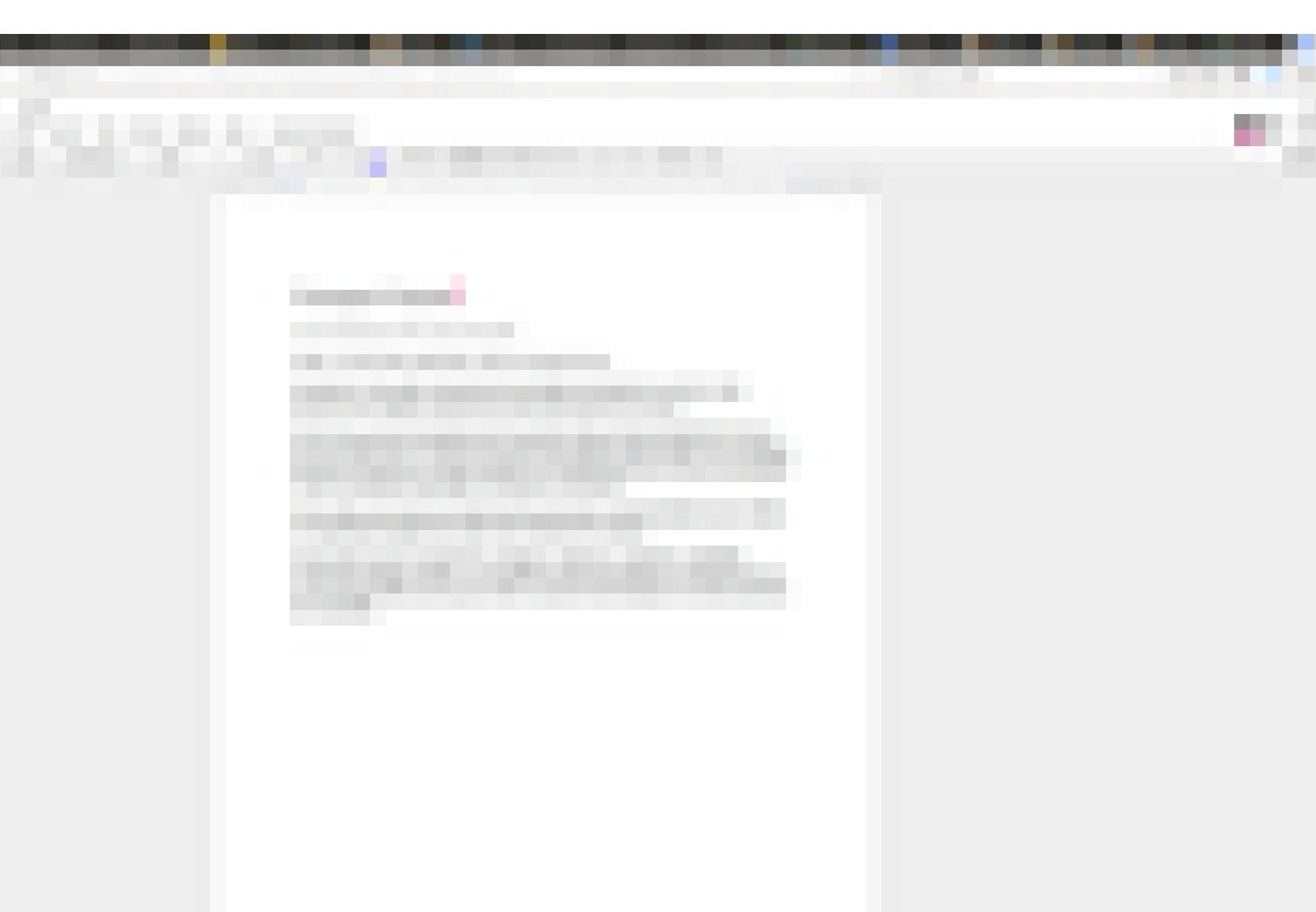
SOFTPIA
www.softpia.com

A photograph showing four people in professional attire (three men in suits and one woman in a teal jacket) standing in front of large windows that look out onto a city skyline with prominent skyscrapers.

