Hybrid Stochastic-Adversarial On-line Learning

Alessandro LAZARIC and Rémi Munos
INRIA, *SequeL* (Sequential Learning) project, France

ECML'09 Workshop on *Learning from non-IID data*,
September 7, 2009, Bled

Motivating example

A (not-so-serious) real-life prediction problem:

Motivating example

A (not-so-serious) real-life prediction problem:

Predict: $\{Yes, No\}$ for dinner tonight

Motivating example

A (not-so-serious) real-life prediction problem:

Predict: $\{Yes, No\}$ for dinner tonight

- Inputs (*girls*) are drawn from a fixed probability distribution
- The relationship between inputs and outputs is complex
- Hybrid stochastic (inputs) adversarial (outputs) problem

Motivating example

A (little-bit-more-serious) real-life prediction problem:

Predict: $\{Buy, NotBuy\}$ the new model of mobile phone

Motivating example

A (little-bit-more-serious) real-life prediction problem:

Predict: $\{Buy, NotBuy\}$ the new model of mobile phone

- Inputs (*potential users*) are drawn from a fixed probability distribution
- The relationship between inputs and outputs is complex
- Hybrid stochastic (inputs) adversarial (outputs) problem

Outline

Online classification problem

Input space $\mathcal{X}$, Output space $\mathcal{Y} = \{0, 1\}$,
Hypothesis space $\mathcal{H} = \{h : \mathcal{X} \to \mathcal{Y}\}$

for $t = 1, 2, \ldots,$

- The learner observes $x_t \in \mathcal{X}$,
- The learner chooses $h_t \in \mathcal{H}$ and predicts $\widehat{y}_t = h_t(x_t) \in \mathcal{Y}$,
- $y_t \in \mathcal{Y}$ is revealed,
- The learner incurs a loss $\ell(\widehat{y}_t, y_t) = \mathbb{I}\{\widehat{y}_t \neq y_t\}$

Goal : minimize the cumulative regret:

$$R_n = \sum_{t=1}^{n} \ell(\widehat{y}_t, y_t) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell(h(x_t), y_t)$$

Known results in different settings

- **Fully stochastic setting:** $(x_t, y_t) \overset{iid}{\sim} P$,

$$R_n = O(\sqrt{VC(\mathcal{H}) n \log n})$$

(by using an Empirical Risk Minimizer on-line)

Known results in different settings

- **Fully stochastic setting:** $(x_t, y_t) \overset{iid}{\sim} P$,

$$R_n = O(\sqrt{VC(\mathcal{H}) n \log n})$$

(by using an Empirical Risk Minimizer on-line)

- **Fully adversarial setting:** (x_t, y_t) chosen by adversary,

$$R_n = O(\sqrt{Ldim(\mathcal{H}) n \log n})$$

(Agnostic Online Learning [Ben-David, Pál and Shalev-Shwartz, 2009])

Known results in different settings

- **Fully stochastic setting:** $(x_t, y_t) \overset{iid}{\sim} P$,

$$R_n = O(\sqrt{VC(\mathcal{H}) n \log n})$$

(by using an Empirical Risk Minimizer on-line)

- **Fully adversarial setting:** (x_t, y_t) chosen by adversary,

$$R_n = O(\sqrt{Ldim(\mathcal{H}) n \log n})$$

(Agnostic Online Learning [Ben-David, Pál and Shalev-Shwartz, 2009])

- **Hybrid stochastic */* adversarial setting:** $x_t \overset{iid}{\sim} P$ and y_t chosen by adversary.

$$R_n = O(\sqrt{VC(\mathcal{H}) n \log n})$$

(This work)

Hybrid Stochastic-Adversarial setting

Hybrid stochastic/adversarial classification problem:
for $t = 1, 2, \ldots,$

- Nature chooses $x_t \overset{iid}{\sim} P$ which is revealed
- Simultaneously,
  - The adversary chooses y_t,
  - The learner chooses $h_t \in \mathcal{H}$ and predicts $\widehat{y}_t = h_t(x_t) \in \mathcal{Y}$,
- y_t is revealed,
- The learner incurs a loss $\ell(\widehat{y}_t, y_t) = \mathbb{I}\{\widehat{y}_t \neq y_t\}$

Goal : minimize the cumulative regret:

$$R_n = \sum_{t=1}^{n} \ell(\widehat{y}_t, y_t) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell(h(x_t), y_t)$$

Outline

Exponentially weighted forecaster (EWF)

Finite number of hypotheses (experts) $|\mathcal{H}| = N < \infty$.

