Multi-agent learning

Prediction, postdiction, and calibration

Gerard Vreeswijk, Intelligent Systems Group, Computer Science Department, Faculty of Sciences, Utrecht University, The Netherlands.

Friday 6$^{\text{th}}$ October, 2017

Terminology

Terminology

1.

Terminology

1. Outcome space Z. (Finite.)

Terminology

1. Outcome space Z. (Finite.)
2.

Terminology

1. Outcome space Z. (Finite.)
2. A discrete-time stochastic process $(z^t)_{t \in \mathbb{N}}$ with values in Z.

Terminology

1. Outcome space Z. (Finite.)
2. A discrete-time stochastic process $(z^t)_{t \in \mathbb{N}}$ with values in Z.
3.

Terminology

1. Outcome space Z. (Finite.)
2. A discrete-time stochastic process $(z^t)_{t \in \mathbb{N}}$ with values in Z.
3. History: $h^t = (z^1, z^2, \dots, z^t)$. Set of all histories of length t: H_t. Set of all histories:

$$H =_{Def} \cup_{t \in \mathbb{N}} H_t.$$

Terminology

1. Outcome space Z. (Finite.)
2. A discrete-time stochastic process $(z^t)_{t\in\mathbb{N}}$ with values in Z.
3. History: $h^t = (z^1, z^2, \dots, z^t)$. Set of all histories of length t: H_t. Set of all histories:
$$H =_{Def} \cup_{t\in\mathbb{N}} H_t.$$
4. Realisation: $\omega = (z^1, z^2, \dots)$. Set of all realisations: Ω.

Terminology

1. Outcome space Z. (Finite.)
2. A discrete-time stochastic process $(z^t)_{t\in\mathbb{N}}$ with values in Z.
3. History: $h^t = (z^1, z^2, \dots, z^t)$. Set of all histories of length t: H_t. Set of all histories:
$$H =_{Def} \cup_{t\in\mathbb{N}} H_t.$$
4. Realisation: $\omega = (z^1, z^2, \dots)$. Set of all realisations: Ω.
5. Actual probability of realisation (by nature): $q : H \to \Delta(Z)$:
$$q(z \mid h) = \text{the actual probability that } z \text{ occurs after } h$$

Terminology

1. Outcome space Z. (Finite.)
2. A discrete-time stochastic process $(z^t)_{t \in \mathbb{N}}$ with values in Z.
3. History: $h^t = (z^1, z^2, \dots, z^t)$. Set of all histories of length t: H_t. Set of all histories:
$$H =_{Def} \cup_{t \in \mathbb{N}} H_t.$$
4. Realisation: $\omega = (z^1, z^2, \dots)$. Set of all realisations: Ω.
5. Actual probability of realisation (by nature): $q : H \to \Delta(Z)$:
$$q(z \mid h) = \text{the actual probability that } z \text{ occurs after } h$$
6. Forecast rule (by agent): $p : H \to \Delta(Z)$:
$$p(z \mid h) = \text{the predicted probability that } z \text{ occurs after } h$$

Being a good predictor

A realisation ω generates

Being a good predictor

A realisation ω generates

1. A sequence of realisation probabilities $q_1, q_2, \dots$

Being a good predictor

A realisation ω generates

1. A sequence of realisation probabilities $q_1, q_2, \ldots$
2. A sequence of predictions $p_1, p_2, \ldots$

Being a good predictor

A realisation ω generates

1. A sequence of realisation probabilities $q_1, q_2, \dots$
2. A sequence of predictions $p_1, p_2, \dots$

Definition (Good predictor).

$$\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} \|p^i - q^i\|^2 = 0 \text{ almost surely}$$

Being a good predictor

A realisation ω generates

1. A sequence of realisation probabilities $q_1, q_2, \ldots$
2. A sequence of predictions $p_1, p_2, \ldots$

Definition (Good predictor).

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \|p^i - q^i\|^2 = 0 \text{ almost surely}$$

For example, if $|Z| = 3$ then $\|p^i - q^i\|^2$ might be

$$\left\| \begin{pmatrix} 0.2 \\ 0.1 \\ 0.7 \end{pmatrix} - \begin{pmatrix} 0.3 \\ 0.1 \\ 1.0 \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} 0.1 \\ 0.0 \\ 0.3 \end{pmatrix} \right\|^2 = 0.1^2 + 0.0^2 + 0.3^2 = 0.01 + 0.09 = 0.1$$

Being a good predictor

Discussion:

Being a good predictor

Discussion:

- Requiring that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \|p^i - q^i\|^2 = 0 \text{ almost surely}$$

is too forgiving.

