Foundations of machine learning

Online convex optimization

Maximilian Kasy

Department of Economics, University of Oxford

Hilary term 2023

Outline

- Setup of online convex optimization:
  - Iteratively choose x_t.
  - Observe loss $f_t(x_t)$ and gradient $\nabla f_t(x_t)$.
- Baseline algorithm:
  Online gradient descent (OGD).
- Adversarial regret guarantee for OGD.
- Connection to related settings:
  - Adversarial online learning.
  - Stochastic gradient descent.
  - Multiarmed bandits.

Takeaways for this part of class

- Online convex optimization provides a natural framework to connect learning theory and optimization theory.
- Adversarial regret guarantee:
  Adversarial regret grows at a rate of $\sqrt{T}$.
- Other settings can be reduced to online convex optimization:
  - Stochastic gradient descent:
    Adversarial bounds imply stochastic bounds.
    Return average of $\boldsymbol{x}_t$ at the end.
  - Bandit settings:
    Form unbiased estimators of loss and gradients using inverse probability weighting.

Online gradient descent

Connection to other learning problems

References

Setup

- Sequential choices $\boldsymbol{x_t} \in \mathscr{K}$, where $\mathscr{K}$ is convex.
- Convex loss functions $\boldsymbol{f_t}(\cdot)$.
- Observable, after choice of $\boldsymbol{x_t}$:
  - Cost $f_t(x_t)$.
  - Gradient $\nabla f_t(x_t)$.
- Regret:

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^*),$$

where

$$x^* = \operatorname*{argmin}_{x \in \mathscr{K}} \sum_{t=1}^{T} f_t(x).$$

Online gradient descent

- For each $t = 1$ to T:
  1. Play x_t.
  2. Observe $\nabla_t = \nabla f_t(x_t)$.
  3. Update with a gradient step:
  $$y_{t+1} = x_t - \eta_t \cdot \nabla_t.$$
  4. Project into $\mathscr{K}$:
  $$x_{t+1} = \Pi_{\mathscr{K}} y_{t+1}.$$
- The stepsizes η_t are tuning parameters, to be specified.

Adversarial regret bound

Theorem

- Consider online gradient descent with step-sizes

$$\eta_t = \frac{D}{G\sqrt{t}},$$

  where

$$\|x-y\| \leq D \quad \forall\, x,y \in \mathscr{K}\,, \qquad \|\nabla f(x)\| \leq G \quad \forall\, x \in \mathscr{K}\,.$$

- Then:

$$R_T \leq \frac{3}{2} GD\sqrt{T}.$$

Proof

- By convexity of f_t:

$$f_t(x_t) - f_t(x^*) \leq \nabla_t \cdot (x_t - x^*).$$

- By orthogonal projection:

$$\|x_{t+1} - x^*\| \leq \|y_{t+1} - x^*\|.$$

- By definition of gradient update:

$$\|y_{t+1} - x^*\|^2 = \|x_t - x^*\|^2 + \eta_t^2 \|\nabla_t\|^2 - 2\eta_t \nabla_t \cdot (x_t - x^*).$$

- Rearrange. By upper bound on ∇_t:

$$2\nabla_t \cdot (x_t - x^*) \leq \frac{\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2}{\eta_t} + \eta_t G^2.$$

Proof continued

Collect bounds and sum across t:

$$
\begin{aligned}
2R_t &\le 2\sum_t \nabla_t \cdot (x_t - x^*) \\
&\le \sum_t \left[\frac{\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2}{\eta_t} + \eta_t G^2 \right] \\
&\le \sum_t \|x_t - x^*\|^2 \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) + G^2 \cdot \sum_t \eta_t && (1/\eta_0 = 0, \|x_{T+1} - x^*\|^2 \ge 0) \\
&\le D^2 \frac{1}{\eta_T} + G^2 \cdot \sum_t \eta_t && \text{(telescoping series)} \\
&\le 3DG\sqrt{T} && \text{(definition of } \eta_t\text{; } \sum_{t=1}^{T} 1/\sqrt{t} \le 2\sqrt{T}\text{)}.
\end{aligned}
$$

$\square$

Online gradient descent

Connection to other learning problems

References

Online learning

- Recall the online learning problem:
  - Expert predictions $\hat{Y}_{h,t}$.
  - Loss $L(\hat{Y}_t, Y_t)$.
- Map into online convex optimization:
  - Weight vector $x_t = (x_{h,t})$ in the simplex $\mathscr{K}$.
  - Prediction:
$$\hat{Y}_t = \sum_h x_{h,t} \cdot \hat{Y}_{h,t}.$$
  - Gradient:
$$\nabla_t = \left(\hat{Y}_{h,t}\right)_h \cdot \partial_{\hat{Y}} L(\hat{Y}_t, Y_t).$$

Stochastic gradient descent

- Recall the stochastic optimization setting:
  - Our goal is to minimize $f(x)$ w.r.t. x.
  - We observe unbiased gradient estimates ∇_t:

$$E[\nabla_t | x_t] = \nabla f(x_t).$$

  Think: $\nabla_t = \nabla m(x, Z_t)$.

- Stochastic gradient descent:
  1. Apply online gradient descent.
  2. Return $\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$.

Regret bound for stochastic gradient descent

Assume $E[\|\nabla_t\|^2] \leq G^2$. Then

$$E[f(\bar{x}_T)] - f(x^*) \leq \frac{3GD}{2\sqrt{T}}.$$

Sketch of proof:

$$\begin{aligned}
E[f(\bar{x}_T)] - f(x^*) &\leq \frac{1}{T}\sum_{t=1}^{T} E\left[f(x_t) - f(x_*)\right] && \text{(convexity)} \\
&\leq \frac{1}{T}\sum_{t=1}^{T} E\left[\nabla_t \cdot (x_t - x_*)\right] && (E[\nabla_t | x_t] = \nabla f(x_t)) \\
&\leq \frac{3GD}{2\sqrt{T}}. && \text{(Theorem for OGD)}
\end{aligned}$$

□

Multi-armed bandits

- Coming up next in class.
- Only observe loss L_t for actions actually chosen.
- For randomized algorithms, we can form unbiased estimators of the gradient of reward:

$$\nabla_t = \left(L_t \cdot \frac{\mathbf{1}(D_t = d)}{x_{d,t}} \right)_d$$

- This allows us to reduce the adversarial bandit problem to an online convex optimization problem.

References

Hazan, E. (2016). Introduction to online convex optimization. Foundations and Trends in Optimization, *2(3-4):157–325, chapter 3.*