Set $w_i^1 = 1$ for all $i \in \{1, \dots, N\}$,
For $t = 1, 2, \dots,$

- Observe the input x_t,
- Select an hypothesis

$$h_t \sim \mathbf{p}^t = (p_1^t, \dots, p_N^t), \text{ where } p_i^t = \frac{w_i^t}{\sum_{j=1}^N w_j^t}$$

  and predict $\hat{y}_t = h_t(x_t)$,
- Observe the label y_t and update the weights:

$$w_i^t = w_i^t \exp\big(-\eta\, \ell(h_i(x_t), y_t)\big), \quad i \in \{1, \dots, N\}$$

Exponentially weighted forecaster

Theorem (Cesa-Bianchi and Lugosi, 2006)

Let $n, N \geq 1$, $0 \leq \beta \leq 1$, $\eta = \sqrt{2 \log N / n}$. With probability at least $1 - \beta$, the regret of EWF satisfies

$$\begin{aligned} R_n &= \sum_{t=1}^{n} \ell(\hat{y}_t, y_t) - \min_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell(h(x_t), y_t) \\ &\leq \sqrt{2n \log N} + \sqrt{\frac{n}{2} \log \frac{1}{\beta}}. \end{aligned}$$

Outline

The EStochAd Forecaster

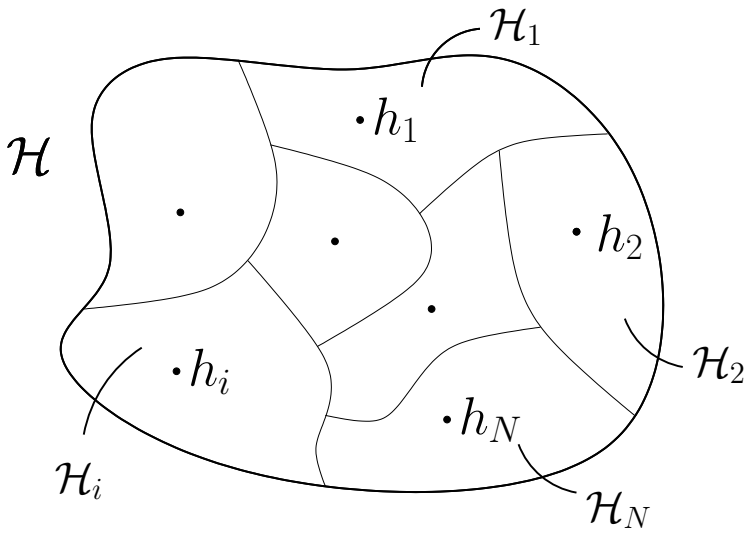
Infinite hypothesis space with finite VC ($VC(\mathcal{H}) = d < \infty$)

Basic idea of the EStochAd Forecaster:

- Epoch-based algorithm,
- At the beginning of each epoch, build a partition of the hypothesis space into a finite number of classes in which all hypotheses have same predictions on past inputs.
- In each epoch, run EWF using a (finite) set of representatives of each class.

Partition of the hypothesis space

Sequence of inputs $\{x_1, \dots, x_t\}$ defines the partition $\mathcal{H} = \{\mathcal{H}_i, 1 \leq i \leq N_t\}$, where $\forall h, h' \in \mathcal{H}_i, h(x_s) = h'(x_s), \forall s \leq t$.



