Je vois avec mes yeux

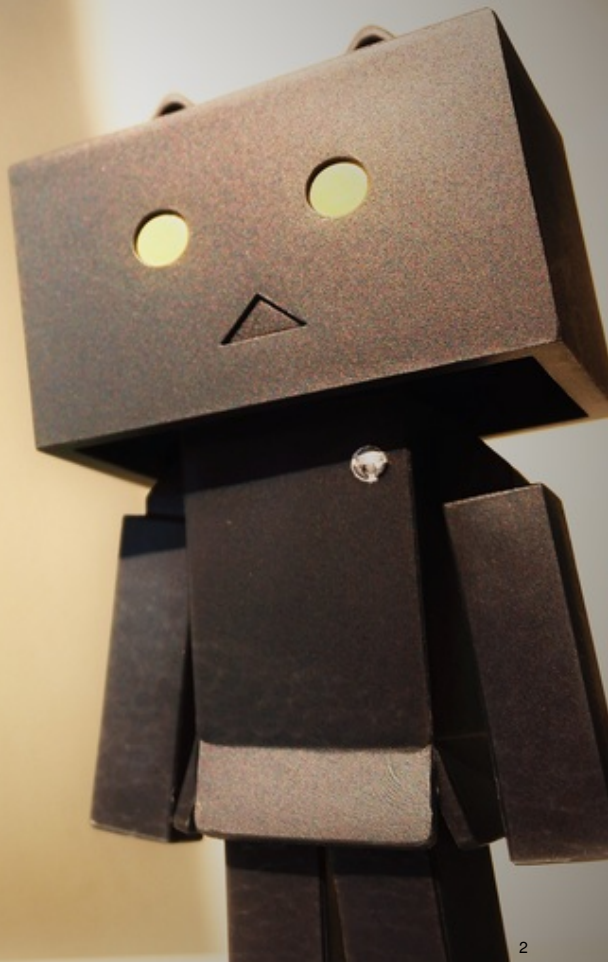
Jeff Abrahamson

20 avril 2017

Copyright 2016 Jeff Abrahamson

Le problème

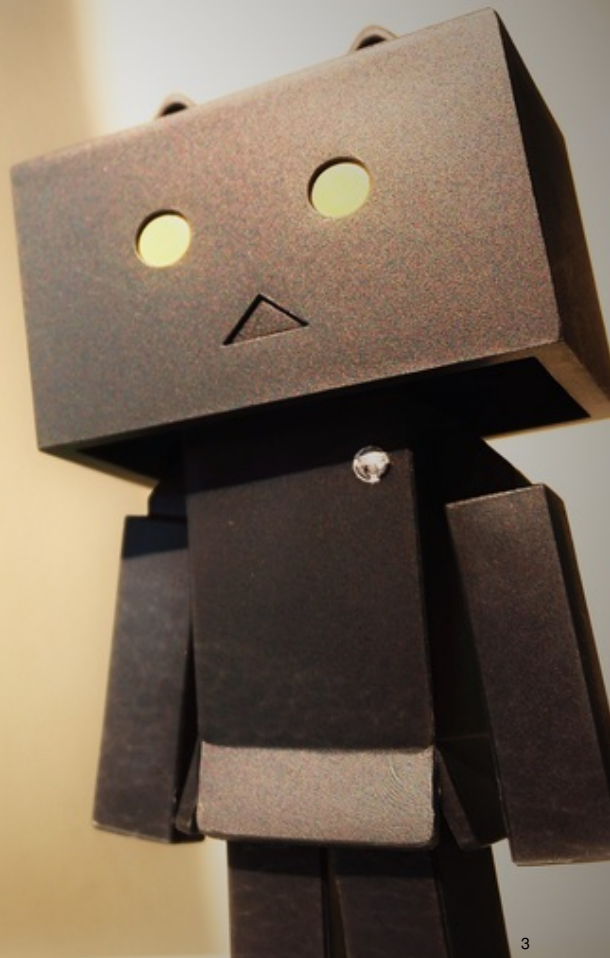
Comprendre ce que je fais



Copyright 2016 Jeff Abrahamson

Le problème

Prédire



Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson

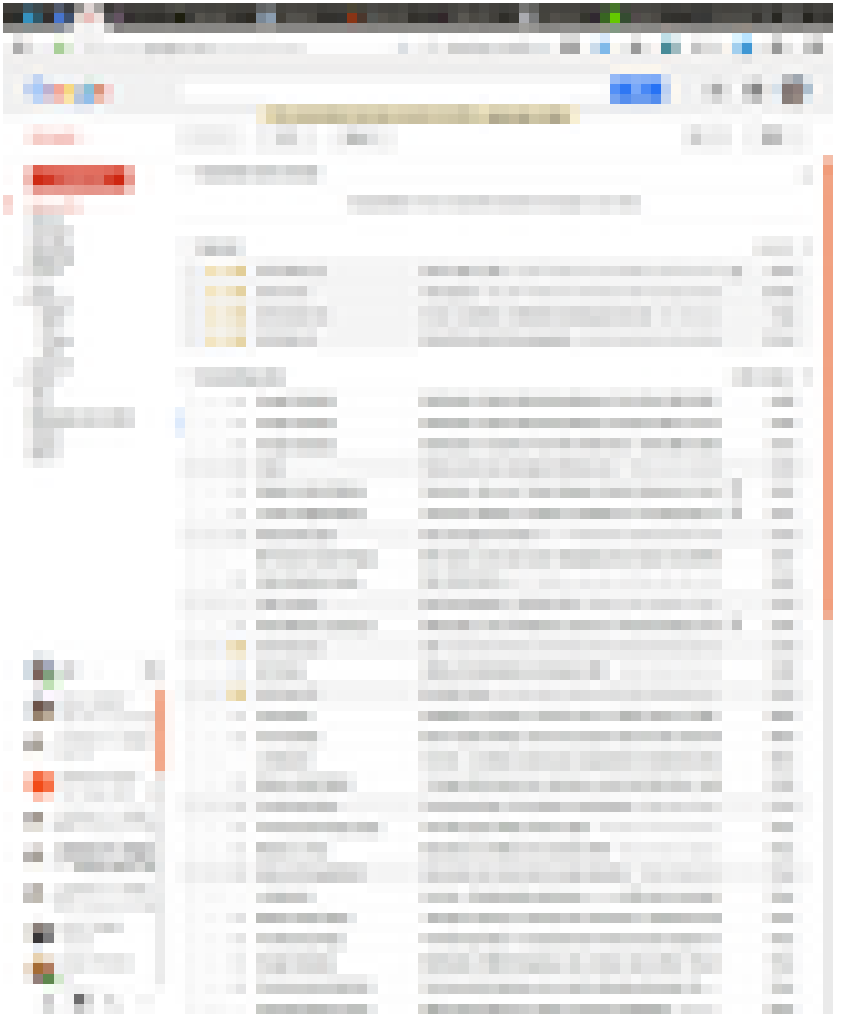
Quelles données ?

Quelles données ?
Mon ordinateur

Quelles données ?
Titre de fenêtre

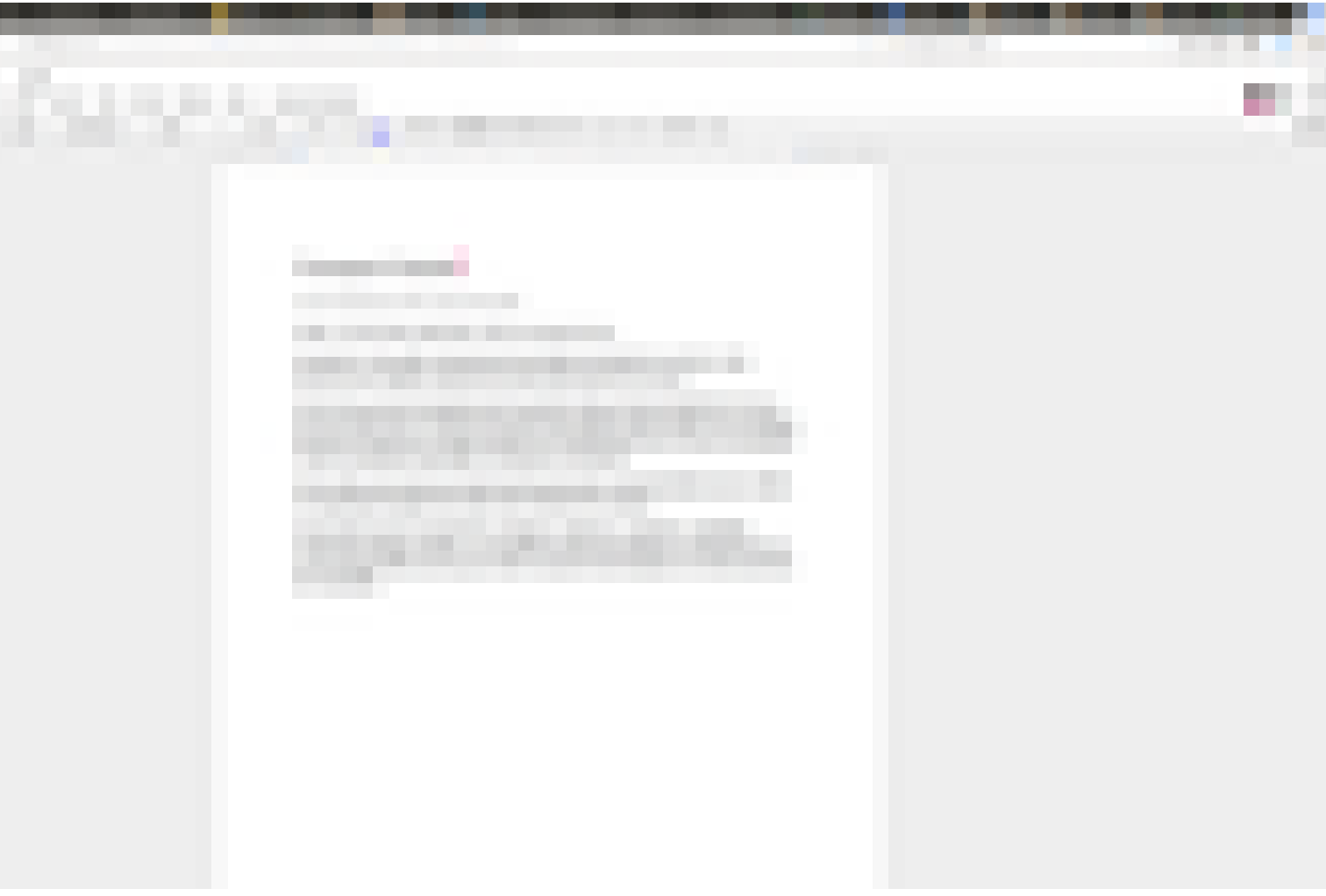
Quelles données ?
Miniature d'image de fenêtre

