Prédire aussi bien que les meilleurs
(Et en plus en faisant des maths!)

Pierre Alquier

École nationale de la statistique et de l'administration économique

Les mercredis mathématiques du CIRM
28 novembre 2018


Orange
Fly
Emirates

Ligue 1 · 02/12, 16:00

OM — contre — Stade de Reims

COMPOSITION | STATISTIQUES | ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%

Ligue 1 · 02/12, 16:00

OM	contre	Stade de Reims

COMPOSITION · **STATISTIQUES** · ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%


Ligue 1 · 02/12, 16:00
OM
contre
Stade de Reims
COMPOSITION
STATISTIQUES
ACTU
PROBABILITÉ DE VICTOIRE
OM
67%
Match nul
21%
Stade de Reims
12%

$2 - 1$


Ligue 1 · 02/12, 16:00
OM
contre
Stade de Reims
COMPOSITION
STATISTIQUES
ACTU
PROBABILITÉ DE VICTOIRE
OM
67%
Match nul
21%
Stade de Reims
12%

$2 - 1$

Ligue 1 · 02/12, 16:00

OM contre Stade de Reims

COMPOSITION STATISTIQUES ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%

$2 - 1$ $1 - 1$

Ligue 1 · 02/12, 16:00

OM contre Stade de Reims

COMPOSITION STATISTIQUES ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%

$2-1$ $1-1$

$2-1$ $\quad$ $1-1$ $\quad$ $2-0$

Ligue 1 · 02/12, 16:00

OM contre Stade de Reims

COMPOSITION | STATISTIQUES | ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%




$2-1$ $1-1$ $2-0$

Ligue 1 · 02/12, 16:00

OM contre Stade de Reims

COMPOSITION | STATISTIQUES | ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%

$2 - 1$

$1 - 1$

$2 - 0$

$3 - 0$

Ligue 1 · 02/12, 16:00

OM contre Stade de Reims

COMPOSITION STATISTIQUES ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%

2 – 1

1 – 1

2 – 0

3 – 0


Ligue 1 · 02/12, 16:00
OM
contre
Stade de Reims
COMPOSITION
STATISTIQUES
ACTU
PROBABILITÉ DE VICTOIRE
OM
67%
Match nul
21%
Stade de Reims
12%





2 – 1	1 – 1	2 – 0	3 – 0	0 – 1

Ligue 1 · 02/12, 16:00

OM contre Stade de Reims

COMPOSITION | STATISTIQUES | ACTU

PROBABILITÉ DE VICTOIRE

OM	Match nul	Stade de Reims
67%	21%	12%





V	ND	V	V	ND

OBJECTIF

Apprendre à prédire aussi bien que le meilleur pronostiqueur.

OBJECTIF

Apprendre à prédire aussi bien que le meilleur pronostiqueur.
De façon générale : on parle plutôt d'expert.

OBJECTIF

Apprendre à prédire aussi bien que le meilleur pronostiqueur.
De façon générale : on parle plutôt d'expert.

1. En général, c'est assez difficile (on verra à la fin).

OBJECTIF

Apprendre à prédire aussi bien que le meilleur pronostiqueur. De façon générale : on parle plutôt d'expert.

1. En général, c'est assez difficile (on verra à la fin).
2. Le problème devient plus facile si il existe un expert qui ne se trompe jamais.

OBJECTIF

Apprendre à prédire aussi bien que le meilleur pronostiqueur.
De façon générale : on parle plutôt d'expert.

1. En général, c'est assez difficile (on verra à la fin).
2. Le problème devient plus facile si il existe un expert qui ne se trompe jamais.

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.






Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

$t = 1$	V	ND	V	V	ND	V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	<u>ND</u>	V	V	ND	V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V

Algorithme consistent

Suivre un expert qui ne s'est jamais trompé.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V
$t = 5$						V

Théorème

En considérant n experts, dont l'un ne se trompe jamais, en utilisant l'algorithme *consistent*, je commetrai au plus $n-1$ erreurs.