Let h_i be a representative of $\mathcal{H}_i$ and write $H_t = \{h_1, \dots, h_{N_t}\}$.
If the VC-dimension of $\mathcal{H}$ is finite, we have $|H_t| = O(t^{VC(\mathcal{H})})$.

EStochAd algorithm

Consider epochs of length $t_k = 2^k$. For (epoch) $k = 0, 1, 2, \ldots$,

- **Partition the hypothesis space** according to the past inputs $\{x_1, \ldots, x_{t_k}\}$ and select a set of representative hypotheses H_k.
- **Run EWF during epoch** k using the set of hypotheses H_k:
  Initialize the weights $w_i^{t_k} = 1$, and set $\eta_k = \sqrt{2 \log |H_k| / t_k}$
  For $t = t_k + 1, \ldots, t_{k+1}$,
  - Observe x_t, select $h_t \sim \mathbf{p}^t$, with $p_i = w_i^t / (\sum_{j=1}^{N_k} w_j^t)$, and predict $\widehat{y}_t = h_i(x_t)$,
  - Observe the true label y_t and update the weights: $w_i^{t+1} = w_i^t \exp\left(-\eta_k \ell(h_i(x_t), y_t)\right)$

EStochAd

Theorem

For any $\beta > 0$, the cumulative regret of EStochAd satisfies

$$R_n \leq c_1 \sqrt{VC(\mathcal{H}) n \log \frac{en}{d}} + c_2 \sqrt{n \log \left[12 (\lfloor \log_2 n \rfloor + 1) \beta^{-1} \right]}$$

with probability at least $1 - \beta$, where $c_1 = 18 + 10\sqrt{2}$, and $c_2 = 18(\sqrt{2} + 1)$.

Sketch of proof (1)

The regret decomposes as

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \ell(h_t(x_t), y_t)) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell(h(x_t), y_t) \\
&\leq \sum_{k=0}^{K-1} \underbrace{\left(\sum_{t=t_k+1}^{t_{k+1}} \ell(h_t(x_t), y_t) - \inf_{h \in \mathcal{H}} \sum_{t=t_k+1}^{t_{k+1}} \ell(h(x_t), y_t) \right)}_{R_k}
\end{aligned}
$$

where R_k is the regret at epoch k.

Sketch of proof (2)

$R_k = R_k^1 + R_k^2$ with:

$$R_k^1 = \sum_{t=t_k+1}^{t_{k+1}} \ell(h_t(x_t), y_t) - \min_{h \in H_k} \sum_{t=t_k+1}^{t_{k+1}} \ell(h(x_t), y_t)$$

$$R_k^2 = \min_{h \in H_k} \sum_{t=t_k+1}^{t_{k+1}} \ell(h(x_t), y_t) - \inf_{h \in \mathcal{H}} \sum_{t=t_k+1}^{t_{k+1}} \ell(h(x_t), y_t).$$

- R_k^1 is algorithm dependent, and running EWF gives:

  $$R_k^1 = O\left(\sqrt{t_k \log(|H_k|\beta^{-1})}\right) = O\left(\sqrt{t_k (VC(\mathcal{H}) \log t_k + \log \beta^{-1})}\right)$$

- R_k^2 depends on how much H_k is representative of $\mathcal{H}$ for prediction at samples that are observed in epoch k
  –> Stochastic input assumption

Sketch of proof (3)