Being a good predictor

Discussion:

- Requiring that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \|p^i - q^i\|^2 = 0 \text{ almost surely}$$

  is too forgiving.

  This is because a low average does not exclude large intervals of forecasting errors.

Being a good predictor

Discussion:

- Requiring that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \|p^i - q^i\|^2 = 0 \text{ almost surely}$$

  is too forgiving.

  This is because a low average does not exclude large intervals of forecasting errors.

- Having

$$\lim_{t\to\infty} \|p^t - q^t\| = 0 \text{ almost surely}$$

  is desirable. Later, it turns out this requirement *is too demanding*.

Calibration score

Calibration score

- Let $\omega = (z_1, z_2, \dots)$ be a realisation of events.

Calibration score

- Let $\omega = (z_1, z_2, \dots)$ be a realisation of events.
- Define

$$n^t(p) =_{Def} \text{the number of times that } p \text{ is forecast}$$

Calibration score

- Let $\omega = (z_1, z_2, \dots)$ be a realisation of events.
- Define

$$n^t(p) =_{Def} \text{the number of times that } p \text{ is forecast}$$

- Define

$$\phi^t(p) =_{Def} \text{the empirical distribution when } p \text{ was forecast}$$

Calibration score

- Let $\omega = (z_1, z_2, \dots)$ be a realisation of events.
- Define

$$n^t(p) =_{Def} \text{the number of times that } p \text{ is forecast}$$

- Define

$$\phi^t(p) =_{Def} \text{the empirical distribution when } p \text{ was forecast}$$

In both definitions the quantities $n^t(p)$ and $\phi^t(p)$ depend on ω but in the notation this is suppressed.

Calibration score

- Let $\omega = (z_1, z_2, \dots)$ be a realisation of events.
- Define

$$n^t(p) =_{Def} \text{the number of times that } p \text{ is forecast}$$

- Define

$$\phi^t(p) =_{Def} \text{the empirical distribution when } p \text{ was forecast}$$

  In both definitions the quantities $n^t(p)$ and $\phi^t(p)$ depend on ω but in the notation this is suppressed.

- Define

$$C^t(\omega) =_{Def} \sum_{p \in \Delta(Z)} \frac{n^t(p)}{t} \|\phi^t(p) - p\|^2$$

Calibration

Definition (Calibration).

Calibration

Definition (Calibration).

- Let $\epsilon > 0$. The forecasts $p_1, p_2, \dots$ are said to be calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) = 0.$$

Calibration

Definition (Calibration).

- Let $\epsilon > 0$. The forecasts $p_1, p_2, \ldots$ are said to be calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) = 0.$$

and ϵ-calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) \leq \epsilon.$$

Calibration

Definition (Calibration).

- Let $\epsilon > 0$. The forecasts $p_1, p_2, \dots$ are said to be calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) = 0.$$

  and ϵ-calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) \leq \epsilon.$$

- A forecast rule f is said to be calibrated (ϵ-calibrated) if the above equality (inequality) holds for all realisations.

Calibration

Definition (Calibration).

- Let $\epsilon > 0$. The forecasts $p_1, p_2, \ldots$ are said to be calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) = 0.$$

  and ϵ-calibrated on ω if

$$\lim_{t\to\infty} C^t(\omega) \leq \epsilon.$$

- A forecast rule f is said to be calibrated (ϵ-calibrated) if the above equality (inequality) holds for all realisations.

Pessimistic result (Oakes, 1985): *For every forecasting rule f there exists a realisation ω and an $\epsilon > 0$ such that f does not ϵ-calibrate.*

Impossibility of ϵ-calibration in deterministic forecasting

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.
2. Take outcome space $Z = \{0, 1\}$.

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.
2. Take outcome space $Z = \{0, 1\}$.
3. A (deterministic) forecast at time t is real number $p^t \in [0, 1]$.

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.
2. Take outcome space $Z = \{0, 1\}$.
3. A (deterministic) forecast at time t is real number $p^t \in [0, 1]$.
4. Construct a worst case realisation $\omega = z^1, z^2, z^3, \ldots$ as follows:

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.
2. Take outcome space $Z = \{0, 1\}$.
3. A (deterministic) forecast at time t is real number $p^t \in [0, 1]$.
4. Construct a worst case realisation $\omega = z^1, z^2, z^3, \ldots$ as follows:
   - Realise 1 at $t + 1$ if $p^t \leq 1/2$.