Copyright 2016 Jeff Abrahamson



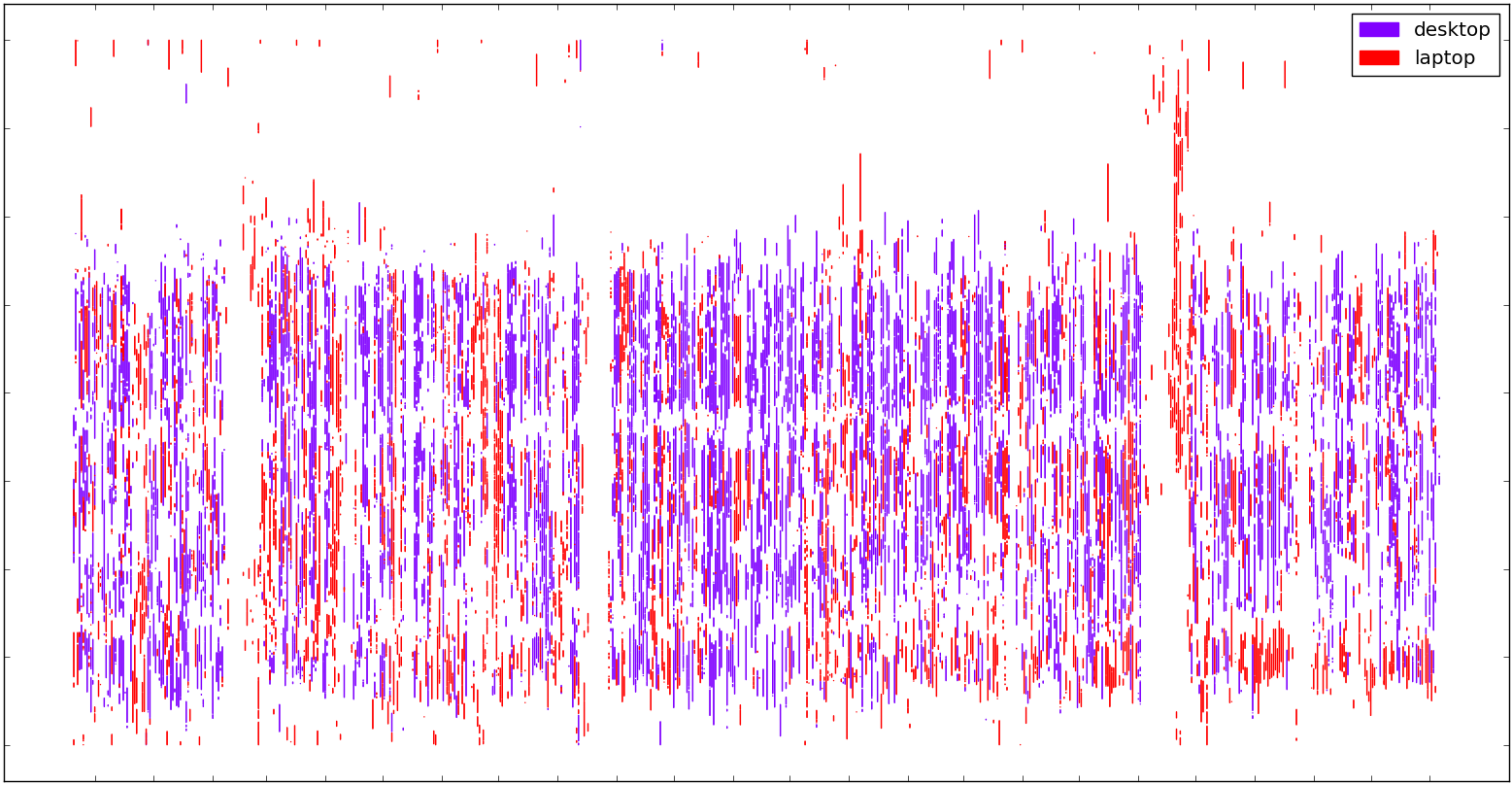
Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson



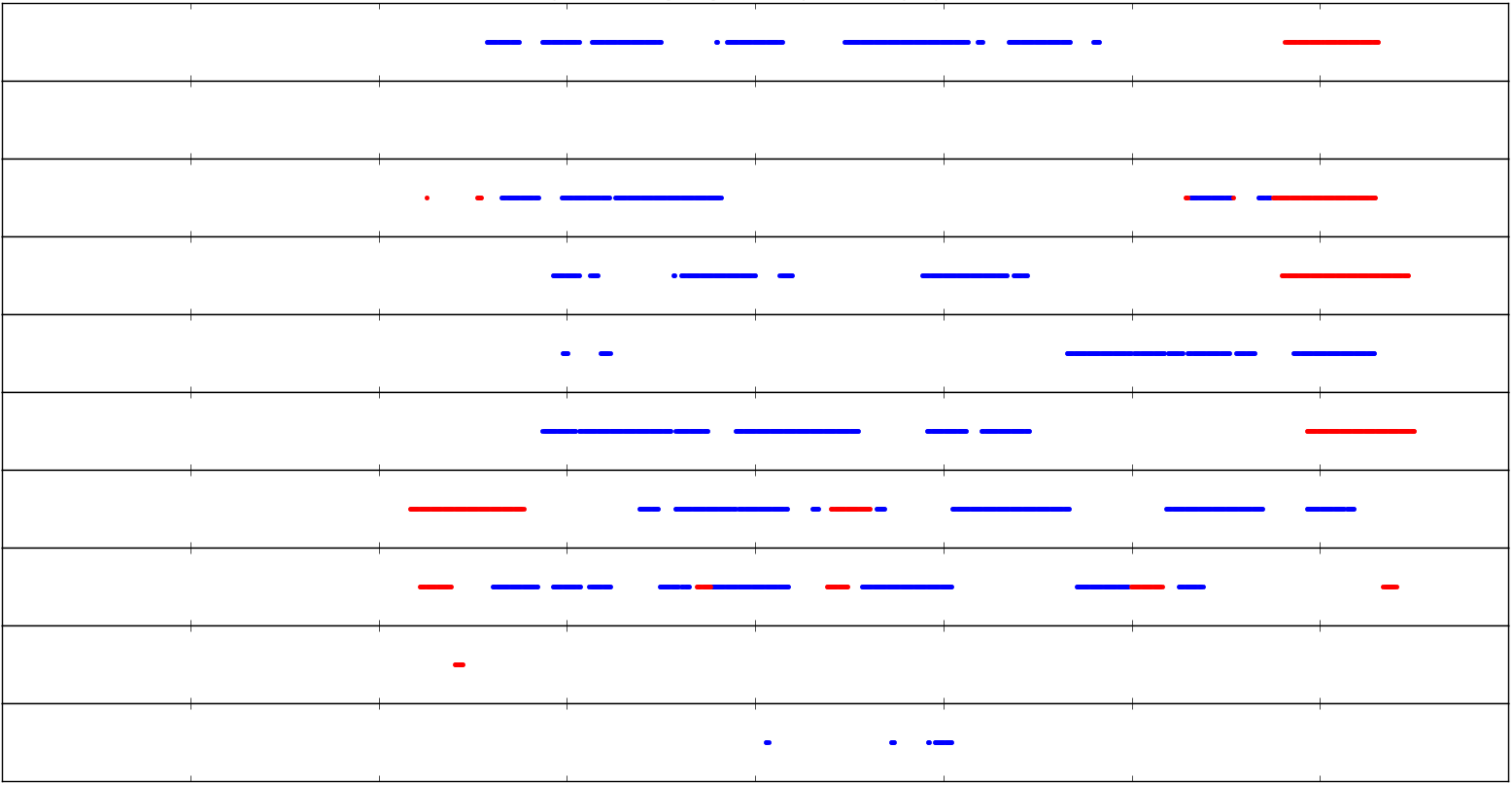
Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson

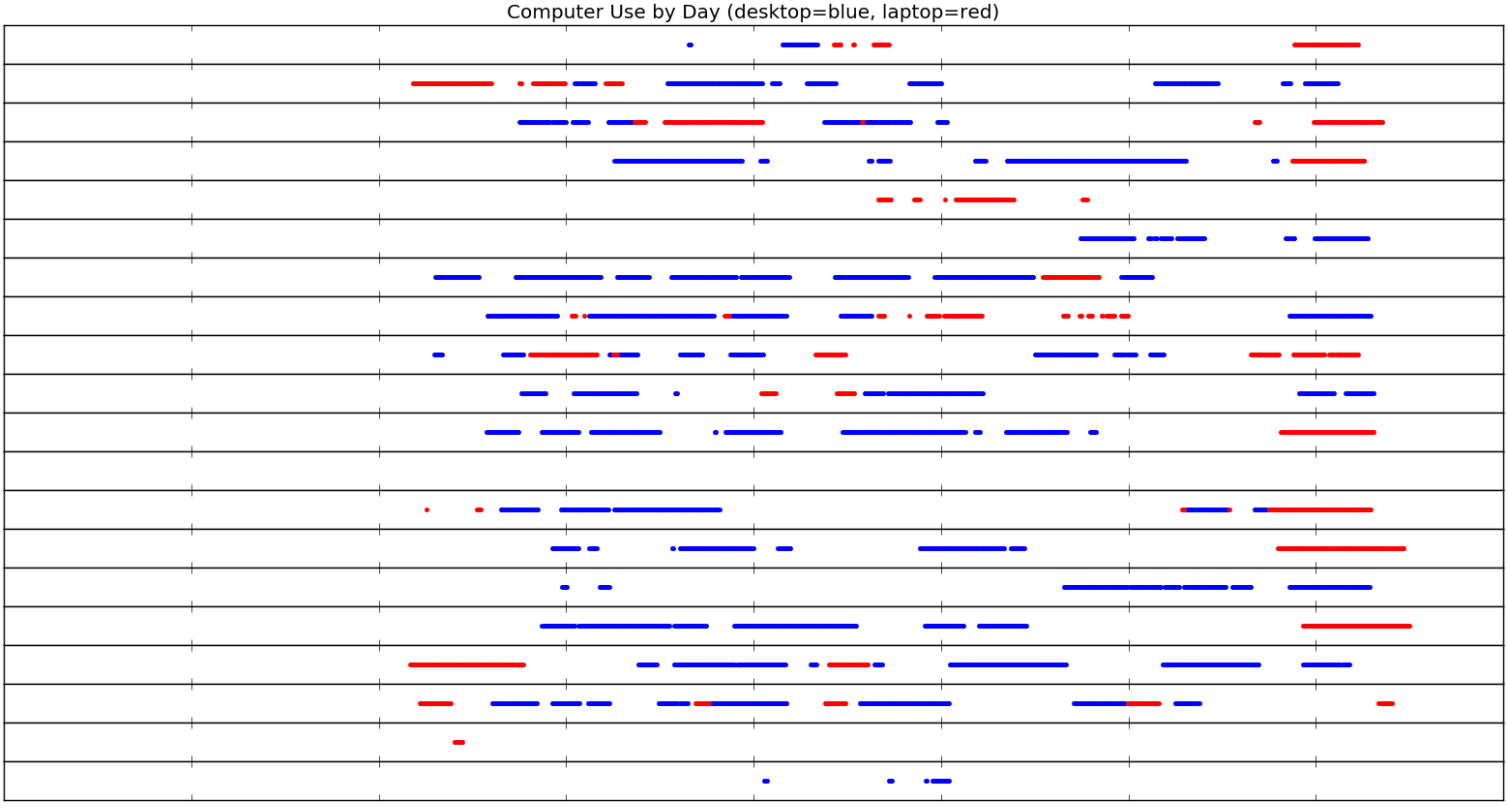


Copyright 2016 Jeff Abrahamson

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



Copyright 2016 Jeff Abrahamson



Copyright 2016 Jeff Abrahamson

supervisé

Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson

model

Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson

strategie générale : critères

Copyright 2016 Jeff Abrahamson

strategie générale : étiqueter

Copyright 2016 Jeff Abrahamson

strategie générale : auto-étiqueter

Copyright 2016 Jeff Abrahamson

strategie générale : séquences

Copyright 2016 Jeff Abrahamson

strategie générale : prédire

Copyright 2016 Jeff Abrahamson

stratégie 1 : sac de mot

Copyright 2016 Jeff Abrahamson

stratégie 1 : tf-idf

Copyright 2016 Jeff Abrahamson

stratégie 2 : word2vec

Copyright 2016 Jeff Abrahamson

deux stratégies : trouver une séquence

Copyright 2016 Jeff Abrahamson

```
Le chat est orange.
Le chien court vite.
```

Copyright 2016 Jeff Abrahamson

```
6   1    7    2
Le chat est orange.
Le chien court vite.
6    3     4    5
```

Copyright 2016 Jeff Abrahamson

6 1 7 2

Le chat est orange.

Le chien court vite.

6 3 4 5

Copyright 2016 Jeff Abrahamson

Bag of Words

```
[[6, 1, 7, 2],

 [6, 3, 4, 5]]
```

Copyright 2016 Jeff Abrahamson

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]]
```

Copyright 2016 Jeff Abrahamson

Bag of Words

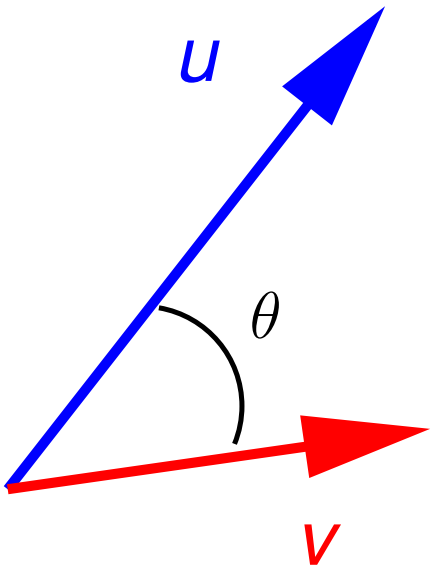
```
[1, 1, 0, 0, 0, 1, 1]
[0, 0, 1, 1, 1, 1, 0]
```

Copyright 2016 Jeff Abrahamson