$$
\begin{aligned}
R_k^2 &= \min_{h\in H_k} \sum_{t=t_k+1}^{t_{k+1}} \ell(h(x_t),y_t) - \inf_{h\in\mathcal{H}} \sum_{t=t_k+1}^{t_{k+1}} \ell(h(x_t),y_t) \\
&= \sup_{h\in\mathcal{H}} \min_{h'\in H_k} \sum_{t=t_k+1}^{t_{k+1}} \ell(h'(x_t),y_t) - \ell(h(x_t),y_t) \\
&\le \sup_{h\in\mathcal{H}} \min_{h'\in H_k} \sum_{t=t_k+1}^{t_{k+1}} \mathbb{I}\left\{h(x_t)\neq h'(x_t)\right\} \\
&\le \sup_{h\in\mathcal{H}} \min_{h'\in H_k} t_k \mathbb{E}_{X\sim P}\left[\mathbb{I}\left\{h(X)\neq h'(X)\right\}\right] + O\left(\sqrt{t_k\left[VC(\mathcal{H}^2)\log t_k + \log\beta^{-1}\right]}\right) \\
&\le \sup_{h\in\mathcal{H}} \min_{h'\in H_k} \sum_{t=1}^{t_k} \mathbb{I}\left\{h(x_t)\neq h'(x_t)\right\} + O\left(\sqrt{t_k\left[VC(\mathcal{H}^2)\log t_k + \log\beta^{-1}\right]}\right) \\
&= O\left(\sqrt{t_k\left[VC(\mathcal{H}^2)\log t_k + \log\beta^{-1}\right]}\right)
\end{aligned}
$$

Sketch of proof (4)

Putting everything together... We have that

$$R_k^1 + R_k^2 = O\left(\sqrt{t_k(VC(\mathcal{H}) \log t_k + \log \beta^{-1})}\right)$$

Thus

$$R_n = \sum_{k=0}^{\log_2 n} R_k^1 + R_k^2 = O\left(\sqrt{n(VC(\mathcal{H}) \log n + \log \beta^{-1})}\right)$$

with high probability.

Outline

Multi-armed Bandit with stochastic side information

There are K arms. For $t = 1, 2, \ldots,$

- Nature chooses $x_t \overset{iid}{\sim} P$ which is revealed
- Simultaneously,
  - Adversary chooses a loss function $\ell_t : \{1, \ldots, K\} \to [0, 1]$,
  - Learner chooses an arm I_t,
- The learner incurs the loss $\ell_t(I_t)$ and no other information is provided.

Goal : minimize the cumulative regret:

$$R_n = \sum_{t=1}^{n} \ell_t(I_t) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{n} \ell_t(h(x_t))$$

Multi-armed Bandit with stochastic side information

- Instead of EWF, we use an Exp4-like algorithm [Auer et al., 2002],

$$R_n(\text{Exp4}) \leq 4\sqrt{nK \log \frac{nN}{\beta}} + 8 \log \frac{nN}{\beta}$$

- Plugging the bandit regret bound into the *EStochAd* algorithm

$$R_n \leq O\left(\sqrt{nK\, Ndim(\mathcal{H}) \log \frac{nK^2}{\beta}} + Ndim(\mathcal{H}) \log \frac{nK^2}{\beta}\right)$$

where $Ndim(\mathcal{H})$ is the Natarajan dimension of $\mathcal{H}$.

Outline

Two-player Game with Stochastic Side Information

For $t = 1, 2, \ldots,$

- Simultaneously,
  - Nature chooses $x_t \overset{iid}{\sim} P$
  - Player A chooses strategy $h_{A,t}$
  - Player B chooses strategy $h_{B,t}$
- Player A plays $\widehat{y}_{A,t} = h_{A,t}(x_t)$, player B plays $\widehat{y}_{B,t} = h_{B,t}(x_t)$
- Return feedback:
  - *Bandit information*: $\ell_A(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$ and $\ell_B(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$
  - *Full information*: $\ell_A(\cdot, \widehat{y}_{B,t}, x_t)$ and $\ell_B(\widehat{y}_{A,t}, \cdot, x_t)$
- Player A incurs a loss $\ell_A(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t)$,
  Player B incurs a loss $\ell_B(\widehat{y}_{A,t}, \widehat{y}_{B,t}, x_t))$

2-players zero-sum Game

- (σ_A^*, σ_B^*) is a Nash equilibrium if

$$\begin{aligned} \overline{\ell}_A(\sigma_A^*, \sigma_B^*) &\leq \overline{\ell}_A(\sigma_A, \sigma_B^*), \quad \forall \sigma_A \in \mathcal{D}(\mathcal{H}) \\ \overline{\ell}_B(\sigma_A^*, \sigma_B^*) &\leq \overline{\ell}_B(\sigma_A^*, \sigma_B), \quad \forall \sigma_B \in \mathcal{D}(\mathcal{H}). \end{aligned}$$