Théorème

En considérant *n* experts, dont l'un ne se trompe jamais, en utilisant l'algorithme *consistent*, je commetrai au plus $n - 1$ erreurs.

$$\mathbf{Err}_t(\mathrm{consistent}) \leq \mathrm{n} - 1.$$

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.






Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t=1$	V	ND	V	V	ND	V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	$\underline{\text{V}}$	ND	$\underline{\text{V}}$	$\underline{\text{V}}$	ND	$\underline{\text{V}}$

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	V	V	ND	V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	V	V	ND	V
$t = 2$	ND		ND	V		ND
$t = 3$	V		ND			V
$t = 4$	ND					V
$t = 5$						V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	ND	ND	ND	V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	<u>ND</u>	<u>ND</u>	<u>ND</u>	<u>ND</u>	V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	ND	ND	ND	V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND
$t = 3$	ND					V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND
$t = 3$	ND					V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND
$t = 3$	ND					V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND
$t = 3$	ND					V
$t = 4$						V

Algorithme halving

Suivre la majorité des experts qui ne se sont pas trompés.

						
$t = 1$	V	ND	ND	ND	ND	V
$t = 2$	ND					ND
$t = 3$	ND					V
$t = 4$						V
$t = 5$						V

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5 \xrightarrow{6} 1.25$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5 \xrightarrow{6} 1.25 \xrightarrow{7} \color{red}{0.625}$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5 \xrightarrow{6} 1.25 \xrightarrow{7} {\color{red}0.625}$$

Donc *halving* commet au plus 6 erreurs si $n = 80$.

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5 \xrightarrow{6} 1.25 \xrightarrow{7} 0.625$$

Donc *halving* commet au plus 6 erreurs si $n = 80$.

Le nombre d'erreurs E vérifie

$$1 \leq \frac{n}{2^E} = \frac{n}{2 \times \cdots \times 2}.$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5 \xrightarrow{6} 1.25 \xrightarrow{7} 0.625$$

Donc *halving* commet au plus 6 erreurs si $n = 80$.

Le nombre d'erreurs E vérifie

$$1 \leq \frac{n}{2^E} = \frac{n}{2 \times \cdots \times 2}.$$

Autrement dit

$$2^E \leq n.$$

Donc : le nombre d'erreurs commises par *halving* est au plus le nombre de fois que je peux diviser n par deux en restant ≥ 1...

Exemple :

$$n = 80 \xrightarrow{1} 40 \xrightarrow{2} 20 \xrightarrow{3} 10 \xrightarrow{4} 5 \xrightarrow{5} 2.5 \xrightarrow{6} 1.25 \xrightarrow{7} \color{red}{0.625}$$

Donc *halving* commet au plus 6 erreurs si $n = 80$.

Le nombre d'erreurs E vérifie

$$1 \leq \frac{n}{2^E} = \frac{n}{2 \times \cdots \times 2}.$$

Autrement dit

$$2^E \leq n.$$

$$E \leq \log_2(n).$$

Théorème

En considérant n experts, dont l'un ne se trompe jamais, en utilisant l'algorithme *halving*, je commetrai au plus $\log_2(n)$ erreurs.

Théorème

En considérant n experts, dont l'un ne se trompe jamais, en utilisant l'algorithme *halving*, je commetrai au plus $\log_2(n)$ erreurs.

$$\mathbf{Err}_t(\mathrm{halving}) \leq \log_2(\mathrm{n}).$$

Théorème

En considérant n experts, dont l'un ne se trompe jamais, en utilisant l'algorithme *halving*, je commetrai au plus $\log_2(n)$ erreurs.

$$\mathbf{Err}_t(\text{halving}) \leq \log_2(\text{n}).$$

Exemple : si on suit $n = 1247$ experts,

$$\mathbf{Err}_t(\text{consistent}) \leq n - 1 = 1246$$

$$\mathbf{Err}_t(\text{halving}) \leq \log_2(n) \simeq 10.28$$

Théorème

En considérant n experts, dont l'un ne se trompe jamais, en utilisant l'algorithme *halving*, je commetrai au plus $\log_2(n)$ erreurs.

$$\mathbf{Err}_t(\text{halving}) \leq \log_2(\text{n}).$$

Exemple : si on suit $n = 1247$ experts,

$$\mathbf{Err}_t(\text{consistent}) \leq n - 1 = 1246$$

$$\begin{aligned} & \mathbf{Err}_t(\text{halving}) \leq \log_2(n) \simeq 10.28 \\ \Rightarrow\ & \mathbf{Err}_t(\text{halving}) \leq 10 \end{aligned}$$





Le gouvernement vous ment !