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.
2. Take outcome space $Z = \{0, 1\}$.
3. A (deterministic) forecast at time t is real number $p^t \in [0, 1]$.
4. Construct a worst case realisation $\omega = z^1, z^2, z^3, \dots$ as follows:
   – Realise 1 at $t + 1$ if $p^t \leq 1/2$.
   – Realise 0 at $t + 1$, otherwise.

Impossibility of ϵ-calibration in deterministic forecasting

1. Let $0 < \epsilon < 1/4$.
2. Take outcome space $Z = \{0, 1\}$.
3. A (deterministic) forecast at time t is real number $p^t \in [0, 1]$.
4. Construct a worst case realisation $\omega = z^1, z^2, z^3, \dots$ as follows:
   – Realise 1 at $t+1$ if $p^t \leq 1/2$.
   – Realise 0 at $t+1$, otherwise.

Now

$$\begin{aligned} C^t(\omega) &= \sum_{p \in \Delta(Z)} \frac{n^t(p)}{t} \|\phi^t(p) - p\|^2 \\ &= \sum_{p \geq 1/2} \frac{n^t(p)}{t} \|0 - p\|^2 + \sum_{p < 1/2} \frac{n^t(p)}{t} \|1 - p\|^2 \\ &\geq \lambda \|\frac{1}{2}\|^2 + (1-\lambda)\|\frac{1}{2}\|^2 = \lambda \frac{1}{4} + (1-\lambda)\frac{1}{4} = \frac{1}{4} > \epsilon. \end{aligned}$$

Randomisation in forecasting

Randomisation in forecasting

1. A finite outcome space Z.

Randomisation in forecasting

1. A finite outcome space Z.
2. A random forecasting rule is a function

$$F : H \to \Delta(Z)$$

Randomisation in forecasting

1. A finite outcome space Z.
2. A random forecasting rule is a function

$$F : H \to \Delta(Z)$$

3. Let e^z be the representation of z in $\{0, 1\}^{|Z|}$.

Randomisation in forecasting

1. A finite outcome space Z.
2. A random forecasting rule is a function
$$F : H \to \Delta(Z)$$
3. Let e^z be the representation of z in $\{0, 1\}^{|Z|}$.
4. If $\omega = z^1, z^2, \ldots$ is a realisation, and $p^1, p^2, \ldots$ a sequence of forecasts, then F is calibrated on ω if the average of
$$\frac{1}{2} \| p^t - e^t \|^2$$
goes to zero.

Randomisation in forecasting

1. A finite outcome space Z.
2. A random forecasting rule is a function
$$F : H \to \Delta(Z)$$
3. Let e^z be the representation of z in $\{0, 1\}^{|Z|}$.
4. If $\omega = z^1, z^2, \ldots$ is a realisation, and $p^1, p^2, \ldots$ a sequence of forecasts, then F is calibrated on ω if the average of
$$\frac{1}{2}\|p^t - e^t\|^2$$
goes to zero.
5. F is calibrated if the average error goes to zero for every realisation almost surely.

Vohra *et al.*'s forecasting rule for random forecasts

Theorem. (Foster and Vohra, 1997-98). *Given any finite set* Z *and any* $\epsilon > 0$*, there exist random forecasting rules that are* ϵ*-calibrated for all sequences on* Z*.*

Vohra *et al.*'s forecasting rule for random forecasts

Theorem. (Foster and Vohra, 1997-98). *Given any finite set Z and any $\epsilon > 0$, there exist random forecasting rules that are ϵ-calibrated for all sequences on Z.*

1. Create an ϵ-covering Δ_ϵ of $\Delta(Z)$. This is a finite subset of $\Delta(Z)$ such that for every $p \in \Delta(Z)$ we have $d(p, \Delta_\epsilon) \leq \epsilon$.

Vohra *et al.*'s forecasting rule for random forecasts

Theorem. (Foster and Vohra, 1997-98). *Given any finite set Z and any $\epsilon > 0$, there exist random forecasting rules that are ϵ-calibrated for all sequences on Z.*

1. Create an ϵ-covering Δ_ϵ of $\Delta(Z)$. This is a finite subset of $\Delta(Z)$ such that for every $p \in \Delta(Z)$ we have $d(p, \Delta_\epsilon) \leq \epsilon$.

