Discounted online Newton method for time-varying time series prediction

Dongsheng Ding

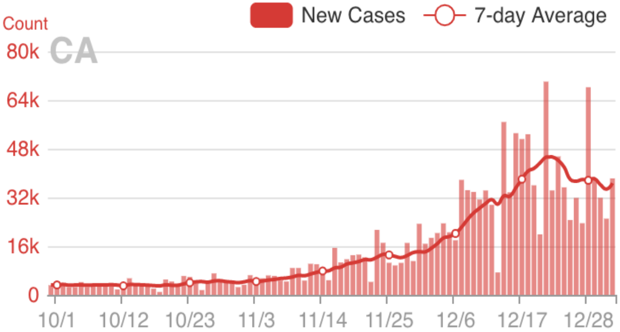
a joint work with

Jianjun Yuan, Mihailo Jovanović

2021 American Control Conference

Motivation

- CALIFORNIA COVID NEW CASES



⋆ Time-dependent statistics, e.g., mean, variance, and covariance

coronavirus.1point3acres.com

Online prediction

- CLASSICAL SETUP
  - ⋆ Batch data
  - ⋆ Static model
  - ⋆ Fix noise distribution, e.g., Guassian
  - ⋆ Quadratic loss function

- ONLINE LEARNING
  - ⋆ Streaming data
  - ⋆ Adaptive model
  - ⋆ Unknown noise distribution, often adversarial
  - ⋆ General loss function

Anava, Hazan, Mannor, Shamir, COLT ’13

Time-varying ARIMA model

- ARIMA (p, d, q)

$$\nabla^d X_t \;=\; \sum_{i=1}^{p} \alpha_t^i \, \nabla_t^d X_{t-i} \;+\; \sum_{j=1}^{q} \beta_t^j \, \epsilon_{t-j} \;+\; \epsilon_t$$

  - ⋆ $\nabla^d X_t = \nabla^{d-1} X_t - \nabla^{d-1} X_{t-1}$ – dth order difference
  - ⋆ $\alpha_t := (\alpha_t^1, \ldots, \alpha_t^p)$, $\beta_t := (\beta_t^1, \ldots, \beta_t^q)$ – model parameters
  - ⋆ ϵ_t – zero-mean noise

Time-varying ARIMA model

- ARIMA (p, d, q)

$$\nabla^d X_t \;=\; \sum_{i=1}^{p} \alpha_t^i \, \nabla_t^d X_{t-i} \;+\; \sum_{j=1}^{q} \beta_t^j \, \epsilon_{t-j} \;+\; \epsilon_t$$

  - ⋆ $\nabla^d X_t = \nabla^{d-1} X_t - \nabla^{d-1} X_{t-1}$ – dth order difference
  - ⋆ $\alpha_t := (\alpha_t^1, \ldots, \alpha_t^p)$, $\beta_t := (\beta_t^1, \ldots, \beta_t^q)$ – model parameters
  - ⋆ ϵ_t – zero-mean noise

known α_t, β_t and observable ϵ_t

$$\hat{X}_t(\alpha_t, \beta_t) \;=\; \nabla^d \hat{X}_t \;+\; \sum_{k=0}^{d-1} \nabla^k X_{t-1}$$

  - ⋆ $\nabla^d \hat{X}_t := \sum_{i=1}^{p} \alpha_t^i \nabla_t^d X_{t-i} + \sum_{j=1}^{q} \beta_t^j \epsilon_{t-j}$

unknown α_t, β_t and unobservable ϵ_t

Online ARIMA prediction

- LEARNING PROTOCOL

  0. environment picks α_t, β_t
  1. environment generates ϵ_t and X_t using the ARIMA model
  2. player predicts $\hat{X}_t$, e.g., $\hat{X}_t(\hat{\alpha}_t, \hat{\beta}_t)$
  3. player observes X_t and suffers loss

$$\ell_t(\hat{\alpha}_t, \hat{\beta}_t) \; := \; \ell_t(X_t, \hat{X}_t(\hat{\alpha}_t, \hat{\beta}_t))$$