Le gouvernement vous ment !

Le Père Noël n'existe pas !

Le gouvernement vous ment !

Le Père Noël n'existe pas !

halving : nombre de “voix” données à chaque expert.

1	1	1	1	1

halving : nombre de “voix” données à chaque expert.





0 1 0 0 1

halving : nombre de “voix” données à chaque expert.

0	1	0	0	1
	1			1

halving : nombre de “voix” données à chaque expert.

0	1	0	0	1
	1			0

halving : nombre de “voix” données à chaque expert.

				
0	1	0	0	1
	1			0
	1			

halving : nombre de “voix” données à chaque expert.

0	1	0	0	1
	1			0
	0			

mult.upt. : ne jamais éliminer complètement un expert.





1	1	1	1	1

mult.upt. : ne jamais éliminer complètement un expert.

$\frac{1}{2}$

1

$\frac{1}{2}$

$\frac{1}{2}$

1

mult.upt. : ne jamais éliminer complètement un expert.

				
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1

mult.upt. : ne jamais éliminer complètement un expert.

$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$

mult.upt. : ne jamais éliminer complètement un expert.

$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$

mult.upt. : ne jamais éliminer complètement un expert.

$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{4}$

Théorème

$$\mathbf{Err}_t(\mathrm{EWA}) \leq \mathbf{Err}_{\mathrm{t}}(\text{meilleur expert}) + \mathrm{Regret}(\mathrm{t})$$

et

$$\frac{\mathrm{Regret}(t)}{t} \xrightarrow[t\to\infty]{} 0.$$

Une version un peu plus générale et formalisée du résultat, extraite de la thèse de Sébastien Gerchinovitz.

Une version un peu plus générale et formalisée du résultat, extraite de la thèse de Sébastien Gerchinovitz.

Lemma 1. *Assume that for some known constant $B_y > 0$,*

$$y_1, \ldots, y_T \in [-B_y, B_y] \, .$$

For all $\tau > 0$, if the algorithm $\mathrm{SeqSEW}_\tau^{B,\eta}$ is used with $B \geqslant B_y$ and $\eta \leqslant 1/(8B^2)$, then it satisfies

$$\sum_{t=1}^{T} (y_t - \widehat{y}_t)^2 \leqslant \inf_{\rho \in \mathcal{M}_1^+(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \sum_{t=1}^{T} \Big(y_t - \big[\boldsymbol{u} \cdot \boldsymbol{\varphi}(x_t) \big]_B \Big)^2 \rho(\mathrm{d}\boldsymbol{u}) + \frac{\mathcal{K}(\rho, \pi_\tau)}{\eta} \right\} \tag{14}$$

$$\leqslant \inf_{\rho \in \mathcal{M}_1^+(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \sum_{t=1}^{T} \big(y_t - \boldsymbol{u} \cdot \boldsymbol{\varphi}(x_t) \big)^2 \rho(\mathrm{d}\boldsymbol{u}) + \frac{\mathcal{K}(\rho, \pi_\tau)}{\eta} \right\} . \tag{15}$$

Proof (of Lemma 1). As is usually done in the online learning setting for the study of the Exponentially Weighted Average Forecaster, our proof relies on the control of $\sum_t \eta^{-1} \ln(W_{t+1}/W_t)$ where we recall that $W_1 \triangleq 1$ and, for all $t \geqslant 2$,

$$W_t \triangleq \int_{\mathbb{R}^d} \exp\left(-\eta \sum_{s=1}^{t-1} \Big(y_s - \big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_s)\big]_B\Big)^2\right) \pi_\tau(\mathrm{d}\boldsymbol{u}) \,.$$