2. Row actions: all forecasts from Δ_ϵ.

Vohra *et al.*'s forecasting rule for random forecasts

Theorem. (Foster and Vohra, 1997-98). *Given any finite set Z and any $\epsilon > 0$, there exist random forecasting rules that are ϵ-calibrated for all sequences on Z.*

1. Create an ϵ-covering Δ_ϵ of $\Delta(Z)$. This is a finite subset of $\Delta(Z)$ such that for every $p \in \Delta(Z)$ we have $d(p, \Delta_\epsilon) \leq \epsilon$.

2. Row actions: all forecasts from Δ_ϵ.

3. Column actions: elements from Z represented as elements of $e^z \in \{0, 1\}^{|Z|}$.

Vohra *et al.*'s forecasting rule for random forecasts

Theorem. (Foster and Vohra, 1997-98). *Given any finite set Z and any $\epsilon > 0$, there exist random forecasting rules that are ϵ-calibrated for all sequences on Z.*

1. Create an ϵ-covering Δ_ϵ of $\Delta(Z)$. This is a finite subset of $\Delta(Z)$ such that for every $p \in \Delta(Z)$ we have $d(p, \Delta_\epsilon) \leq \epsilon$.
2. Row actions: all forecasts from Δ_ϵ.
3. Column actions: elements from Z represented as elements of $e^z \in \{0, 1\}^{|Z|}$.
4. Utility for row:
$$u(p, e^z) =_{Def} -\|p - e^z\|^2.$$

Vohra *et al.*'s forecasting rule for random forecasts

Theorem. (Foster and Vohra, 1997-98). *Given any finite set Z and any $\epsilon > 0$, there exist random forecasting rules that are ϵ-calibrated for all sequences on Z.*

1. Create an ϵ-covering Δ_ϵ of $\Delta(Z)$. This is a finite subset of $\Delta(Z)$ such that for every $p \in \Delta(Z)$ we have $d(p, \Delta_\epsilon) \leq \epsilon$.
2. Row actions: all forecasts from Δ_ϵ.
3. Column actions: elements from Z represented as elements of $e^z \in \{0,1\}^{|Z|}$.
4. Utility for row:
$$u(p, e^z) =_{Def} -\|p - e^z\|^2.$$
5. Minimise conditional regret.

Vohra *et al.*'s forecasting rule for random forecasts

Vohra *et al.*'s forecasting rule for random forecasts

- Let $\omega = z_1, z_2, z_3, \dots$ be a realisation. Consider a forecast $p^* \in \Delta_\epsilon$.

Vohra *et al.*'s forecasting rule for random forecasts

- Let $\omega = z_1, z_2, z_3, \dots$ be a realisation. Consider a forecast $p^* \in \Delta_\epsilon$.
- Because of conditional regret-minimisation, the mean square error w.r.t. p^*:

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| p^* - e^z \|^2$$

is minimal with respect to alternative forecasts.

Vohra *et al.*'s forecasting rule for random forecasts

- Let $\omega = z_1, z_2, z_3, \dots$ be a realisation. Consider a forecast $p^* \in \Delta_\epsilon$.
- Because of conditional regret-minimisation, the mean square error w.r.t. p^*:

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| p^* - e^z \|^2$$

is minimal with respect to alternative forecasts.

- Let $\phi(p^*) =_{Def} \lim_{t\to\infty} \phi^t(p^*)$. Then the absolute minimum is

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| \phi(p^*) - e^z \|^2$$

Vohra *et al.*'s forecasting rule for random forecasts

- Let $\omega = z_1, z_2, z_3, \dots$ be a realisation. Consider a forecast $p^* \in \Delta_\epsilon$.
- Because of conditional regret-minimisation, the mean square error w.r.t. p^*:

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| p^* - e^z \|^2$$

  is minimal with respect to alternative forecasts.
- Let $\phi(p^*) =_{Def} \lim_{t\to\infty} \phi^t(p^*)$. Then the absolute minimum is

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| \phi(p^*) - e^z \|^2$$

- Choose $p' \in \Delta_\epsilon$ with $\| p' - \phi(p^*) \| \leq \epsilon$.

