Aug 5, 2020

Contextual Bandits

- in many Bandit problems, learner possesses extra/side information to "predict" quality of actions.
- all algorithms + regret def^n thus far ignore these contextual data.
- here we look at <u>better models</u>

- Eg: movie recommendation...

Interaction Protocol

- Adversary secretly chooses $(x_t)_{t=1}^{n}$, $x_t \in [0,1]^k$
- Adversary secretly chooses $(c_t)_{t=1}^{n}$, $c_t \in \mathcal{C}$ arbitrary, fixed
- for $t=1, \dots, n$

  learner observes c_t

  learner selects $P_t \in \mathcal{P}_{k-1}$, $A_t \sim P_t$

  learner observes $X_t = x_{tA_t}$

Regret:
$$R_n = \mathbb{E}\left[\sum_{c \in \mathcal{C}} \max_{i \in [k]} \sum_{t, c_t = c} (x_{ti} - X_t)\right]$$

$$R_{nc} = \mathbb{E}\left[\max_{i} \sum_{t, c_t = c} (x_{ti} - X_t)\right]$$

Scanned with CamScanner

if exp3 is used for each context seperately,

$$R_{nc} \leq 2\sqrt{k \log k \sum_{t=1}^{n} \mathbb{1}\{c_t = c\}}$$

$$R_n \leq 2 \sum_{c \in C} \sqrt{k \log k \sum_{t} \mathbb{1}\{c_t = c\}}$$

- if $|C| = 1$, then same as adv. bandit
- if all $c \in C$ are equally likely

$$R_n \leq 2\sqrt{nk|C| \log k}$$

Bandits w/ expert advice

- if $|C|$ is large, then exp3 on each context not useful, unless n is enormous.
- however, C is "structured" in real-life
- ex: movie recommendation - users with similar demographics have similar preferences w high likelihood

let Φ be set of all functions from $C \to [k]$

$$R_n = \mathbb{E}\left[\max_{\phi \in \Phi} \sum_{t=1}^{n} x_{t\phi(c_t)} - X_t\right]$$

- if Φ is small, we can get better reward.

1 $\boxed{\text{partitions}}$

- let $\mathcal{P} \subset 2^C$ be a partition of C
- define Φ as set of functions from $C \to [k]$ s.t, they are constant on each part in $\mathcal{P}$
- now, if exp3 is run on each part, regret $\approx |\mathcal{P}|$

2 $\boxed{\text{similarity functions}}$

let $s: C \times C \to [0,1]$ similarity b/w pairs of contexts

let Φ be set of $C \to [k]$ s.t "average dissimilarity" is below a threshold $\theta \in (0,1)$

$$\frac{1}{|C|^2} \sum_{c,d \in C} \left(1 - s(c,d)\right) \mathbb{1}\{\phi(c) \neq \phi(d)\}$$

- not clear how to use exp3, but regret will be small

3 $\boxed{\text{Supervised learning to expert advice}}$

- train on batch data to find $\Phi_1, \ldots, \Phi_M : C \to [k]$
- then use bandit algo to compete w/ best of these in an online fashion

4. ---- many more

Bandits w/ experts framework

- k arms
- M experts predict most promising action in each round. (generally prob. distributions)

- predictions of M experts in round t,

$E^{(t)} \in [0,1]^{M \times k}$, $E_m^{(t)}$ - row vector of probability reported by m^{th} expert at t

$$R_n = E\left[\max_{m \in [M]} \sum_{t=1}^{n} E_m^{(t)} x_t - X_t\right]$$

[compete w/ best expert]

Algo : Exp 4

- Input n, k, M, η, γ
- let $Q_1 = (1/M, \ldots, 1/M) \in [0,1]^{1 \times M}$
- for $t = 1, \ldots, n$
  - recieve advice $E^{(t)}$
  - choose $A_t \sim P_t$, $P_t = Q_t E^{(t)}$
  - observe $X_t = x_{tA_t}$
  - estimate $\hat{X}_{ti} = 1 - \frac{\mathbb{1}\{A_t = i\}}{P_{ti} + \gamma}(1 - X_t)$
  - propagate $\tilde{X}_t = E^{(t)} \hat{X}_t$
  - update $Q_{t+1,i} = \dfrac{\exp(\eta \tilde{X}_{ti}) Q_{ti}}{\sum_j \exp(\eta \tilde{X}_{tj}) Q_{tj}}$

Thm: Let $\gamma = 0$, $\eta = \sqrt{2(\log M)/nk}$. Then

$$R_n \leq \sqrt{2nk \log M}$$