Foundations of machine learning

Overview of online learning and active learning

Maximilian Kasy

Department of Economics, University of Oxford

Winter 2024

Common framework

- Sequential decisions D_t at times $t = 1, 2, \ldots$:
  Predictions/forecasts, treatment choices, moves in a game, ...
- Decision D_t can depend on the history of observed information up to time $t-1$.
- Decisions result in a period-specific loss $L(D_t, Y_t)$,
  which depends on some variable/vector Y_t.
- The goal is to minimize cumulative loss

$$\sum_t L(D_t, Y_t).$$

- This is often evaluated in terms of regret relative to some optimal decision D^*:

$$\sum_t [L(D_t, Y_t) - L(D^*, Y_t)]$$

Observability

How to evaluate algorithms

What is observable?

1. *Online learning* (e.g. forecasting):
   - Observability does not depend on choices $\Rightarrow$ no motive to experiment/explore!
   - Y_t are observed for past periods t.
   - $\Rightarrow$ Counterfactual loss $L(d, Y_t)$ is known for all values of d.
   - Loss is often given by a function of the prediction error, e.g. $L(D_t, Y_t) = (D_t - Y_t)^2$.
2. *Multi-armed bandits* (e.g. treatment assignment):
   - Observability does depend on choices $\Rightarrow$ there is a motive to experiment/explore! Tradeoff with the motive to "exploit" (do well now).
   - C.f. causal inference / potential outcomes: $D \in \{1, \ldots, k\}$, $Y = (Y^1, \ldots, Y^k)$. We observe only Y^D.
   - $\Rightarrow$ Loss is only observed for the realized choice D_t, but not for any counter-factual choice $d \neq D_t$.
   - Loss is often equal to (minus) realized outcomes, i.e., $L(D_t, Y_t) = -Y_t^{D_t}$.

What is observable? - continued

3. *Online convex optimization*:
   - Like mulit-armed bandits for convex action spaces and loss functions,

     but additionally we observe the gradient ∇_t of loss.
   - Online learning and bandits can be reduced to online convex optimization.
4. *Semi-bandits*
   - