Statistical Learning Theory: Weighted Experts and Bandits

CS4780/5780 – Machine Learning
Fall 2014

Thorsten Joachims
Cornell University

Reading: Mitchell Chapter 7.5

Expert Learning Model

- Setting
  - N experts named $\mathrm{H} = \{h_1, \ldots, h_N\}$
  - Each expert h_i takes an action $y = h_i(x_t)$ in each round t and incurs loss $\Delta_{t,i}$
  - Algorithm can select which expert's action to follow in each round
- Interaction Model
  - FOR t from 1 to T
    - Algorithm selects expert h_{i_t} according to strategy $A(w_t)$ and follows its action y
    - Experts incur losses $\Delta_{t,1} \ldots \Delta_{t,N}$
    - Algorithm incurs loss Δ_{t,i_t}
    - Algorithm updates w_t to w_{t+1} based on $\Delta_{t,1} \ldots \Delta_{t,N}$

Regret

- Idea
  - N experts named $\mathrm{H} = \{h_1, \ldots, h_N\}$
  - Compare performance of A to best expert i^* in hindsight.
- Regret
  - Overall loss of best expert i^* in hindsight is

$$\Delta_T^* = \min_{i^* \in [1..N]} \sum_{t=1}^{T} \Delta_{t,i^*}$$

  - Loss of algorithm A at time t is

$$\Delta_{t,i}$$

    for algorithm that picks recommendation of expert $i = A(w_t)$ at time t.
  - Regret is difference between loss of algorithm and best fixed expert in hindsight

$$Regret(T) = \sum_{t=1}^{T} \Delta_{t,A(w_t)} - \min_{i^* \in [1..N]} \sum_{t=1}^{T} \Delta_{t,i^*}$$

Weighted Majority Algorithm

- Setting
  - N experts named $\mathrm{H} = \{h_1, \ldots, h_N\}$
  - Binary actions $y = \{+1, -1\}$ given input x, zero/one loss
  - There may be no expert in H that acts perfectly
- Algorithm
  - Initialize $\mathrm{w}_1 = (1,1, \ldots, 1)$
  - FOR t = 1 TO T
    - Predict the same y as majority of $h_i \in H$, each weighted by $w_{t,i}$
    - FOREACH $h_i \in H$
      - IF h_i incorrect THEN $w_{t+1,i} = w_{t,i} * \beta$
        ELSE $w_{t+1,i} = w_{t,i}$
- Mistake Bound
  - How close is the number of mistakes the Weighted Majority Algorithm makes to the number of mistakes of the best expert in hindsight?

Exponentiated Gradient Algorithm for Expert Setting (EG)

- Setting
  - N experts named $\mathrm{H} = \{h_1, \ldots, h_N\}$
  - Any actions, any loss function
  - There may be no expert in H that acts perfectly
- Algorithm
  - Initialize $w_1 = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$
  - FOR t from 1 to T
    - Algorithm randomly picks i_t from $P(I_t = i_t) = w_{t,i}$
    - Experts incur losses $\Delta_{t,1} \ldots \Delta_{t,N}$
    - Algorithm incurs loss Δ_{t,i_t}
    - Algorithm updates w for all experts i as
      $\forall i, w_{t+1,i} = w_{t,i} \exp(-\eta \Delta_{t,i})$
      Then normalize w_{t+1} so that $\sum_j w_{t+1,j} = 1$.

Expected Regret

- Idea
  - Compare performance to best expert in hindsight
- Regret
  - Overall loss of best expert i^* in hindsight is

$$\Delta_T^* = \min_{i^* \in [1..N]} \sum_{t=1}^{T} \Delta_{t,i^*}$$

  - Expected loss of algorithm $A(w_t)$ at time t is

$$E_{A(w_t)}[\Delta_{t,i}] = w_t \Delta_t$$

    for randomized algorithm that picks recommendation of expert i at time t with probability $w_{t,i}$.
  - Regret is difference between expected loss of algorithm and best fixed expert in hindsight

$$ExpectedRegret(T) = \sum_{t=1}^{T} w_t \Delta_t - \min_{i^* \in [1..N]} \sum_{t=1}^{T} \Delta_{t,i^*}$$

Regret Bound for Exponentiated Gradient Algorithm

- Theorem

  The expected regret of the exponentiated gradient algorithm in the expert setting is bounded by

$$\text{Expected}Regret(T) \leq \Delta^{\max}\sqrt{2\,T\,log(|H|)}$$

where $\Delta^{\max} = max\{\Delta_{t,i}\}$ and $\eta = \frac{\sqrt{\log(N)}}{\Delta\sqrt{2T}}$.

Bandit Learning Model

- Setting
  - N bandits named $\mathrm{H} = \{h_1, \dots, h_N\}$
  - Each bandit h_i takes an action in each round t and incurs loss $\Delta_{t,i}$
  - Algorithm can select which bandit's action to follow in each round
- Interaction Model
  - FOR t from 1 to T
    - Algorithm selects expert h_{i_t} according to strategy A_{w_t} and follows its action y
    - Bandits incur losses $\Delta_{t,1}$ … $\Delta_{t,N}$
    - Algorithm incurs loss Δ_{t,i_t}
    - Algorithm updates w_t to w_{t+1} based on Δ_{t,i_t}

Key difference compared to Expert Model

Bar

- Setting
  - N bandits named $\mathrm{H} = \{$
  - Each bandit h_i takes an loss $\Delta_{t,i}$
  - Algorithm can select wh each round
- Interaction Model
  - FOR t from 1 to T
    - Algorithm selects expert h_{i_t} according to strategy A_{w_t} and follows its action y
    - Bandits incur losses $\Delta_{t,1}$ … $\Delta_{t,N}$
    - Algorithm incurs loss Δ_{t,i_t}
    - Algorithm updates w_t to w_{t+1} based on Δ_{t,i_t}

Key difference compared to Expert Model

Exponentiated Gradient Algorithm for Bandit Setting (EXP3)

- Initialize $w_1 = \left(\frac{1}{N}, \dots, \frac{1}{N}\right), \gamma = \min\left\{1, \sqrt{\frac{N \log N}{(e-1)\Delta\mathrm{T}}}\right\}$
- FOR t from 1 to T
  - Algorithm randomly picks i_t with probability $P(i_t) = (1-\gamma)w_{t,i} + \gamma/N$
  - Experts (aka Bandits) incur losses $\Delta_{t,1} \dots \Delta_{t,N}$
  - Algorithm incurs loss Δ_{t,i_t}
  - Algorithm updates w for bandit i_t as
    $$w_{t+1,i_t} = w_{t,i_t} \exp\left(-\eta \Delta_{t,i_t}/P(i_t)\right)$$
    Then normalize w_{t+1} so that $\sum_j w_{t+1,j} = 1$.

Other Online Learning Problems

- Stochastic Experts
- Stochastic Bandits
- Online Convex Optimization
- Partial Monitoring