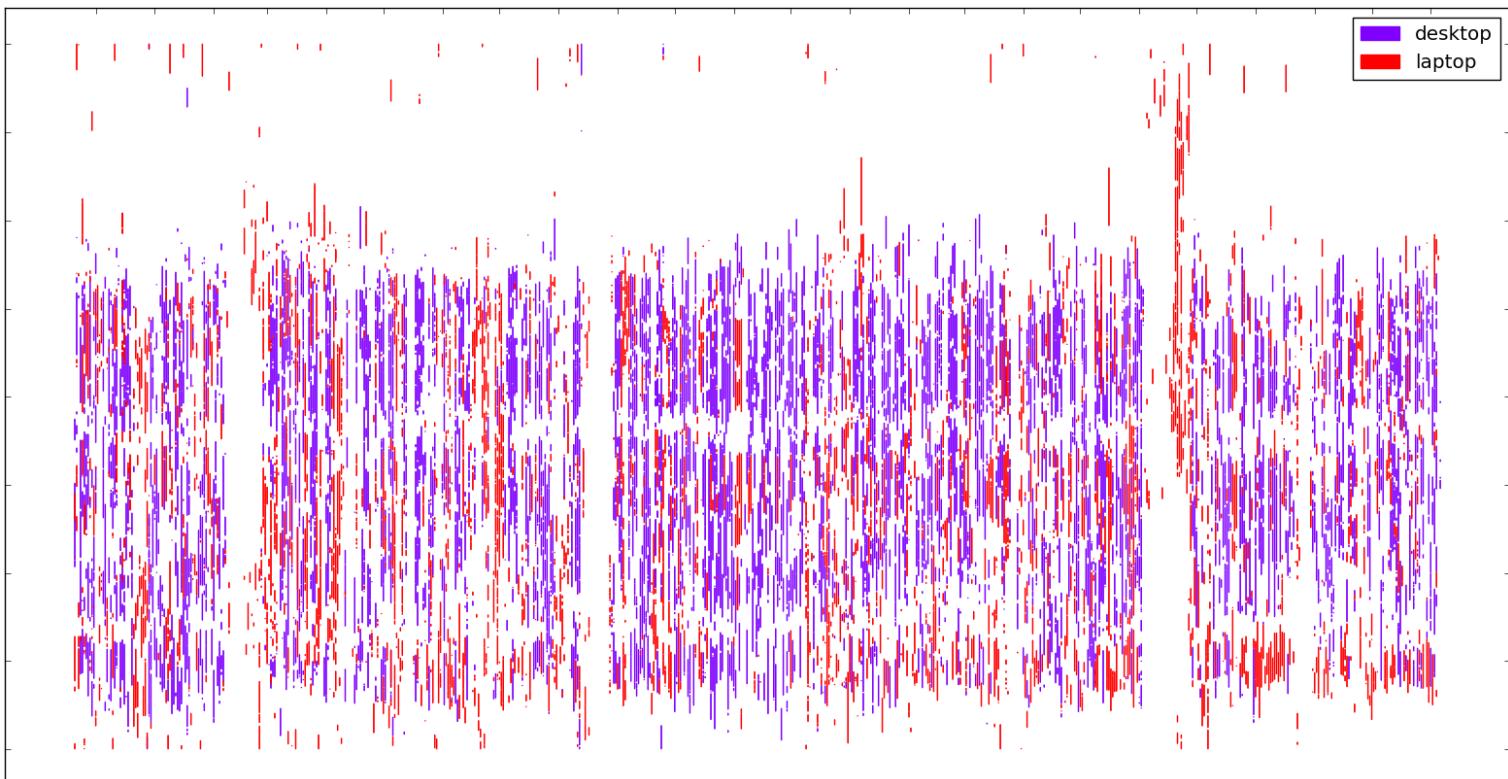
1492626991emacs@birdsong-talk.
tex<2017-04_
_breizhcamp-je-vois-avec-mes-yeux:
/home/jeff/src/jma/talks/2017-04_
_breizhcamp-je-vois-avec-mes-yeux/
talk.tex