```
Le chat est orange.
  [1, 1, 0, 0, 0, 1, 1]
  [0, 0, 1, 1, 1, 1, 0]
Le chien court vite.
```

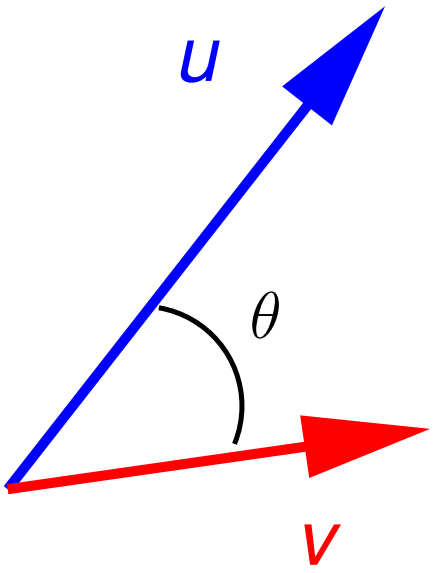
Copyright 2016 Jeff Abrahamson

Cosine Similarity



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

Copyright 2016 Jeff Abrahamson



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

(if u and v have norm 1)

Copyright 2016 Jeff Abrahamson

Le chat est orange.

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

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

Le chien court vite.

Copyright 2016 Jeff Abrahamson

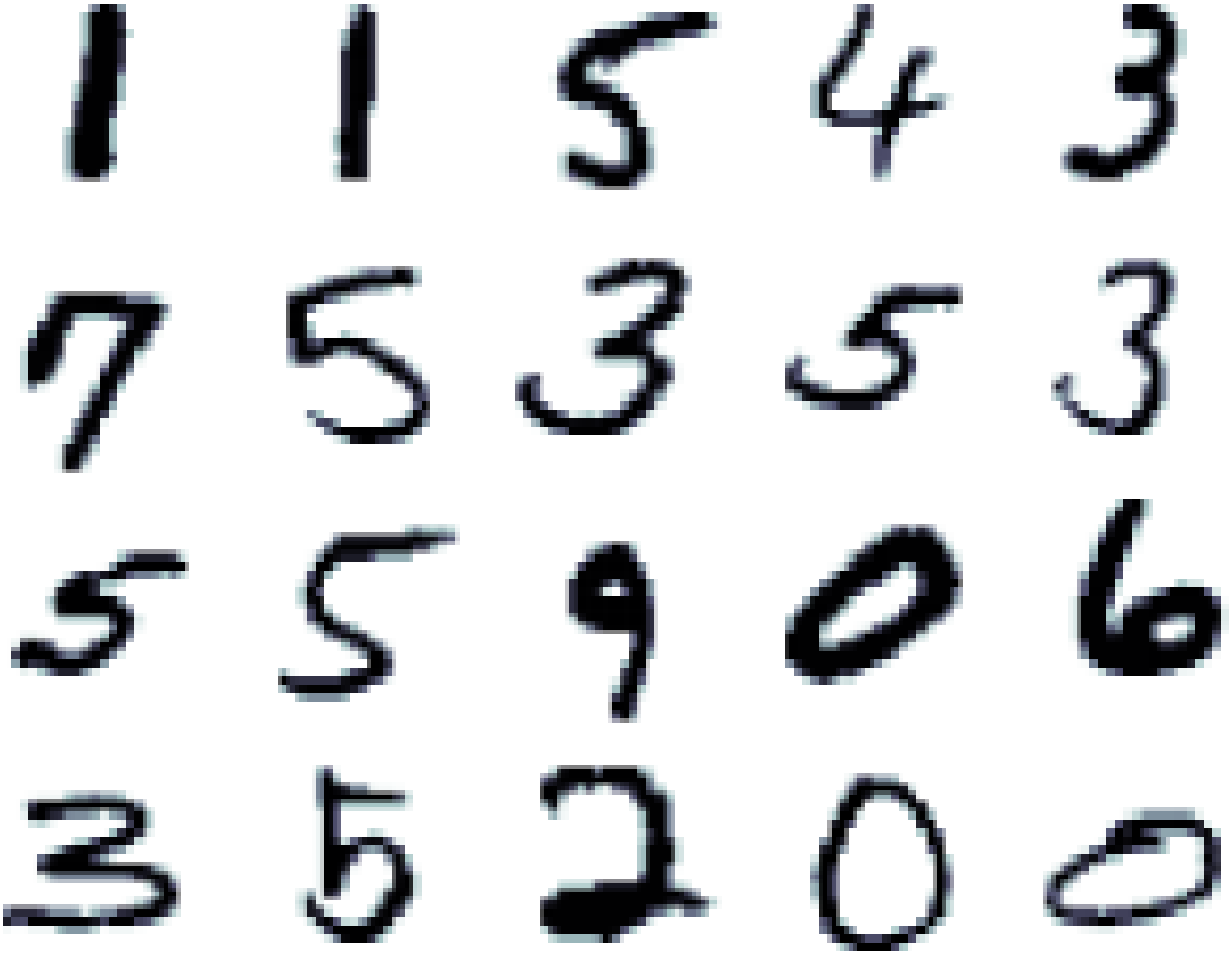
$u = [1,\ 1,\ 0,\ 0,\ 0,\ 1,\ 1]$

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

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

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

Copyright 2016 Jeff Abrahamson



Copyright 2016 Jeff Abrahamson

Un exemple (trop) vite

Copyright 2016 Jeff Abrahamson

Linear neuron

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

Copyright 2016 Jeff Abrahamson