On the one hand, we have

$$\begin{aligned}
\frac{1}{\eta}\ln\frac{W_{T+1}}{W_1} &= \frac{1}{\eta}\ln\int_{\mathbb{R}^d}\exp\left(-\eta\sum_{t=1}^{T}\Big(y_s - \big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_s)\big]_B\Big)^2\right)\pi_\tau(\mathrm{d}\boldsymbol{u}) - \frac{1}{\eta}\ln 1 \\
&= \frac{1}{\eta}\sup_{\rho\in\mathcal{M}_1^+(\mathbb{R}^d)}\left\{\int_{\mathbb{R}^d}\left(-\eta\sum_{t=1}^{T}\Big(y_t - \big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\big]_B\Big)^2\right)\rho(\mathrm{d}\boldsymbol{u}) - \mathcal{K}(\rho,\pi_\tau)\right\} - 0 \qquad (16)\\
&= -\inf_{\rho\in\mathcal{M}_1^+(\mathbb{R}^d)}\left\{\int_{\mathbb{R}^d}\sum_{t=1}^{T}\Big(y_t - \big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\big]_B\Big)^2\rho(\mathrm{d}\boldsymbol{u}) + \frac{\mathcal{K}(\rho,\pi_\tau)}{\eta}\right\}, \qquad (17)
\end{aligned}$$

where (16) follows from a convex duality argument for the Kullback-Leibler divergence (cf., e.g., [DZ98, p. 264] or [Cat04, p. 159]) which we recall in Proposition 5 in the Appendix.

On the other hand, we can rewrite W_{T+1}/W_1 as a telescopic product and get

$$\frac{1}{\eta}\ln\frac{W_{T+1}}{W_1} = \sum_{t=1}^{T}\frac{1}{\eta}\ln\frac{W_{t+1}}{W_t}$$

$$= \sum_{t=1}^{T}\frac{1}{\eta}\ln\int_{\mathbb{R}^d}\frac{\exp\left(-\eta\Big(y_t-\big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\big]_B\Big)^2\right)\exp\left(-\eta\sum_{s=1}^{t-1}\Big(y_s-\big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_s)\big]_B\Big)^2\right)}{W_t}\,\pi_\tau(\mathrm{d}\boldsymbol{u})$$

$$= \sum_{t=1}^{T}\frac{1}{\eta}\ln\int_{\mathbb{R}^d}\exp\left(-\eta\Big(y_t-\big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\big]_B\Big)^2\right)p_t(\mathrm{d}\boldsymbol{u})\,, \tag{18}$$

where (18) follows from the definition of p_t.

Let $t\in\{1,\ldots,T\}$. First note that by assumption $y_t\in[-B_y,B_y]\subset[-B,B]$ so that both y_t and $\big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\big]_B$ are $[-B,B]$-valued for all $\boldsymbol{u}\in\mathbb{R}^d$. Moreover, from Proposition 6 in the Appendix, the square loss is $1/(8B^2)$-exp-concave on $[-B,B]$ and thus η-exp-concave (since $\eta\leqslant 1/(8B^2)$ by assumption). Therefore, by Jensen's inequality,

$$\int_{\mathbb{R}^d} e^{-\eta\left(y_t-\left[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\right]_B\right)^2}p_t(\mathrm{d}\boldsymbol{u})\leqslant\exp\left(-\eta\left(y_t-\int_{\mathbb{R}^d}\big[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\big]_B\,p_t(\mathrm{d}\boldsymbol{u})\right)^2\right)\,.$$

Taking the logarithms of both sides of the inequality yields

$$\begin{aligned} \ln \int_{\mathbb{R}^d} e^{-\eta\left(y_t - \left[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\right]_B\right)^2} p_t(\mathrm{d}\boldsymbol{u}) &\leqslant -\eta\left(y_t - \int_{\mathbb{R}^d} \left[\boldsymbol{u}\cdot\boldsymbol{\varphi}(x_t)\right]_B p_t(\mathrm{d}\boldsymbol{u})\right)^2 \\ &= -\eta(y_t - \widehat{y}_t)^2 \,. \end{aligned} \tag{19}$$