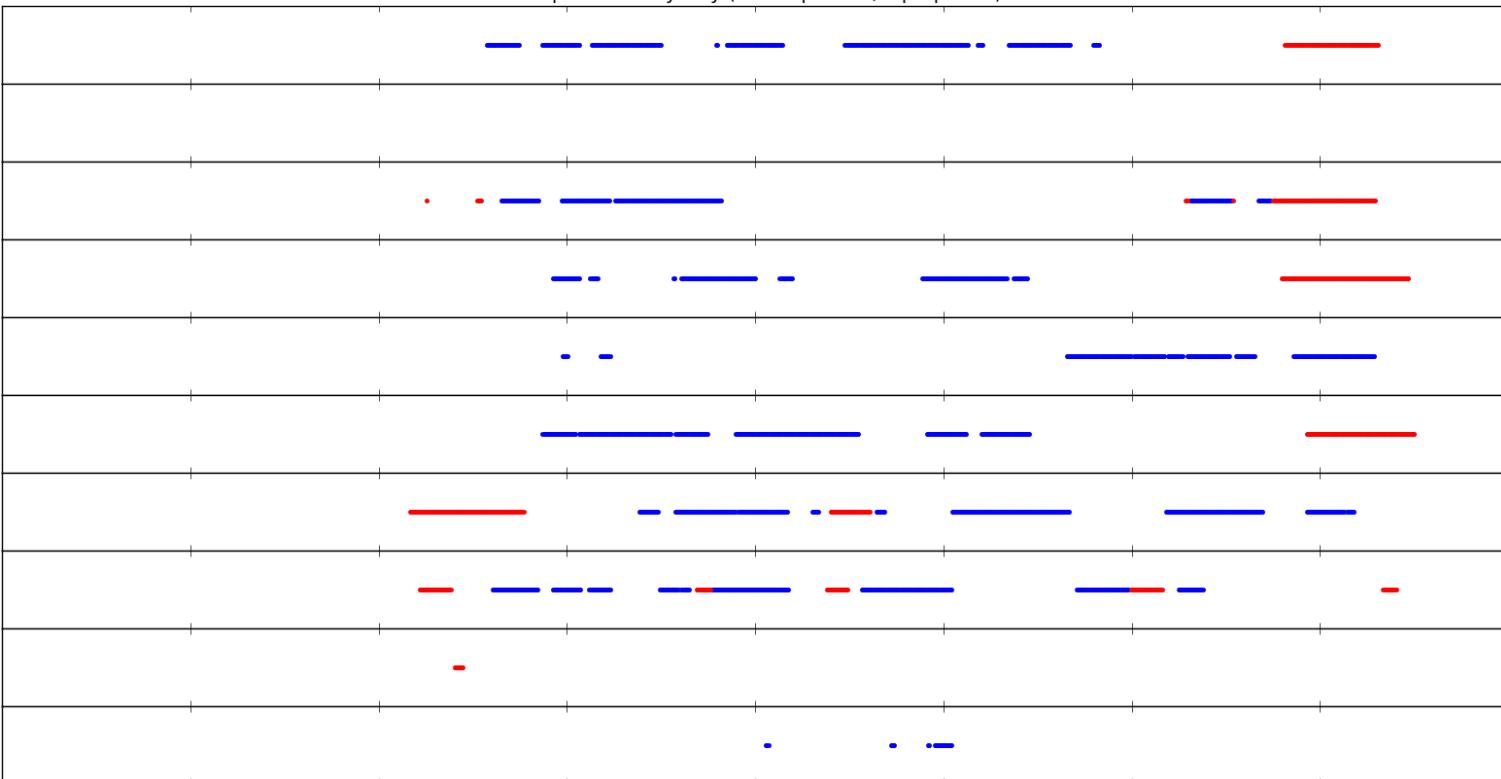




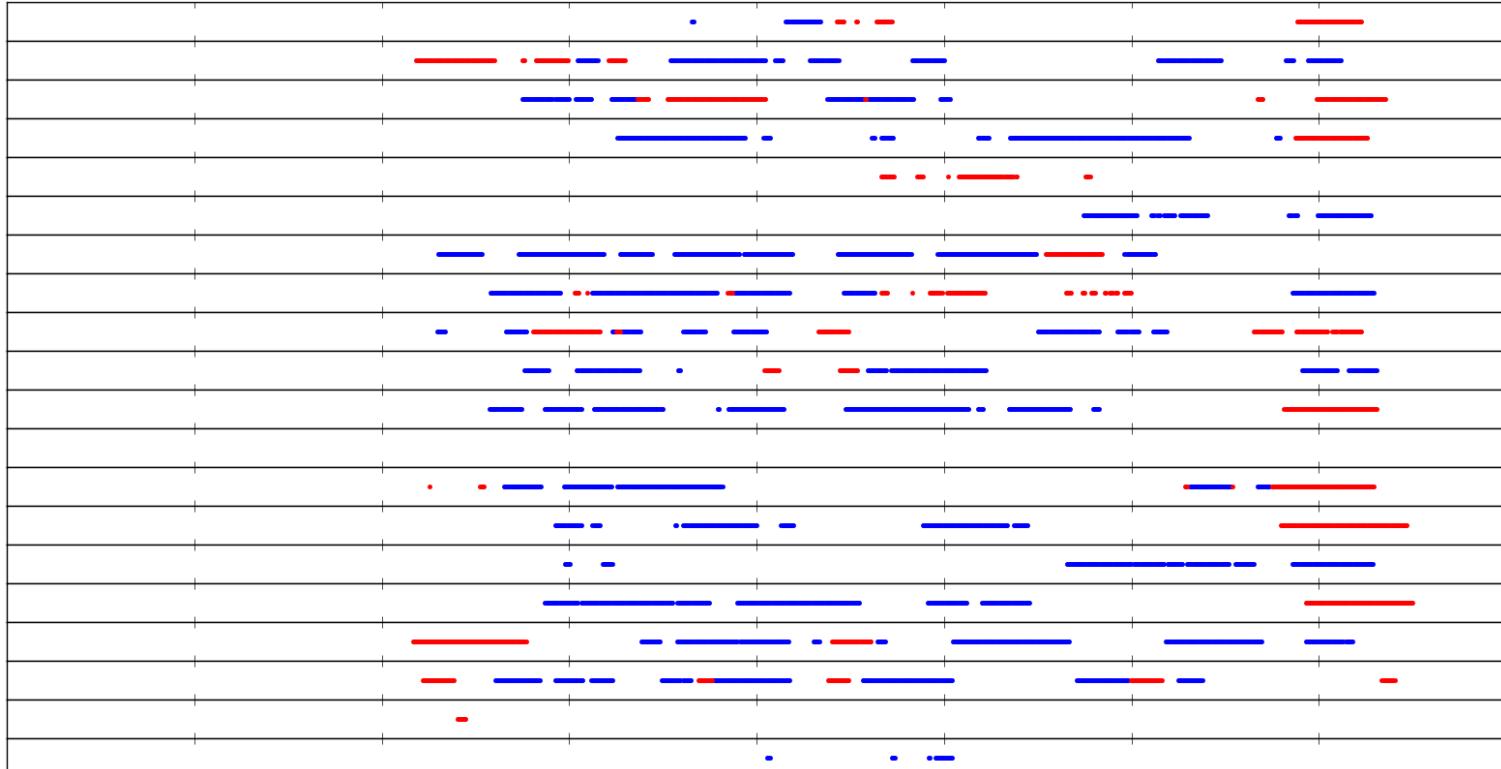




Computer Use by Day (desktop=blue, laptop=red)



Computer Use by Day (desktop=blue, laptop=red)





supervisé



non-supervisé

model



y, \hat{y}



stratégie générale : critères



strategie générale : étiqueter



stratégie générale : auto-étiqueter



stratégie générale : séquences



stratégie générale : prédire



stratégie 1 : sac de mot



stratégie 1 : tf-idf



stratégie 2 : word2vec



deux stratégies : trouver une séquence



Le chat est orange.

Le chien court vite.

6 1 7 2

Le chat est orange.

Le chien court vite.

6 3 4 5

Bag of Words

6 1 7 2

Le chat est orange.

Le chien court vite.

6 3 4 5

Bag of Words

[[6 , 1 , 7 , 2] ,

[6 , 3 , 4 , 5]]

Bag of Words

```
[ [ 6, 1, 7, 2 ]  
[ 1, 1, 0, 0, 0, 1, 1 ]  
[ 0, 0, 1, 1, 1, 1, 0 ]  
[ 6, 3, 4, 5 ] ]
```

Bag of Words

```
[1, 1, 0, 0, 0, 1, 1]  
[0, 0, 1, 1, 1, 1, 0]
```

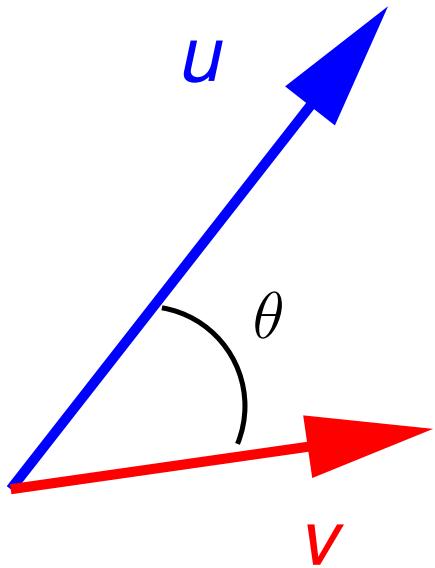
Bag of Words

Le chat est orange.