- If $\mathcal{H}$ is a compact metric set then Minimax theorem:

$$\begin{aligned} V &= \sup_{\sigma_B \in \mathcal{D}(\mathcal{H})} \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B) \\ &= \inf_{\sigma_A \in \mathcal{D}(\mathcal{H})} \sup_{\sigma_B \in \mathcal{D}(\mathcal{H})} \overline{\ell}_A(\sigma_A, \sigma_B), \end{aligned}$$

Two-player Game with Stochastic Side Information

Theorem

Let losses ℓ_A, ℓ_B be bounded in $[0,1]$, $\mathcal{H}$ be a compact metric set. If both players run (Bandit-)EStochAd in a zero-sum game with stochastic side information as defined above, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \ell_A(h_{A,t}(x_t), h_{B,t}(x_t), x_t) = V$$

almost surely, and the frequencies of played strategies by both players converge to the set of Nash equilibria. The convergence rate is of order $O(\sqrt{\frac{N\dim(\mathcal{H})}{n} \log(nK^2)})$ (full information case).

Outline

Comparison

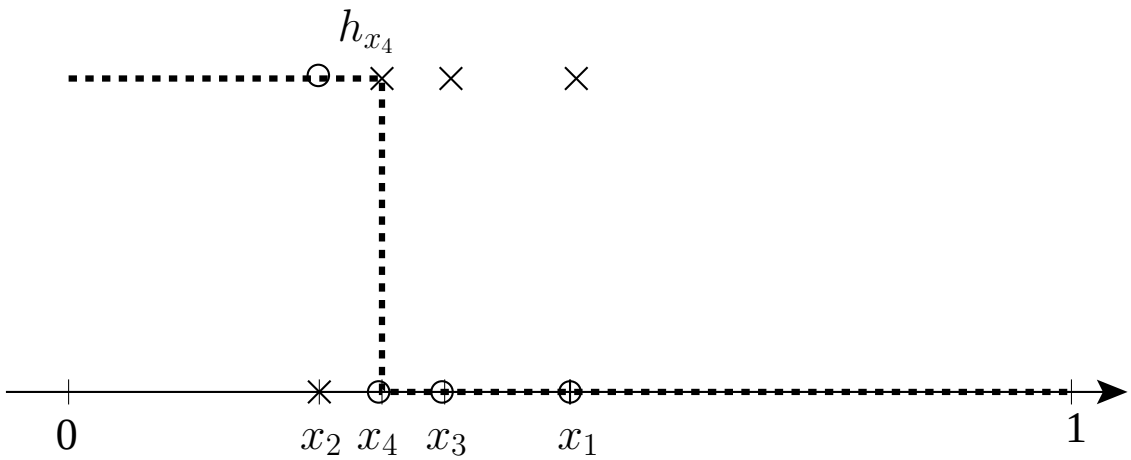
	EStochAd	Online-ERM	AOL
Input	Stochastic	Stochastic	Adversarial
Output	Adversarial	Stochastic	Adversarial
Complexity	VC	VC	Ldim
Bound	$\sqrt{nVC(\mathcal{H}) \log n}$	$\sqrt{nVC(\mathcal{H}) \log n}$	$\sqrt{nLdim(\mathcal{H}) \log n}$

- The VC dimension is already an *adversarial* measure of complexity!
- We have: $VC(\mathcal{H}) \leq Ldim(\mathcal{H})$ and in very simple problems adversarial inputs may lead to an arbitrarily poor performance.