Dividing the latter inequality by η, summing over $t \in \{1, \dots, T\}$ and combining with Equation (18), we get

$$\frac{1}{\eta} \ln \frac{W_{T+1}}{W_1} \leqslant -\sum_{t=1}^{T} (y_t - \widehat{y}_t)^2 \,. \tag{20}$$

Putting Inequalities (17) and (20) together, we get Inequality (14) of Lemma 1. As for Inequality (15), it follows from (14) by noting that

$$\forall y \in [-B, B], \quad \forall x \in \mathbb{R}, \qquad \left|y - [x]_B\right| \leqslant |y - x| \,,$$

so that truncation to $[-B, B]$ can only improve prediction under the square loss if the observations are $[-B, B]$-valued, which is the case here since by assumption $y_t \in [-B_y, B_y] \subset [-B, B]$ for all $t = 1, \dots, T$. □


8-2=

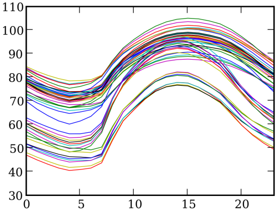

POURQUOI,
POURQUOI,
POURQUOI ?


8-2 =

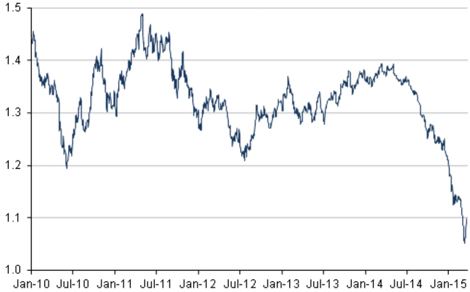

POURQUOI,
POURQUOI,
POURQUOI ?


8-2 =


PUTAIN,MAIS POURQUOI
TU FAIS ÇA ?



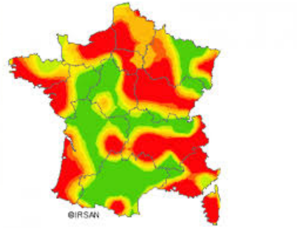
Prédiction de la qualité de l'air par différents experts, utilisant différents modèles physiques.



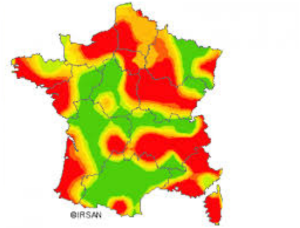
Cours du change Euro/Dollar.

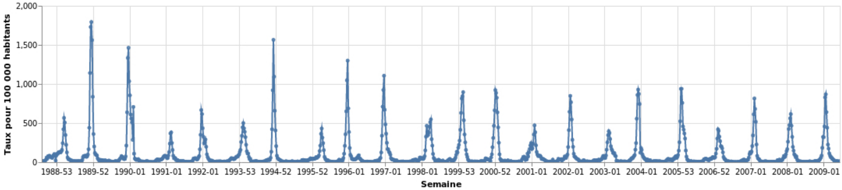
Ce qu'il faut savoir sur
la grippe saisonnière
La grippe est une maladie causée par les virus grippaux saisonniers.
Ces virus se propagent d'une personne à l'autre.
Comment reconnaître la grippe?
Fièvre élevée soudaine
Mal de tête
Toux ou mal de gorge
Douleurs musculaires
Que faire quand vous avez la grippe?
Couvrez votre bouche quand vous toussez, ou éternuez dans votre bras / coude ou dans un mouchoir
Lavez-vous les mains fréquemment
Reposez-vous
Buvez beaucoup d'eau et mangez des aliments nutritifs
Consultez un médecin, si vous ne vous améliorez pas, ou si vous faites partie d'un groupe à risque élevé
Organisation mondiale de la Santé
Programme D'URGENCES SANITAIRES
Comment prévenir la gripppe?
Se faire vacciner contre la grippe chaque année est le meilleur moyen de prévenir la grippe
La vaccination est particulièrement importante pour les personnes à haut risque de complications secondaires suite à une grippe :
femmes enceintes
les personnes de plus de 65 ans
les enfants de 6 mois à 5 ans
les personnes souffrant de problèmes de santé chroniques
les personnes qui vivent avec, ou prennent soin des personnes à haut risque


