Online gradient descent for least squares regression: Non-asymptotic bounds and application to bandits

Prashanth. L.A.

INRIA Lille - Nord Europe, Team SequeL

September, 2013

Joint work with Nathaniel Korda and Rémi Munos,

INRIA Lille - Nord Europe, Team SequeL

Motivation [1]

Want to solve Ordinary Least Squares (OLS): $\hat{\theta}_n = \arg\min_\theta \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta^\top x_i)^2$

Complexity

- $O(d^2)$ using the Sherman-Morrison lemma or
- $O(d^{2.807})$ using the Strassen algorithm or $O(d^{2.375})$ the Coppersmith-Winograd algorithm

What we propose: Use online gradient descent (GD) to estimate $\hat{\theta}_t$

Why:

- Efficient with complexity of only $O(d)$ (Well-known)
- High probability bounds with explicit constants can be derived (not fully known)

Motivation [2]

A typical linear bandit algorithm

Given: arms x in a compact subset D of $\mathbb{R}^d$.
For $n = 1, 2, \ldots$ **do**

- Step 1 Compute an OLS estimate $\hat{\theta}_n$ based on arms x_i chosen and losses y_i seen so far, $i = 1, \ldots, n-1$
- Step 2 Construct an ellipsoid B_n^2 centered at $\hat{\theta}_t$
- Step 3 Choose x_n that gives the minimum estimated loss over B_n^2
- Step 4 Observe the reward y_n.

Motivation [2]

A typical linear bandit algorithm

Given: arms x in a compact subset D of $\mathbb{R}^d$.

For $n = 1, 2, \ldots$ **do**

- Step 1 Compute an OLS estimate $\hat{\theta}_n$ based on arms x_i chosen and losses y_i seen so far, $i = 1, \ldots, n-1$
- Step 2 Construct an ellipsoid B_n^2 centered at $\hat{\theta}_t$
- Step 3 Choose x_n that gives the minimum estimated loss over B_n^2
- Step 4 Observe the reward y_n.

Our contribution

Use online GD in Step 1 and study the impact on regret performance:

- Strongly-convex arms no impact on regret (barring log-factors) vis-a-vis PEGE algorithm
- Non-strongly convex arms $O\left(n^{1/5}\right)$ deterioration of the regret vis-a-vis ConfidenceBall algorithm

Outline

1. Bandits with strongly convex arms
   - Random online algorithm for OLS
   - Regret bounds
2. Bandits with non-strongly convex arms
   - Random online-regularized algorithm
   - Regret bounds
3. Conclusions

OUTLINE

1. BANDITS WITH STRONGLY CONVEX ARMS
   - Random online algorithm for OLS
   - Regret bounds
2. BANDITS WITH NON-STRONGLY CONVEX ARMS
   - Random online-regularized algorithm
   - Regret bounds
3. CONCLUSIONS

Random online algorithm

Pick a sample (x_{i_n}, y_{i_n}) uniformly randomly from the set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.
Update the iterate θ_n as

$$\theta_n = \theta_{n-1} + \gamma_n (y_{i_n} - \theta_{n-1}^{\mathsf{T}} x_{i_n}) x_{i_n}. \tag{1}$$

We assume:

(A1) Boundedness of x_n, i.e., $\sup_n \|x_n\|_2 \leq 1$.

(A2) The noise $\{\xi_n\}$ is i.i.d. and $|\xi_n| \leq 1, \forall n$.

(A3) For all n, $\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^{\mathsf{T}}\right) \geq \mu$.[a]

[a] $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix.

Error bound

Theorem

With $\gamma_n = c/n$ and $c > 1/(2\mu)$, we have, for any $\delta > 0$,

$$P\left(\left\|\theta_n - \hat{\theta}_n\right\|_2 \leq \sqrt{\frac{K_{\mu,c}}{n}\log\frac{1}{\delta}} + \left(\frac{\left\|\theta_0 - \hat{\theta}_0\right\|_2}{n^{\mu c}} + \frac{h_1(n)}{\sqrt{n}}\right)\right) \geq 1 - \delta. \qquad (2)$$

$h_1(n)$ hides log factors, $K_{\mu,c}$ depends on μ and c

By averaging the iterates, the dependency on μ can be removed while obtaining optimal rate of convergence.