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

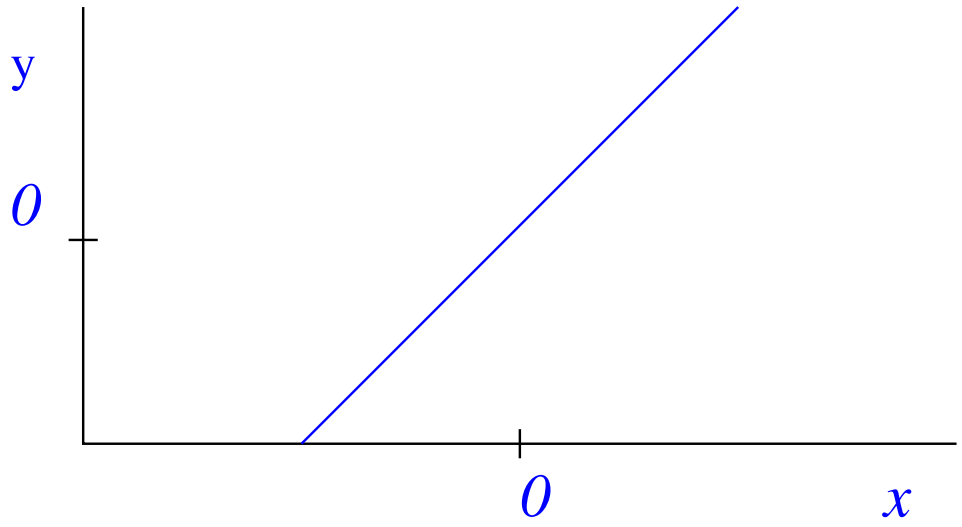
where

$$y = \text{output}$$
$$b = \text{bias}$$
$$x_i = i^{\text{th}} \text{ input}$$
$$w_i = \text{weight on } i^{\text{th}} \text{ input}$$

Copyright 2016 Jeff Abrahamson

Linear neuron

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



Copyright 2016 Jeff Abrahamson

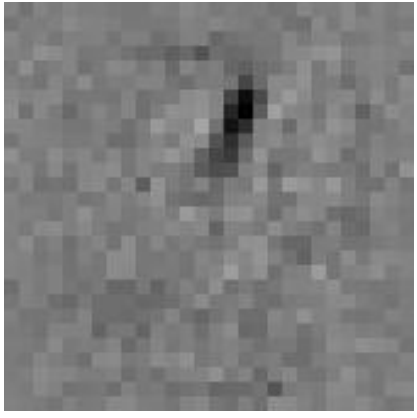
- Input neurons: pixels
- Output neurons: classes (digits)
- Connect them all! *(bipartite)*

Copyright 2016 Jeff Abrahamson

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

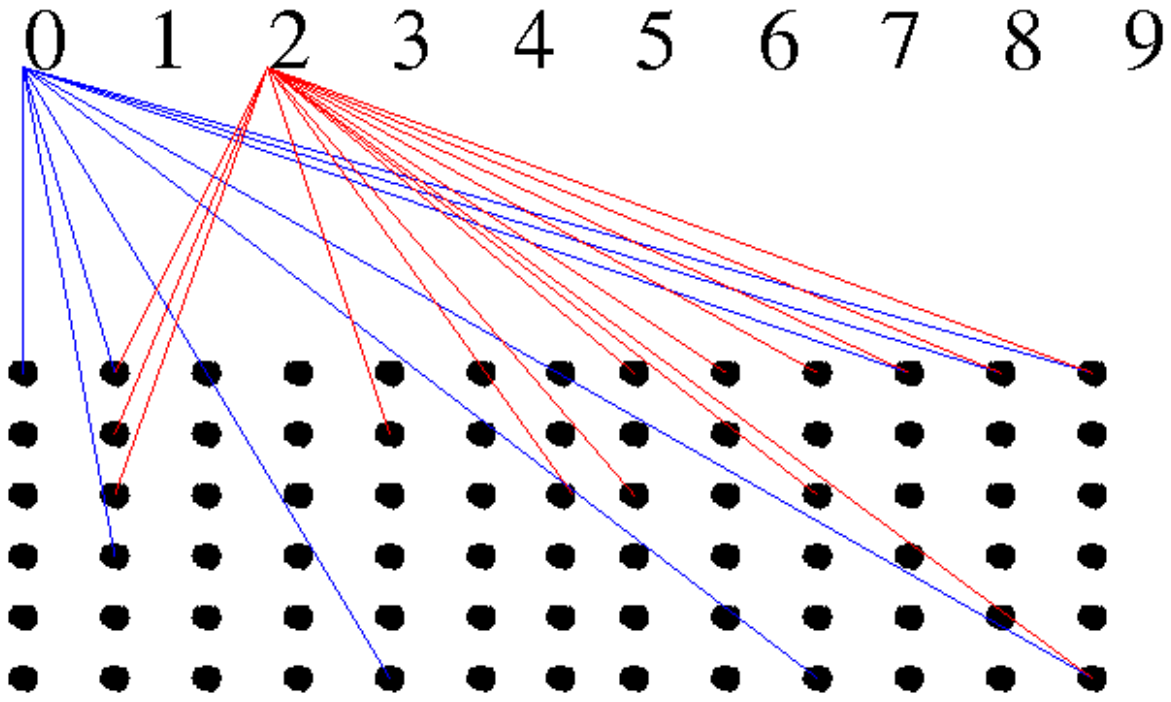
Copyright 2016 Jeff Abrahamson

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



Copyright 2016 Jeff Abrahamson

Example: handwriting recognition of digits



Copyright 2016 Jeff Abrahamson

To train this ANN:

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

Copyright 2016 Jeff Abrahamson

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.

Copyright 2016 Jeff Abrahamson

stratégie 1 : sac de mot

Copyright 2016 Jeff Abrahamson

stratégie 1 : tf-idf

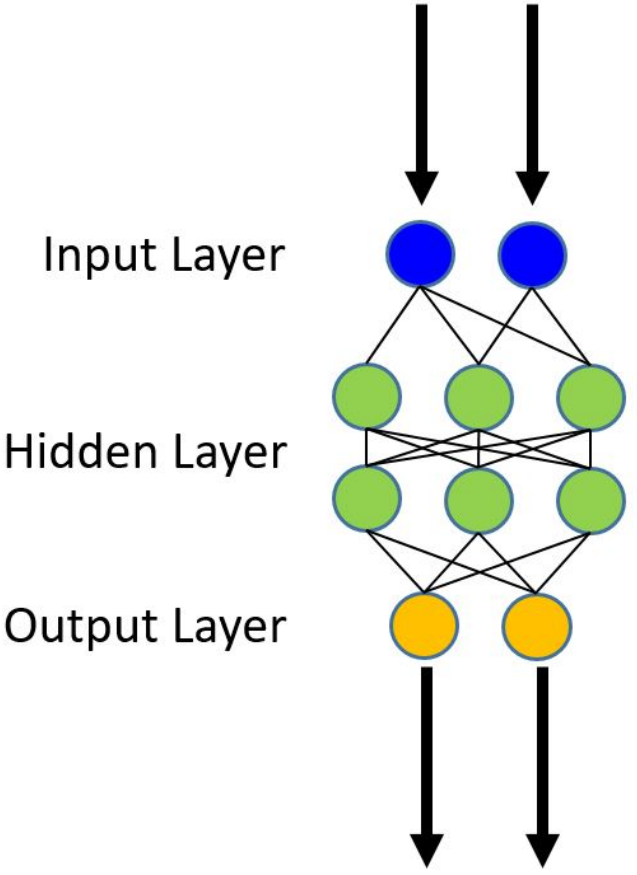
Copyright 2016 Jeff Abrahamson

stratégie 2 : word2vec

Copyright 2016 Jeff Abrahamson

deux stratégies : trouver une séquence

Copyright 2016 Jeff Abrahamson



Copyright 2016 Jeff Abrahamson

stratégie 1 : sac de mot

Copyright 2016 Jeff Abrahamson

stratégie 1 : tf-idf

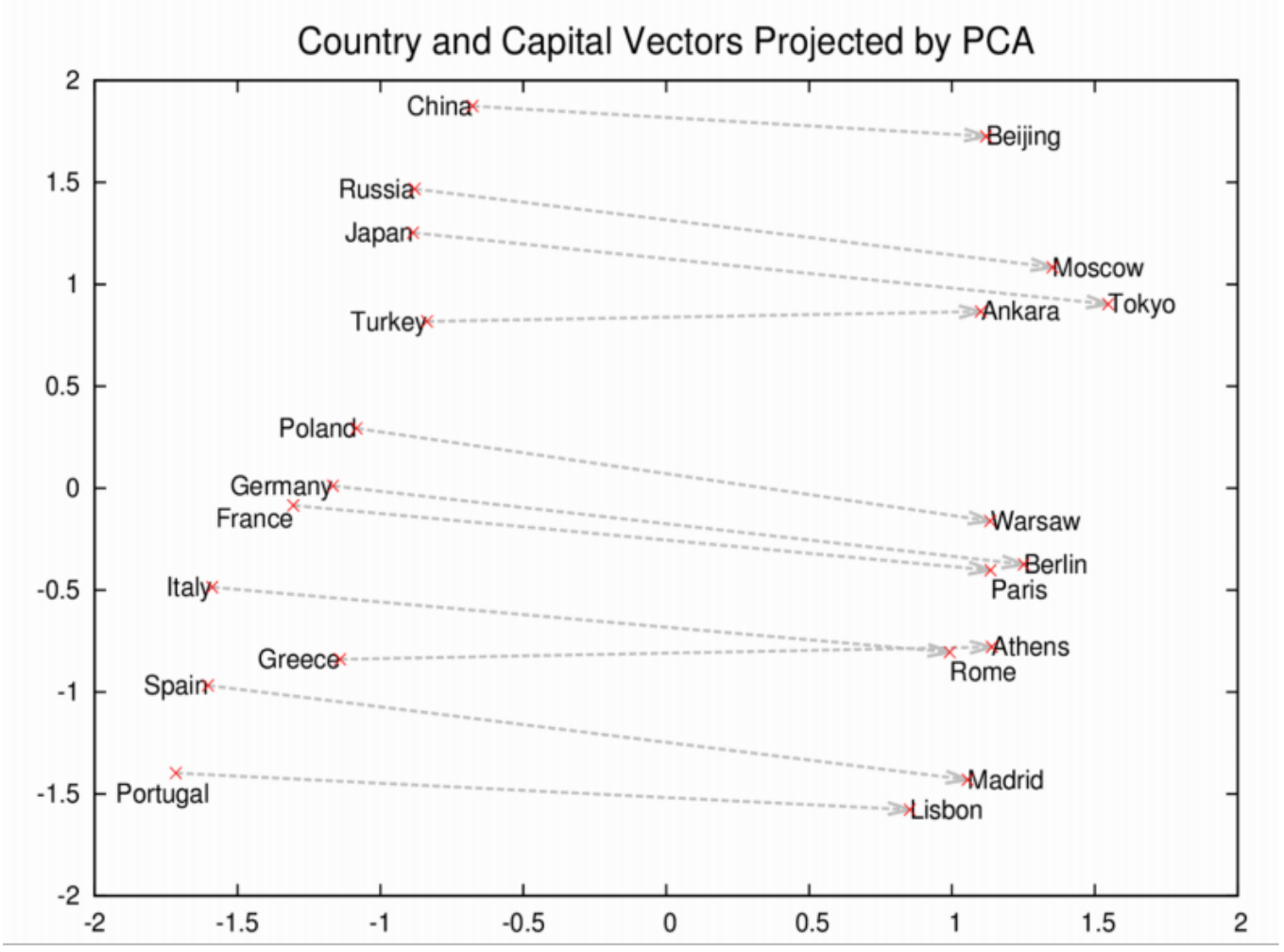
Copyright 2016 Jeff Abrahamson

stratégie 2 : word2vec

Copyright 2016 Jeff Abrahamson

deux stratégies : trouver une séquence

Copyright 2016 Jeff Abrahamson



Copyright 2016 Jeff Abrahamson

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

Similar to ‘dress’

Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson

Copyright 2016 Jeff Abrahamson

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

Copyright 2016 Jeff Abrahamson

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

Copyright 2016 Jeff Abrahamson

Resources

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

Copyright 2016 Jeff Abrahamson

Questions?

Copyright 2016 Jeff Abrahamson