[1, 1, 0, 0, 0, 1, 1]
[0, 0, 1, 1, 1, 1, 0]

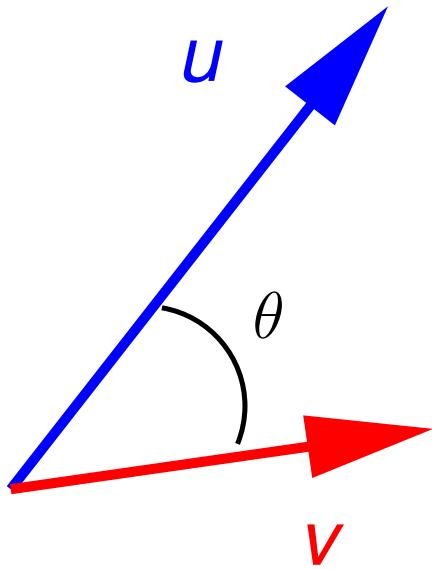
Le chien court vite.

Cosine Similarity



$$\cos \theta = \frac{u \cdot v}{\| u \| \| v \|}$$

Cosine Similarity



$$\cos \theta = u \cdot v$$

(if u and v have norm 1)

Le chat est orange.

[1, 1, 0, 0, 0, 1, 1]

[0, 0, 1, 1, 1, 1, 0]

Le chien court vite.

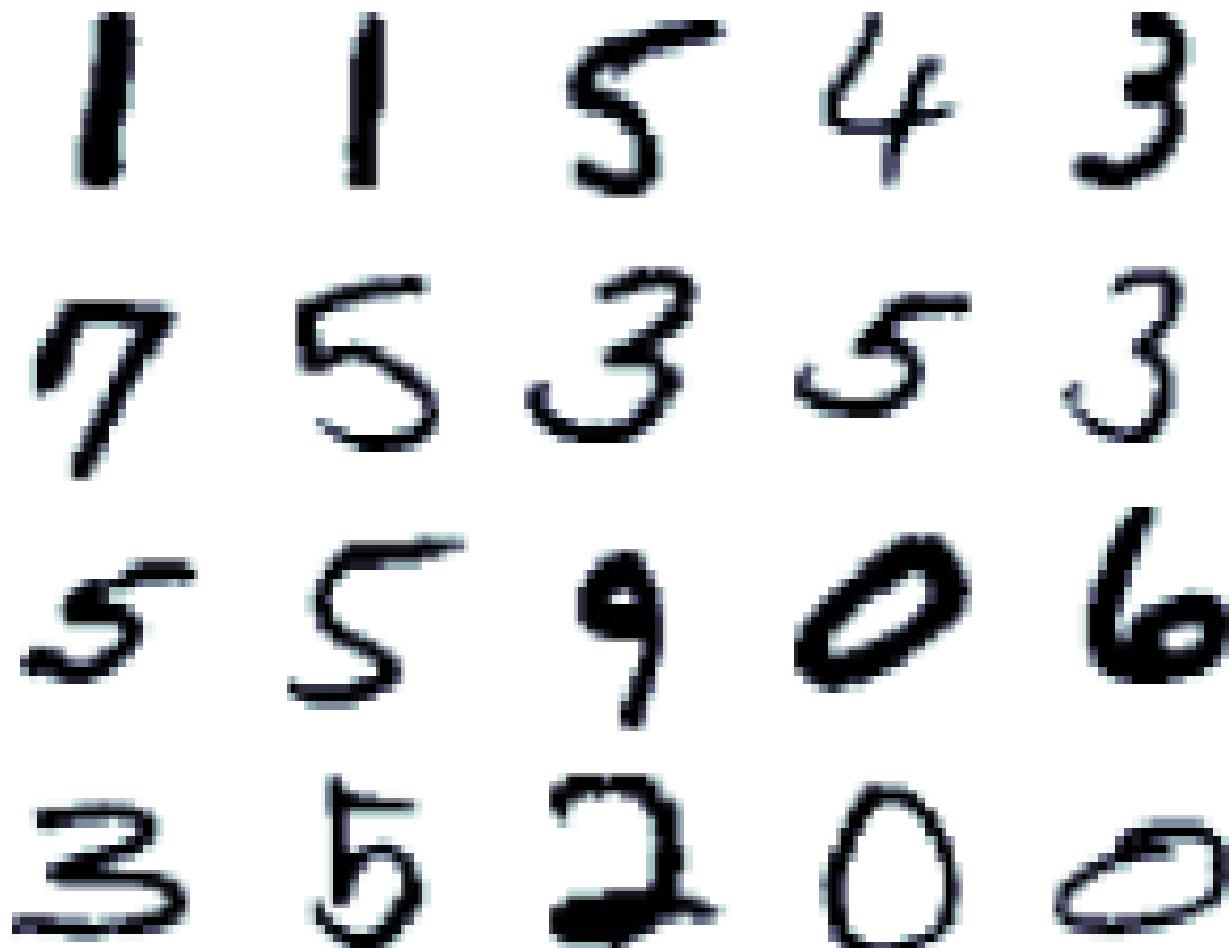
Cosine Similarity

$$u = [1, 1, 0, 0, 0, 1, 1]$$

$$v = [0, 0, 1, 1, 1, 1, 0]$$

$$u \cdot v = 0 + 0 + 0 + 0 + 0 + 1 + 0 = 1$$

$$\cos \theta = \frac{u \cdot v}{\| u \| \| v \|} = \frac{1}{\sqrt{4} \cdot \sqrt{4}} = \frac{1}{4}$$