Fully Adversarial vs Hybrid Setting

Let consider a binary classification problem with $X = [0, 1]$ and a hypothesis space $\mathcal{H}$ containing functions of the form

$$h_\vartheta(x) = \begin{cases} 0 & \text{if } x \geq \vartheta \\ 1 & \text{otherwise,} \end{cases} \quad \text{with } \vartheta \in [0, 1].$$



Here $Ldim(\mathcal{H}) = \infty$ whereas $VC(\mathcal{H}) = 1$.

Conclusions

- Motivation for hybrid stochastic-adversarial setting
- Regret upper-bound of *EStochAd* : $O(\sqrt{nVC(\mathcal{H})\log n})$
- Same complexity measure as in fully stochastic setting (VC = adversary measure of complexity)
- Straightforward extensions to multi-class classification, bandits and games with stochastic side information
- Other extensions: regression
- Future work:
  - Smoothed analysis (noisy adversary?)
  - Computationally efficient version (at the cost of the regret?)
  - Algorithmic implementations

Thank you

Questions?

Bibliography I


Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire.
The nonstochastic multiarmed bandit problem.
SIAM J. Comput., 32(1):48–77, 2003.


O. Bousquet, S. Boucheron, and G. Lugosi.
Introduction to statistical learning theory.
Advanced Lectures on Machine Learning Lecture Notes in Artificial Intelligence, 3176:169–207, 2004.


S. Ben-David, N. Cesa-Bianchi, D. Haussler, and P. M. Long.
Characterizations of learnability for classes of {0...n}-valued functions.
Journal of Computer and System Sciences, 50:74–86, 1995.


N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. Shapire, and M. Warmuth.
How to use expert advice.
Journal of the ACM, 44(3):427–485, 1997.


N. Cesa-Bianchi and G. Lugosi.
Prediction, Learning, and Games.
Cambridge University Press, 2006.


Koby Crammer and Yoram Singer.
Ultraconservative online algorithms for multiclass problems.
J. Mach. Learn. Res., 3:951–991, 2003.

Bibliography II


Luc Devroye, Laszlo Györfi, and Gabor Lugosi.
A Probabilistic Theory of Pattern Recognition (Stochastic Modelling and Applied Probability).
Springer, February 1997.


Michael Fink, Shai Shalev-Shwartz, Yoram Singer, and Shimon Ullman.
Online multiclass learning by interclass hypothesis sharing.
In *Proceedings of the 23rd international conference on Machine learning*, pages 313–320, New York, NY, USA, 2006. ACM.


Sham M. Kakade, Shai Shalev-Shwartz, and Ambuj Tewari.
Efficient bandit algorithms for online multiclass prediction.
In *Proceedings of the 25th international conference on Machine learning*, pages 440–447, New York, NY, USA, 2008. ACM.


Nick Littlestone.
Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm.
Mach. Learn., 2(4):285–318, 1988.


N. Littlestone and M. Warmuth.
The weighted majority algorithm.
Information and Computation, 108:212–261, 1994.


J. Langford and T. Zhang.
The epoch greedy algorithm for contextual multi-armed bandits.
In *Advances in Neural Information Processing Systems*, 2007.

Bibliography III


B. K . Natarajan.
On larning sets and functions.
Mach. Learn., 4:67–97, 1989.


F. Rosenblatt.
The perceptron : A probabilistic model for information storage and organization in the brain.
Psychological Review, 65(6):386–408, 1958.


Daniil Ryabko.
Pattern recognition for conditionally independent data.
J. Mach. Learn. Res., 7:645–664, 2006.


G. Stoltz and G. Lugosi.
Learning correlated equilibria in games with compact sets of strategies.
Games and Economic Behavior, 59(1):187 – 208, 2007.


Shai Shalev-Shwartz.
Agnostic online learnability.
Technical Report TTIC-TR-2008-2, Toyota Technological Institute, 2008.


V. Vovk.
A game of prediction with expert advice.
Journal of Computer and System Sciences, 56:153–173, 1998.

Bibliography IV

J. Weston and C. Watkins.

Support vector machines for multi-class patter recognition.
In *Proceedings of the Seventh European Symposium on Artificial Neural Networks*, 1999.