Prediction with Expert Advice for the Brier Game

Vladimir Vovk and Fedor Zhdanov

Department of Computer Science
Royal Holloway, University of London

ICML 2008

Outline

Problem Area

- Prediction with expert advice.
- Competitive online learning.
- Agent accumulates losses and competes with other agents.

Example

- We predict the results of football matches game by game.

Example

- We predict the results of football matches game by game.
- Possible outcomes: Home win, Draw, Away win.

Example

- We predict the results of football matches game by game.
- Possible outcomes: Home win, Draw, Away win.
- Our prediction: 3 probabilities $p_{\text{home}} + p_{\text{draw}} + p_{\text{away}} = 1$.

Example

- We predict the results of football matches game by game.
- Possible outcomes: Home win, Draw, Away win.
- Our prediction: 3 probabilities $p_{\text{home}} + p_{\text{draw}} + p_{\text{away}} = 1$.
- We compete with bookmakers.

Example

- We predict the results of football matches game by game.
- Possible outcomes: Home win, Draw, Away win.
- Our prediction: 3 probabilities $p_{\text{home}} + p_{\text{draw}} + p_{\text{away}} = 1$.
- We compete with bookmakers.
- Goal: To be close to the best bookmaker according to some loss function.

Framework: Notation

- We predict events $\omega_t \in \Omega$.
- Our predictions are $\gamma_t \in \Gamma$.
- The quality of predictions is measured by a loss function $\lambda(\omega_t, \gamma_t) \geq 0$.

Framework: Protocol

$L_t^k := 0, \; k = 1, \ldots, K$
$L_t := 0$

FOR $t = 1, 2, \ldots$
  Experts: predictions $\gamma_t^k \in \Gamma, \; k = 1, \ldots, K$
  Learner: prediction $\gamma_t \in \Gamma$
  Reality: the actual outcome $\omega_t \in \Omega$

  Experts $L_t^k := L_{t-1}^k + \lambda(\omega_t, \gamma_t^k), \; k = 1, \ldots, K$
  Learner $L_t := L_{t-1} + \lambda(\omega_t, \gamma_t)$
END FOR

Framework: Goal

At each step t for any expert k cumulative loss

$$L_t \leq L_t^k + R$$

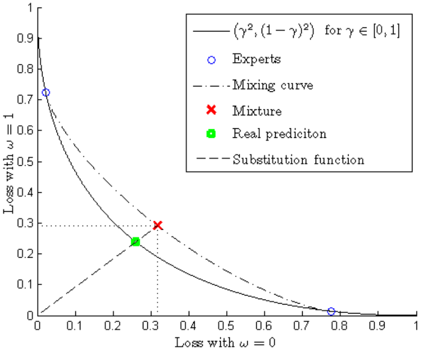
R is some small regret term

Algorithms

- Aggregating Algorithm (Vovk, 1990)
- Weighted Average Algorithm (Kivinen and Warmuth, 1999)
- Hedge Algorithm (Freund and Schapire, 1997)
- Weak Aggregating Algorithm (Kalnishkan and Vyugin, 2005)

Aggregating Algorithm

Initialize weights $w^k := 1/K$, $k = 1, \dots, K$
FOR $t = 1, 2, \dots$:
Mixture:
$$g(\omega) = -\frac{1}{\eta} \ln \sum_{k=1}^{K} e^{-\eta\lambda(\omega,\gamma^k)} w^k,$$
$\omega \in \Omega$.
Prediction $\gamma = S(g)$.
Reality announces the actual outcome ω^*.
Update weights:
$$w^k := w^k e^{-\eta\lambda(\omega^*,\gamma^k)}$$
END FOR.



Outline

Brier game of prediction

- n possible outcomes.
- ω is represented as $(0, \ldots, 1, \ldots, 0)$.
- γ is a probability distribution on outcomes.
- The Brier loss function $\lambda(\omega, \gamma) = \sum_{i=1}^{n} (\omega_i - \gamma_i)^2$.
- Brier game: probability forecasting for meteorology (1950).