Un exemple (trop) vite

Linear neuron

$$y = b + \sum_i x_i w_i$$

$$y = b + \sum_i x_i w_i$$

where

y = output

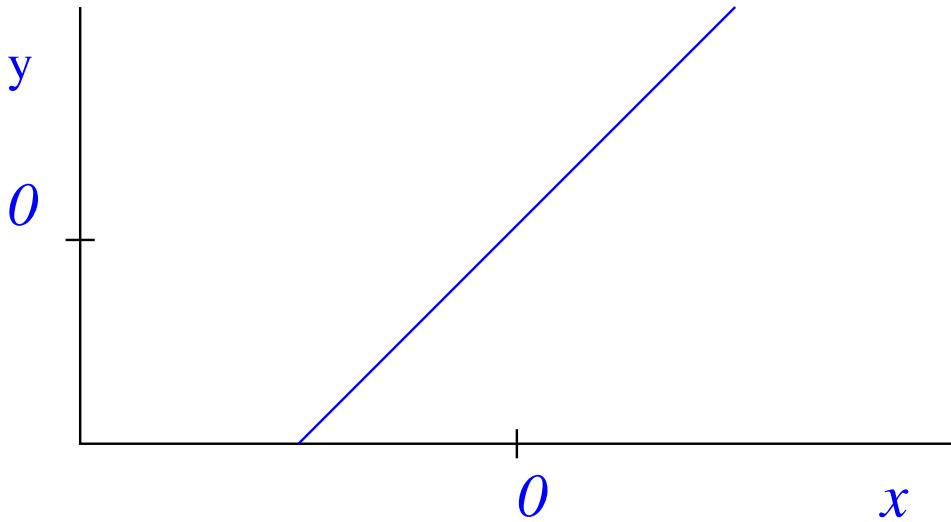
b = bias

x_i = i^{th} input

w_i = weight on i^{th} input

Linear neuron

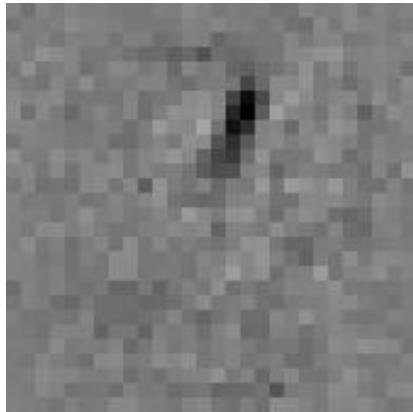
$$y = b + \sum_i x_i w_i$$



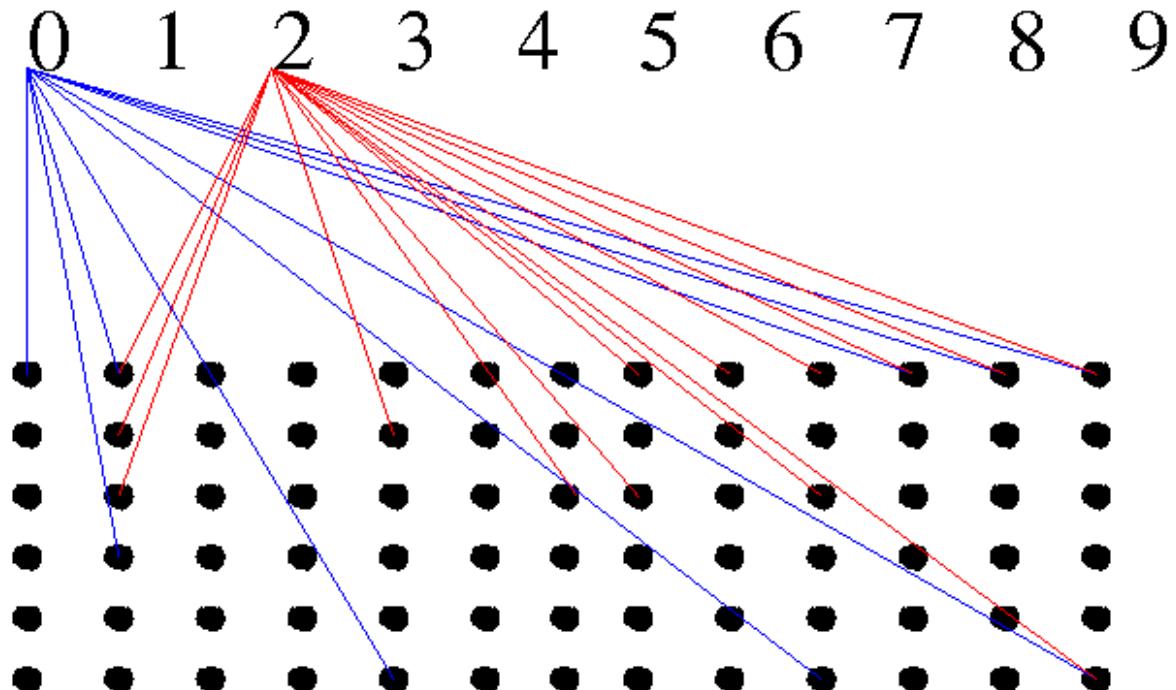
- Input neurons: pixels
- Output neurons: classes (digits)
- Connect them all! (*bipartite*)

- Input neurons: pixels
- Output neurons: classes (digits)
- Connect them all! (*bipartite*)
- Initialize input weights to random

- Input neurons: pixels
- Output neurons: classes (digits)
- Connect them all! (*bipartite*)
- Initialize input weights to random



Example: handwriting recognition of digits



To train this ANN:

- Increment weights from active pixels going to correct class
- Decrement weights from active pixels going to predicted class

To train this ANN:

- Increment weights from active pixels going to correct class
- Decrement weights from active pixels going to predicted class

When it's right, nothing happens. This is good.