Application to bandits[1]

- Arms x_n evolve in a set $\mathcal{D} \subset \mathbb{R}^d$ such that a basis $\{b_1, \dots, b_d\} \in \mathcal{D}$ for $\mathbb{R}^d$ is known
- Losses $y_n = l_n(x_n)$ satisfy $E[\, l_n(x_n) |\, x_n] = x_n^\mathsf{T}\theta^*$
- Aim: minimise the *expected cumulative regret*:

$$R_n = \sum_{i=1}^{n} x_i^\mathsf{T}\theta^* - \min_{x \in \mathcal{D}} x^\mathsf{T}\theta^*$$

- Assume: the "best action" function $G(\theta) := \arg\min_{x \in D}\{\theta_{md}^\mathsf{T} x\}$ is smooth

[1] Rusmevichientong, Paat, and John N. Tsitsiklis. "Linearly parameterized bandits". *Mathematics of Operations Research*, 2010.

PEGE Algorithm with online GD

Input and Initialisation

Get a basis $\{b_1, \ldots, b_d\} \in D$ for $\mathbb{R}^d$.

Set $c = \frac{4d}{3\lambda_{\min}(\sum_{i=1}^{d} b_i b_i^\mathsf{T})}$ and $\theta_0 = 0$.

For $m = 1, 2, \ldots$ **do**

Exploration Phase

For $n = (m-1)d$ **to** $md - 1$

1. Choose arm $x_n = b_{n \bmod md}$ and observe y_n.
2. Update θ as follows: $\theta_n = \theta_{n-1} + \frac{c}{n}((y_j - \theta_{n-1}^\mathsf{T} x_j)x_j)$, where $j \sim \mathcal{U}(1, \ldots, n)$.

Exploitation Phase

Find $x = G(\theta_{md}) := \arg\min_{x \in D}\{\theta_{md}^\mathsf{T} x\}$.

Choose arm x m times consecutively.

Regret bound

We require the following extra assumptions from [Rusmevichientong 2010]

(A3') A basis $\{b_1, \dots, b_d\} \in \mathcal{D}$ for $\mathbb{R}^d$ is made known to the algorithm.

(A4) The function $G : \theta \to \arg\min_{x \in \mathcal{D}}\{\theta^\top x\}$ is J-Lipschitz.

Theorem

Under (A1), (A2), (A3'), and (A4), the cumulative regret R_n satisfies

$$R_n \le C_1(\|\theta^*\|_2 + \|\theta^*\|_2^{-1})h_3(n)dn^{1/2},$$

where constant C_1 depends on $\lambda_{\min}(\sum_{i=1}^d b_i b_i^\top)$ and J, and h_3 hides log factors.

Outline

1. Bandits with strongly convex arms
   - Random online algorithm for OLS
   - Regret bounds
2. Bandits with non-strongly convex arms
   - Random online-regularized algorithm
   - Regret bounds
3. Conclusions

ADAPTIVE REGULARIZATION

Problem: In many setting, $\lambda_{\min}(\bar{A}_n) \geq \mu$ may not hold.
Solution: adaptively regularize with λ_n

$$\tilde{\theta}_n := \arg\min_{\theta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \theta^\mathsf{T} x_i)^2 + \lambda_n \|\theta\|_2^2$$

RANDOM ONLINE-REGULARIZED ALGORITHM

Shadow the solutions $\tilde{\theta}_n$ more and more closely as $n \to \infty$ using θ_n as:

$$\theta_n = \theta_{n-1} + \gamma_n((y_{i_n} - \theta_{n-1}^\mathsf{T} x_{i_n})x_{i_n} - \lambda_n \theta_{n-1}), \text{ where } i_n \sim \mathcal{U}(1, \ldots, n). \quad (3)$$

Error bound

For the bandit application, we need to bound $\theta_n - \theta^*$ in the A_n norm, where $A_n = \sum_{i=1}^{n-1} x_i x_i^\mathsf{T} + n\lambda_n I_d$.

Theorem

Under (A1)-(A2), with $\theta_0 = 0$ and step-sizes $\gamma_n = \dfrac{c}{n^\alpha}$ with $c > \dfrac{1}{2\mu}$ and regularisation parameter $\lambda_n = \mu/n^{1-\alpha}$, with $\alpha \in (1/2, 1)$, we have for any $\delta > 0$

$$P\left(\|\theta_n - \theta^*\|_{A_n,2} \leq \kappa_n + \beta'_n\right) \geq 1 - \delta,$$

where $\kappa_n = \sqrt{\dfrac{K_{\mu,c}}{n^{2\alpha-1}} \log \dfrac{1}{\delta}} + \left(\dfrac{C_{\theta^*}}{\sqrt{n}} + \sqrt{\dfrac{\beta_n}{n}} + \dfrac{h_2(n)}{n^{-\alpha/4+1/2}} + \dfrac{h_1(n)}{n^{\alpha-1/2}}\right)$, C_{θ^*} *bounds* $\|\theta^*\|_2$, $h_2(n) = 2(\sqrt{\beta_n} n^{-\alpha/4} + 1)$ *and* $\beta_n = \max\left(128 d \log n \log \dfrac{n^2}{\delta}, \left(\dfrac{8}{3} \log \dfrac{n^2}{\delta}\right)^2\right)$.