Ce qu'il faut savoir sur
la grippe saisonnière
La grippe est une maladie causée par les virus grippaux saisonniers.
Ces virus se propagent d'une personne à l'autre.
Comment reconnaître la grippe?
Fièvre élevée soudaine
Mal de tête
Toux ou mal de gorge
Douleurs musculaires
Que faire quand vous avez la grippe?
Couvrez votre bouche quand vous toussez, ou éternuez dans votre bras / coude ou dans un mouchoir
Lavez-vous les mains fréquemment
Reposez-vous
Buvez beaucoup d'eau et mangez des aliments nutritifs
Consultez un médecin, si vous ne vous améliorez pas, ou si vous faites partie d'un groupe à risque élevé
Organisation mondiale de la Santé
Programme D'URGENCES SANITAIRES
Comment prévenir la grippe?
Se faire vacciner contre la grippe chaque année est le meilleur moyen de prévenir la grippe
La vaccination est particulièrement importante pour les personnes à haut risque de complications secondaires suite à une grippe :
femmes enceintes
les personnes de plus de 65 ans
les enfants de 6 mois à 5 ans
les personnes souffrant de problèmes de santé chroniques
les personnes qui vivent avec, ou prennent soin des personnes à haut risque


©IRSAN


Ce qu'il faut savoir sur
la grippe saisonnière
La grippe est une maladie causée par les virus grippaux saisonniers.
Ces virus se propagent d'une personne à l'autre.
Comment reconnaître la grippe?
Fièvre élevée soudaine
Mal de tête
Toux ou mal de gorge
Douleurs musculaires
Que faire quand vous avez la grippe?
Couvrez votre bouche quand vous toussez, ou éternuez dans votre bras / coude ou dans un mouchoir
Lavez-vous les mains fréquemment
Reposez-vous
Buvez beaucoup d'eau et mangez des aliments nutritifs
Consultez un médecin, si vous ne vous améliorez pas, ou si vous faites partie d'un groupe à risque élevé
Organisation mondiale de la Santé
Programme D'URGENCES SANITAIRES
Comment prévenir la grippe?
Se faire vacciner contre la grippe chaque année est le meilleur moyen de prévenir la grippe
La vaccination est particulièrement importante pour les personnes à haut risque de complications secondaires suite à une grippe :
femmes enceintes
les personnes de plus de 65 ans
les enfants de 6 mois à 5 ans
les personnes souffrant de problèmes de santé chroniques
les personnes qui vivent avec, ou prennent soin des personnes à haut risque


©IRSAN


Taux pour 100 000 habitants
2,000
1,500
1,000
500
0
1988-53 1989-52 1990-01 1991-01 1992-01 1993-53 1994-52 1995-52 1996-01 1997-01 1998-01 1999-53 2000-52 2001-01 2002-01 2003-01 2004-01 2005-53 2006-52 2007-01 2008-01 2009-01
Semaine

CONCLUSION

FAITES DES MATHS
FAITES DE L'INFORMATIQUE
Vous en aurez besoin PARTOUT

CONCLUSION

FAITES DES MATHS

FAITES DE L'INFORMATIQUE

Vous en aurez besoin PARTOUT

References :

1. L'étude de l'algorithme halving est tirée de [1].
2. L'exemple sur la pollution de l'air est tiré de [2].
3. L'étude de l'algorithme mult.up est copié-collée de [3].
4. Les cartes et séries épidémiologiques sont tirées du site *https ://sites.sentiweb.fr/*.

[1] S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 2012.

[2] G. Stoltz. Agrégation séquentielle de prédicteurs : méthodologie générale et applications à la prévision de la qualité de l'air et à celle de la consommation électrique. *Journal de la SFDS*, 2010.

[3] S. Gerchinovitz. Prédiction de suites individuelles et cadre statistique classique : étude de quelques liens autour de la régression parcimonieuse et des techniques d'agrégation. *Thèse de doctorat, Université Paris Sud*, 2011.