Vohra *et al.*'s forecasting rule for random forecasts

- Let $\omega = z_1, z_2, z_3, \dots$ be a realisation. Consider a forecast $p^* \in \Delta_\epsilon$.
- Because of conditional regret-minimisation, the mean square error w.r.t. p^*:
$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| p^* - e^z \|^2$$
is minimal with respect to alternative forecasts.
- Let $\phi(p^*) =_{Def} \lim_{t\to\infty} \phi^t(p^*)$. Then the absolute minimum is
$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| \phi(p^*) - e^z \|^2$$
- Choose $p' \in \Delta_\epsilon$ with $\| p' - \phi(p^*) \| \leq \epsilon$.
- $\| p' - p^* \|$ is small because the difference in conditional regrets is small.

Vohra *et al.*'s forecasting rule for random forecasts

- Let $\omega = z_1, z_2, z_3, \dots$ be a realisation. Consider a forecast $p^* \in \Delta_\epsilon$.
- Because of conditional regret-minimisation, the mean square error w.r.t. p^*:

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| p^* - e^z \|^2$$

  is minimal with respect to alternative forecasts.
- Let $\phi(p^*) =_{Def} \lim_{t\to\infty} \phi^t(p^*)$. Then the absolute minimum is

$$\frac{1}{n^t(p^*)} \lim_{t\to\infty} \sum \frac{1}{2} \| \phi(p^*) - e^z \|^2$$

- Choose $p' \in \Delta_\epsilon$ with $\| p' - \phi(p^*) \| \leq \epsilon$.
- $\| p' - p^* \|$ is small because the difference in conditional regrets is small. It follows that $\| \phi(p^*) - p^* \|$ is small. (Triangle inequality.)

Foster's forecasting rule

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.
- A forecast is any $p^t \in \Delta_\epsilon$, representing $P\{z^t = 1\}$.

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.
- A forecast is any $p^t \in \Delta_\epsilon$, representing $P\{z^t = 1\}$.
- Let $\phi^t(p)$ the empirical frequency of 1's at times p was predicted.

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.
- A forecast is any $p^t \in \Delta_\epsilon$, representing $P\{z^t = 1\}$.
- Let $\phi^t(p)$ the empirical frequency of 1's at times p was predicted.
- The empirical frequency is updated geometrically:

$$\phi^{t+1}(p) =_{Def} (1-\lambda)\phi^t(p) + \lambda z^t.$$

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.
- A forecast is any $p^t \in \Delta_\epsilon$, representing $P\{z^t = 1\}$.
- Let $\phi^t(p)$ the empirical frequency of 1's at times p was predicted.
- The empirical frequency is updated geometrically:

$$\phi^{t+1}(p) =_{Def} (1-\lambda)\phi^t(p) + \lambda z^t.$$

- A predictor p is said to be good at t if $|p - \phi^t(p)| \leq \epsilon$.

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.
- A forecast is any $p^t \in \Delta_\epsilon$, representing $P\{z^t = 1\}$.
- Let $\phi^t(p)$ the empirical frequency of 1's at times p was predicted.
- The empirical frequency is updated geometrically:

$$\phi^{t+1}(p) =_{Def} (1-\lambda)\phi^t(p) + \lambda z^t.$$

- A predictor p is said to be good at t if $|p - \phi^t(p)| \leq \epsilon$.
- A pair $(p, p+\epsilon)$ is said to be skew at t if

$$\phi^t(p) > p + \epsilon \quad \text{and} \quad \phi^t(p+\epsilon) < p.$$

Foster's forecasting rule

- Suppose a binary outcome space $Z = \{0, 1\}$.
- Let $\Delta_\epsilon = \{\epsilon, 2\epsilon, 3\epsilon, \dots, (N-1)\epsilon\}$, for some $N \geq 3$.
- A forecast is any $p^t \in \Delta_\epsilon$, representing $P\{z^t = 1\}$.
- Let $\phi^t(p)$ the empirical frequency of 1's at times p was predicted.
- The empirical frequency is updated geometrically:

$$\phi^{t+1}(p) =_{Def} (1-\lambda)\phi^t(p) + \lambda z^t.$$

- A predictor p is said to be good at t if $|p - \phi^t(p)| \leq \epsilon$.
- A pair $(p, p+\epsilon)$ is said to be skew at t if

$$\phi^t(p) > p + \epsilon \quad \text{and} \quad \phi^t(p+\epsilon) < p.$$

- If all predictors are bad, there must be a skew pair. (Check!)