stratégie 1 : sac de mot



stratégie 1 : tf-idf

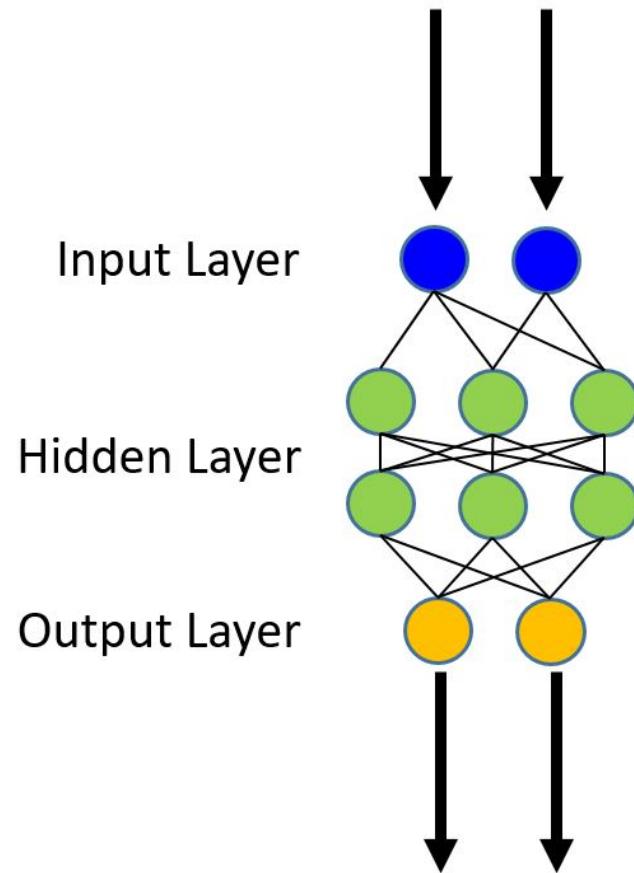


stratégie 2 : word2vec



deux stratégies : trouver une séquence





stratégie 1 : sac de mot



stratégie 1 : tf-idf



stratégie 2 : word2vec



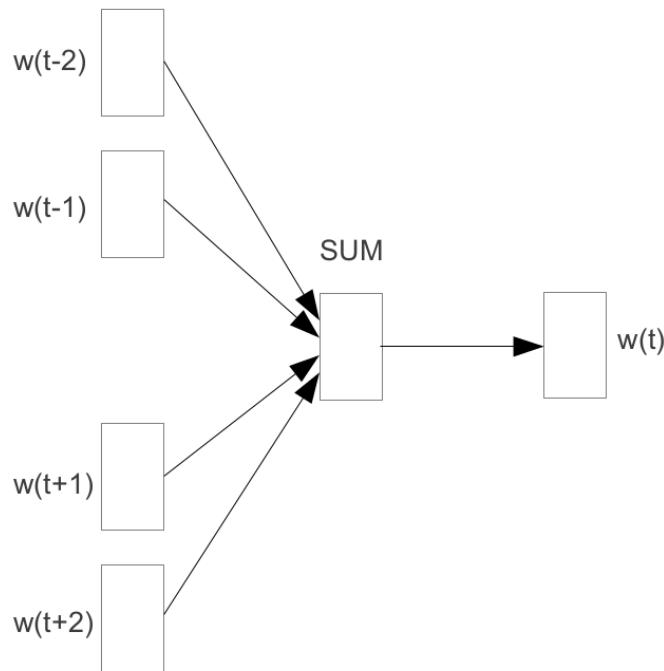
deux stratégies : trouver une séquence



INPUT

PROJECTION

OUTPUT

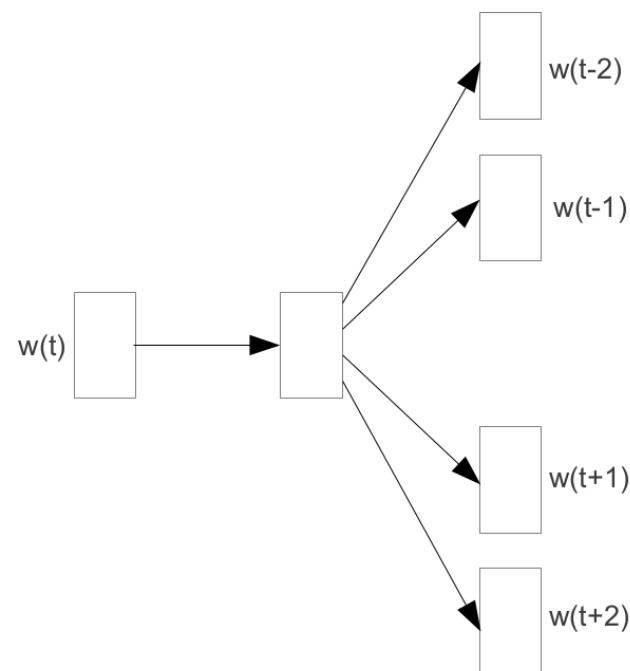


CBOW

INPUT

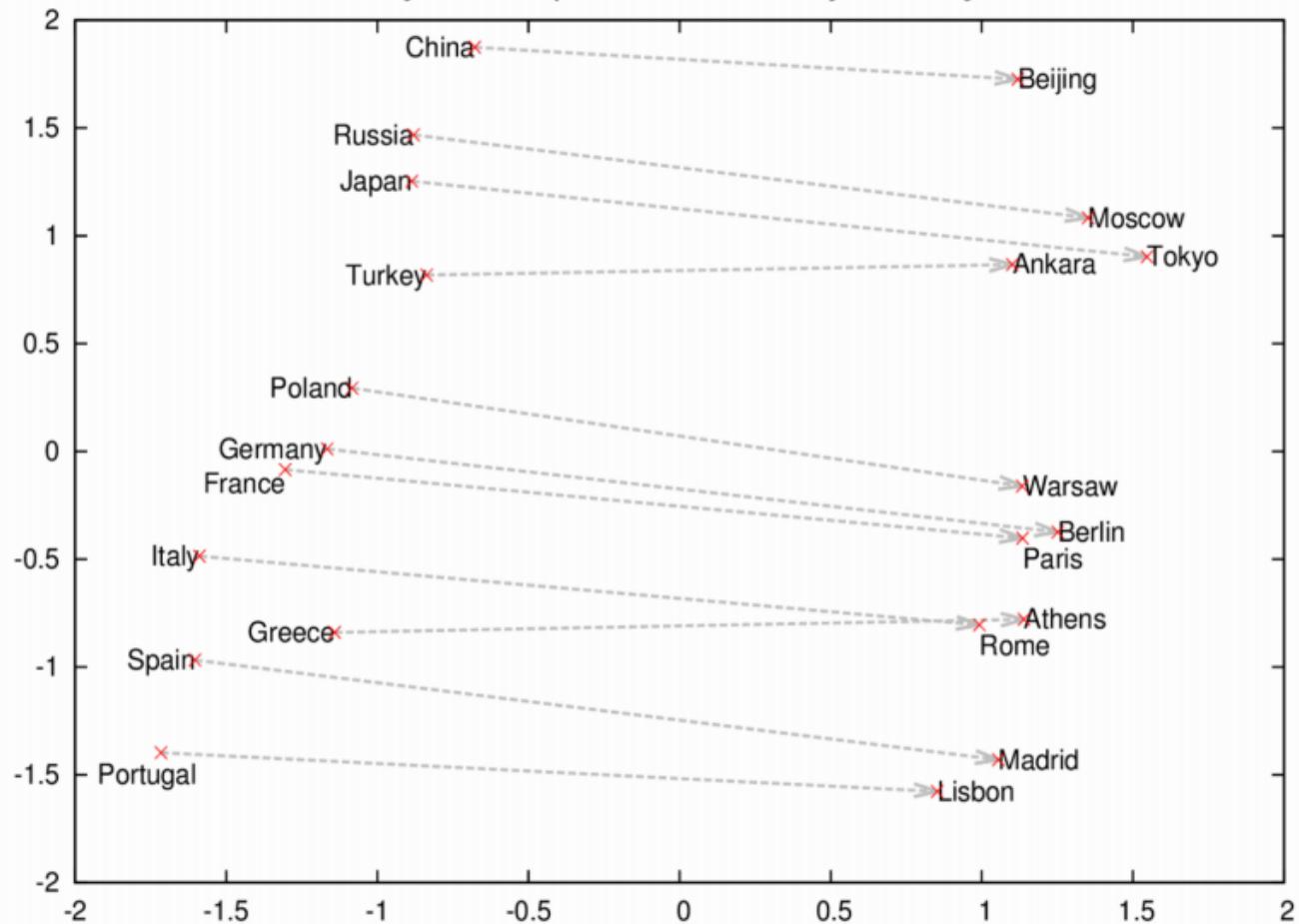
PROJECTION

OUTPUT



Skip-gram

Country and Capital Vectors Projected by PCA



Term	Similarity
	"shift" 0.933104
	"gown" 0.887743
	"skirt" 0.881672
	"bandage" 0.880162
	"midi" 0.869786

Similar to 'dress'