Online ARIMA prediction

- Learning protocol

  0. environment picks α_t, β_t
  1. environment generates ϵ_t and X_t using the ARIMA model
  2. player predicts $\hat{X}_t$, e.g., $\hat{X}_t(\hat{\alpha}_t, \hat{\beta}_t)$
  3. player observes X_t and suffers loss

$$\ell_t(\hat{\alpha}_t, \hat{\beta}_t) \; := \; \ell_t(X_t, \hat{X}_t(\hat{\alpha}_t, \hat{\beta}_t))$$

- Dynamic regret

$$\mathcal{R}_T^{\epsilon} \; = \; \sum_{t=1}^{T} \left(\ell_t(\hat{\alpha}_t, \hat{\beta}_t) \; - \; \min_{\alpha, \beta} \ell_t(\alpha, \beta) \right)$$

not computable ℓ_t

Improper learning

- APPROXIMATE ARIMA (p, d, q) BY ARIMA $(p+m, d, 0)$

$$\hat{X}_t(\hat{\theta}_t) \;=\; \sum_{i=1}^{p+m} \hat{\theta}_t^i \, \nabla^d X_{t-i} \,+\, \sum_{k=0}^{d-1} \nabla^k X_{t-1}$$

$\hat{\theta}_t := (\hat{\theta}_t^1, \ldots, \hat{\theta}_t^{p+m})$ – predicted model parameters

$\ell_t(\hat{\theta}_t) := \ell_t(X_t, \hat{X}_t(\hat{\theta}_t))$ – computable

Improper learning

- APPROXIMATE ARIMA (p, d, q) BY ARIMA $(p + m, d, 0)$

$$\hat{X}_t(\hat{\theta}_t) = \sum_{i=1}^{p+m} \hat{\theta}_t^i \nabla^d X_{t-i} + \sum_{k=0}^{d-1} \nabla^k X_{t-1}$$

$\hat{\theta}_t := (\hat{\theta}_t^1, \ldots, \hat{\theta}_t^{p+m})$ – predicted model parameters

$\ell_t(\hat{\theta}_t) := \ell_t(X_t, \hat{X}_t(\hat{\theta}_t))$ – computable

- (PRACTICAL) DYNAMIC REGRET

$$\mathcal{R}_T = \sum_{t=1}^{T} \left(\ell_t(\hat{\theta}_t) - \min_{\alpha, \beta} \ell_t(\alpha, \beta) \right)$$

Discounted online Newton step

- PREDICT $\hat{X}_t$

$$\hat{X}_t(\hat{\theta}_t) \;=\; \sum_{i=1}^{p+m} \hat{\theta}_t^i \, \nabla^d X_{t-i} \;+\; \sum_{k=0}^{d-1} \nabla^k X_{t-1}$$

Discounted online Newton step

- PREDICT $\hat{X}_t$

$$\hat{X}_t(\hat{\theta}_t) \;=\; \sum_{i=1}^{p+m} \hat{\theta}_t^i \, \nabla^d X_{t-i} \;+\; \sum_{k=0}^{d-1} \nabla^k X_{t-1}$$

- OBSERVE X_t AND SUFFER LOSS

$$\ell_t(\hat{\theta}_t) \;=\; \ell_t(X_t, \hat{X}_t(\hat{\theta}_t))$$

Discounted online Newton step

- PREDICT $\hat{X}_t$

$$\hat{X}_t(\hat{\theta}_t) \;=\; \sum_{i=1}^{p+m} \hat{\theta}_t^i \, \nabla^d X_{t-i} \;+\; \sum_{k=0}^{d-1} \nabla^k X_{t-1}$$

- OBSERVE X_t AND SUFFER LOSS

$$\ell_t(\hat{\theta}_t) \;=\; \ell_t(X_t, \hat{X}_t(\hat{\theta}_t))$$

- NEWTON STEP $\hat{\theta}_{t+1}$ WITH DISCOUNT FACTOR γ

$$\hat{\theta}_{t+1} \;=\; \Pi_{\mathcal{S}}^{P_t}\left(\hat{\theta}_t \;-\; \frac{1}{\eta} P_t^{-1} \nabla_t\right)$$

  - ⋆ $\nabla_t = \nabla \ell_t(\hat{\theta}_t)$ – gradient
  - ⋆ $P_t = (1-\gamma) P_0 + \gamma P_{t-1} + \nabla_t \nabla_t^\top$ – estimated Hessian for $\ell_t(\hat{\theta}_t)$