Foster's forecasting rule

Algorithm:

Foster's forecasting rule

Algorithm:

- Initialise $\phi^0(p)$ to a random number in $[0, 1]$, for example $1/2$.

  To let p start at "the right place", setting $\phi^0(p) =_{Def} p$ is also good.

Loop:

Foster's forecasting rule

Algorithm:

- Initialise $\phi^0(p)$ to a random number in $[0,1]$, for example $1/2$.

  To let p start at "the right place", setting $\phi^0(p) =_{Def} p$ is also good.

Loop:

1. If there is a good forecaster p, use it and update $\phi^t(p)$.

Foster's forecasting rule

Algorithm:

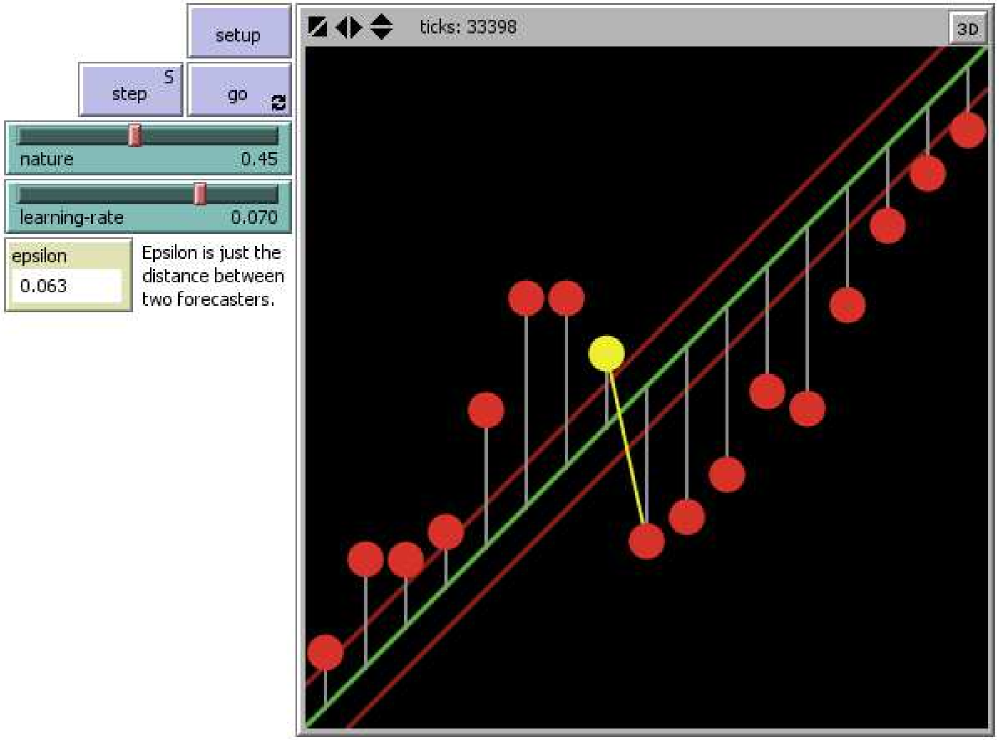
- Initialise $\phi^0(p)$ to a random number in $[0, 1]$, for example $1/2$.

  To let p start at "the right place", setting $\phi^0(p) =_{Def} p$ is also good.

Loop:

1. If there is a good forecaster p, use it and update $\phi^t(p)$.

2. Else, pick a skew pair and forecast with either one of them (choose randomly).



Original algorithm in SLaiL

Original algorithm in SLaiL

1. For each forecaster compute empirical mean (rather than geometric empirical mean).

Original algorithm in SLaiL

1. For each forecaster compute empirical mean (rather than geometric empirical mean).

2. If there is no good forecaster, pick a skew pair $(p, p + \epsilon)$

Original algorithm in SLaiL

1. For each forecaster compute empirical mean (rather than geometric empirical mean).
2. If there is no good forecaster, pick a skew pair $(p, p + \epsilon)$
3. Make action vector orthogonal to error vector:

$$\underbrace{\begin{pmatrix} b^t \\ a^t \end{pmatrix}}_{\text{Actions}} \perp \underbrace{\begin{pmatrix} a^t \\ -b^t \end{pmatrix}}_{\text{Error}}.$$

Now forecast with odds $b^t : a^t$, i.e., with p a $b^t/(a^t + b^t)$ of the times.