ConfidenceBall with online GD [2]

Input and Initialisation
Choose μ and c so that $\mu c > 1/2$, $\alpha \in (1/2, 1)$ and set $\theta_0 = 0$.

For $n = 1, 2, \ldots$ **do**

1. Construct ellipsoid $B_n^2 = \left\{ v : \|v - \theta_t\|_{A_n,2} \leq \kappa_n + \beta_n' \right\}$
2. Choose x_n that gives the minimum estimated loss over B_n^2, i.e., $x_n = \arg\min_{x \in \mathcal{D}} \min_{v \in B_n^2} v^\mathsf{T} x$
3. Observe loss y_t.
4. Update θ_n using random online-regularized algorithm:

$$\theta_n = \theta_{n-1} + cn^{-\alpha}((y_{i_n} - \theta_{n-1}^\mathsf{T} x_{i_n}) x_{i_n} - \mu n^{\alpha-1} \theta_{n-1})$$

[2] Dani, Varsha, Thomas P. Hayes, and Sham M. Kakade. "Stochastic Linear Optimization under Bandit Feedback." COLT. 2008.

Regret bound

Theorem

Assuming an upper bound, C_{θ^*}, for $\|\theta^*\|_2$ is known and under (A1), (A2), and (A4'), with $\gamma_n = c/n^\alpha$ and $\lambda_n = \mu/n^{1-\alpha}$ where $\alpha = 4/5$, the cumulative regret R_T satisfies

$$R_T \leq 2d\sqrt{\ln T}\; T^{1/2+1/5} \text{ w.p. } 1-\delta.$$

Note:

- A vanilla confidence ball algorithm has a complexity $O(nd^2)$ per time step, whereas our proposed enhancement has complexity $O(nd)$.
- However, this comes at a loss of $n^{1/5}$ in the regret R_n.

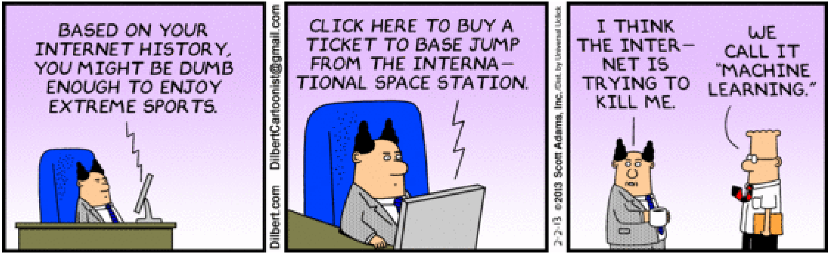
Outline

1. Bandits with strongly convex arms
   - Random online algorithm for OLS
   - Regret bounds
2. Bandits with non-strongly convex arms
   - Random online-regularized algorithm
   - Regret bounds
3. Conclusions

Conclusions

- We proposed two schemes with randomisation for solving least squares
  - The first algorithm assumed strong convexity, while second uses adaptive regularisation.
  - We provide bounds on the error both in expectation and high probability
- We apply our schemes to the linear bandit algorithms PEGE and ConfidenceBall
  - In both settings, there is a significant gains in complexity.
  - While there is no loss in regret for PEGE, in the ConfidenceBall algorithm there is a deterioration of $O(n^{1/5})$ in the regret.
- Future work:
  - Whether the gap in the regret bound for ConfidenceBall algorithm can be eliminated?
  - Experiments on news-feed application - coming soon!

WHAT NEXT?


BASED ON YOUR INTERNET HISTORY, YOU MIGHT BE DUMB ENOUGH TO ENJOY EXTREME SPORTS.
CLICK HERE TO BUY A TICKET TO BASE JUMP FROM THE INTERNA–TIONAL SPACE STATION.
I THINK THE INTER–NET IS TRYING TO KILL ME.
WE CALL IT "MACHINE LEARNING."
Dilbert.com DilbertCartoonist@gmail.com
©2013 Scott Adams, Inc.

References I

Prashanth L.A., Nathaniel Korda and Rémi Munos,
Analysis of stochastic approximation for efficient least squares regression and LSTD.
arXiv preprint arXiv:1306.2557, 2013.

Nathaniel Korda, Prashanth L.A. and Rémi Munos,
Online gradient descent for least squares regression: Non-asymptotic bounds and application to bandits.
arXiv preprint arXiv:1307.3176, 2013.