stratégie générale : critères



strategie générale : étiqueter



stratégie générale : auto-étiqueter



stratégie générale : séquences



stratégie générale : prédire





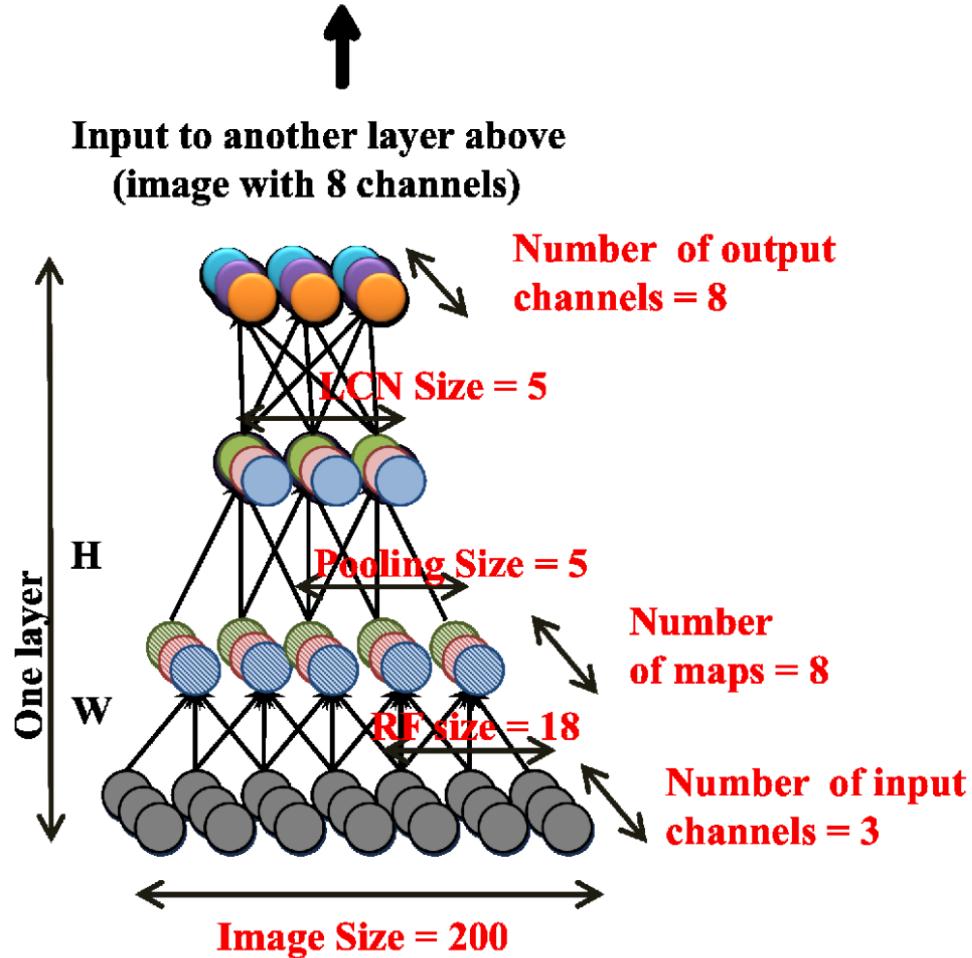
stratégie : étiqueter

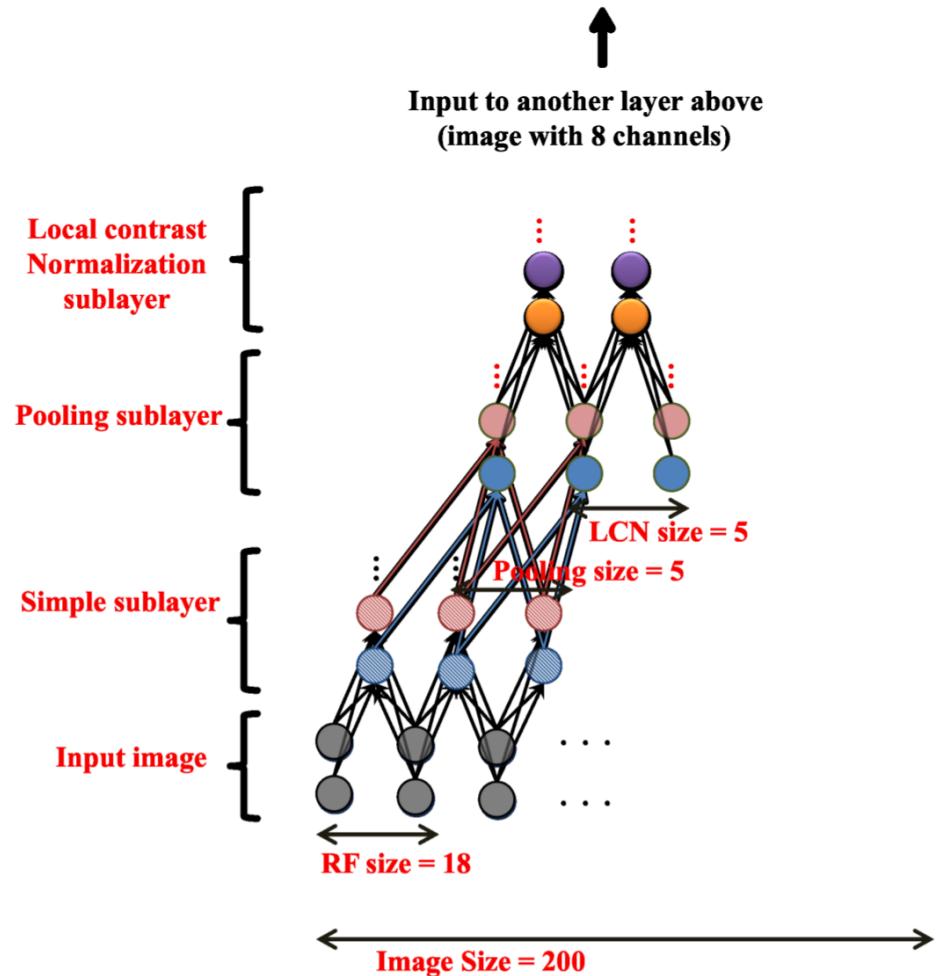
A close-up photograph of a ginger cat with prominent stripes. The cat is lying on its side, facing the camera with wide, intense brown eyes. Its fur is a rich orange-red color. The background is dark and textured, possibly a wall or a piece of furniture.

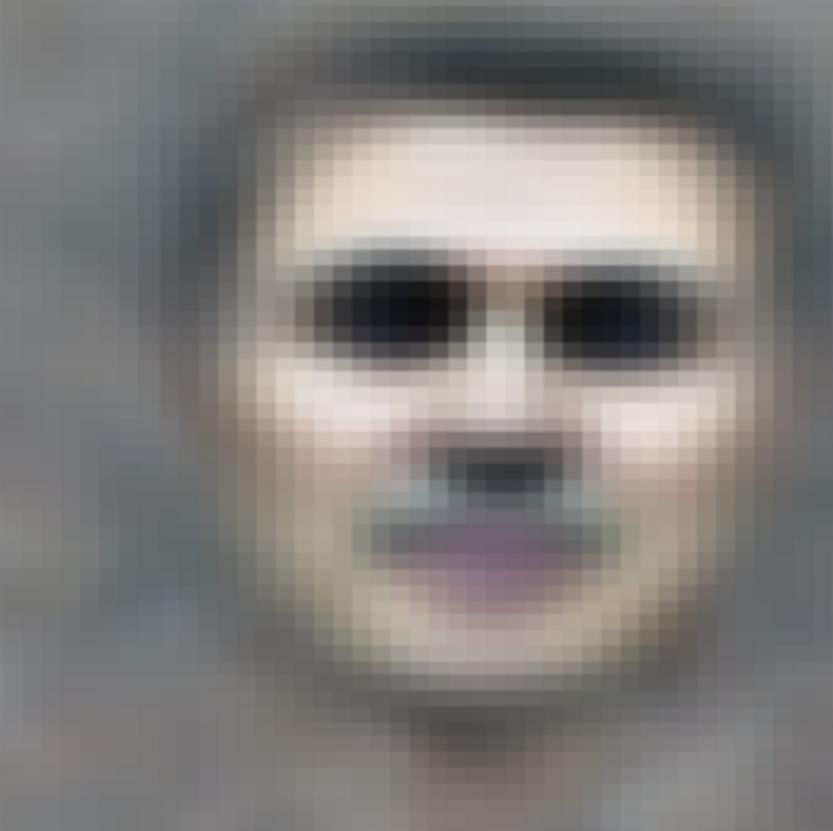
stratégie : auto-étiqueter

A close-up photograph of a ginger tabby cat lying on its side, looking directly at the camera with wide, intense brown eyes. The cat's fur is a vibrant orange-red color with distinct dark stripes. It is positioned against a textured, reddish-brown background, possibly a wall or a piece of furniture. The lighting highlights the texture of the cat's fur and the depth of its gaze.

stratégie : séquence









A close-up photograph of a ginger tabby cat lying on its side, looking directly at the camera with wide, intense brown eyes. The cat's fur is a mix of orange and brown stripes. It is positioned against a dark, textured background, possibly a wall or a piece of furniture.

stratégie : ImageNet



stratégie : DGG-19, resnet

A close-up photograph of a ginger tabby cat lying on its side, looking directly at the camera with wide, intense brown eyes. The cat's fur is a vibrant orange-red color with distinct dark stripes. It is positioned against a textured, reddish-brown background, possibly a wall or a piece of furniture. The lighting highlights the texture of the cat's fur and the depth of its gaze.

stratégie : séquence

$a_1, a_2, \dots, a_k \Rightarrow a_{k+1}$

Meetup Machine Learning Rennes

<https://www.meetup.com/Meetup-Machine-Learning-Rennes/>



<http://www.meetup.com/Nantes-Machine-Learning-Meetup/>



<http://www.ml-week.com/>



Questions?