Our theoretical results

- $\exists S$ such that for any mixture g,
  $\gamma = S(g(\omega)) : \lambda(\omega, \gamma_i) \leq g(\omega), \forall \omega \in \Omega$, for $\eta \leq 1$.
  This means λ is *mixable* for such η.
- The optimal $\eta = 1$ and corresponding substitution function was found.

Theorem

Learner can use the Aggregating Algorithm for the Brier game to guarantee

$$L_T \leq \min_{k=1,\dots,K} L_T^k + \ln K$$

for all $T = 1, 2, \dots,$ *where* K *is the number of experts. Moreover, this bound is optimal: this constant cannot be decreased.*

Outline

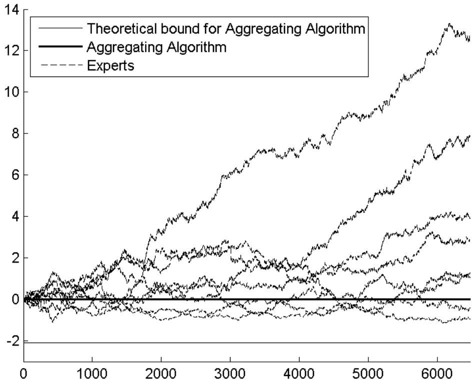
Football Experiment

English Championship database:

- 6473 matches in the last 3 seasons.
- 3 possible outcomes: Home win, Draw, Away win.
- 8 bookmakers (= 8 experts) give their decimal odds. We need to convert these odds into probabilities of each result:

$$p_i = \frac{1/a_i}{1/a_1 + 1/a_2 + 1/a_3}, i = 1, 2, 3.$$

Football Results



$$L_t^{AA} \le L_t^k + \ln 8 \quad \Leftrightarrow \quad L_t^k - L_t^{AA} \ge -\ln 8$$

The theoretical bound holds.

Football Results

The maximal difference between the cumulative loss of the algorithm and the cumulative loss of the best expert for football data: $\max_t(L_t^{AA} - \min_k L_t^k)$.

Algorithm	Maximal difference	Theoretical bound
Aggregating	1.1562	2.0794
Weighted Average	1.8697	16.6355
Hedge	4.5662	234.1159
Weak Aggregating	2.4755	464.0728

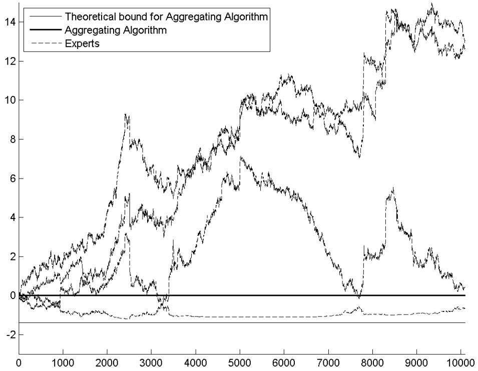
Tennis Experiment

All tournaments database:

- 10087 matches in the last 4 seasons.
- 2 possible outcomes: Winner, Loser.
- 4 bookmakers (= 4 experts) give their decimal odds. We need to convert these odds into probabilities of each result:

$$p_i = \frac{1/a_i}{1/a_1 + 1/a_2}, i = 1, 2.$$

Tennis Results



$$L_t^{AA} \le L_t^k + \ln 4 \quad \Leftrightarrow \quad L_t^k - L_t^{AA} \ge -\ln 4$$

The theoretical bound is very tight.

Tennis Results

The maximal difference between the cumulative loss of the algorithm and the cumulative loss of the best expert for tennis data $\max_t (L_t^{AA} - \min_k L_t^k)$.

Algorithm	Maximal difference	Theoretical bound
Aggregating	1.2021	1.3863
Weighted Average	3.0566	11.0904
Hedge	9.0598	237.8904
Weak Aggregating	3.6101	473.0083

Outline

Summary

- Vovk, V., Zhdanov, F.: Prediction with expert advice for the Brier game, arXiv:0710.0485v1 [cs.LG] (2008)
- The probability forecasting game with the Brier loss function is considered.
- The Aggregating Algorithm is applied to this game. The optimal theoretical bound is found.
- First experiments in the online competitive prediction setting.
- The theoretical bound is quite tight in both experiments.
- Other algorithms give bigger maximal loss difference with the best expert.

Thank you

Thank you