Dynamic regret bound

- ASSUMPTIONS
  - ⋆ noise ϵ_t, model parameters α_t, β_t are bdd
  - ⋆ loss function ℓ_t is smooth, exp-concave
  - ⋆ bounded path length $\sum_{t=2}^{T} \|\phi_t - \phi_{t-1}\| \leq V$

$\phi_t = \mathrm{argmin}_{\theta \in \mathcal{S}} \, \ell_t(\theta)$

- REGRET BOUND

$$\mathcal{R}_T \;\leq\; -b_1 T \log\gamma \;-\; b_1 \log(1-\gamma) \;+\; \frac{b_2}{1-\gamma} V \;+\; b_3$$

$m = O(q \log(T))$

b_1, b_2, b_3 – constants

γ – discount factor

Different discount factors

- UNKNOWN PATH LENGTH V

$$\mathcal{R}_T \leq O(T^{1-s} + T^s V)$$

  - ⋆ $\gamma = 1 - T^{-s}$, $s \in (0,1)$

Static regret $O(T^{1-s})$ when $V = 0$

Different discount factors

- UNKNOWN PATH LENGTH V

$$\mathcal{R}_T \;\leq\; O(\,T^{1-s} \,+\, T^s V\,)$$

  - $\star$ $\gamma = 1 - T^{-s}$, $s \in (0,1)$

Static regret $O(T^{1-s})$ when $V = 0$

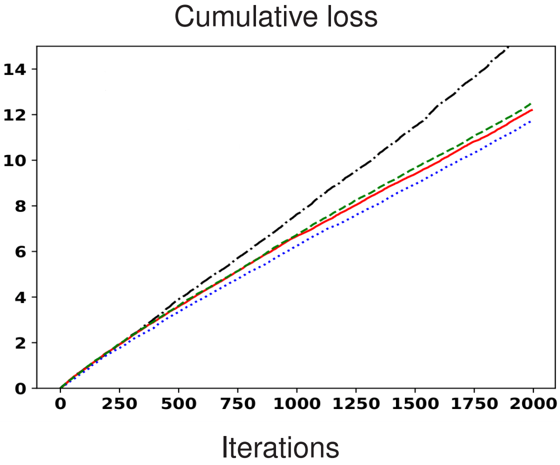
- KNOWN PATH LENGTH V

$$\mathcal{R}_T \;\leq\; \max\Big(\,O(\,\log T\,),\; O(\,\sqrt{TV}\,)\,\Big)$$

  - $\star$ $\gamma = 1 - O(\sqrt{V/T})$

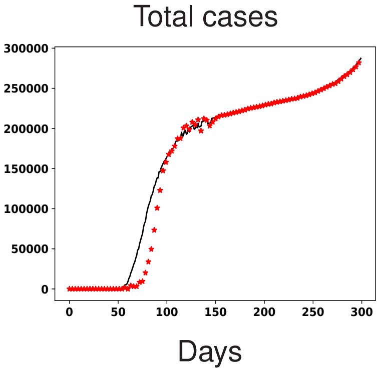
Static regret $O(\log T)$ when $V = 0$

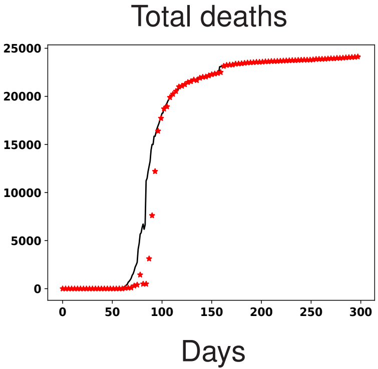
Synthetic data



- ⋆ OGD (Liu, et al. '16) (– - –)
- ⋆ D-ONS with: $\gamma = 0.98$ (—), $\gamma = 0.5$ (····), and $\gamma = 0.1$ (– –).

NYC COVID-19 Data





- ⋆ NYC data from 1/23/2020 to 11/15/2020: real observation (−)
- ⋆ D-ONS's prediction (⋆⋆) (display for every 3 days)

healthdata.org

Summary

- RESULTS
  - ⋆ Discounted online Newton method
  - ⋆ Dynamic regret
- ONGOING EFFORT
  - ⋆ Multi-step prediction
  - ⋆ Automatic parameter tuning

Thank You !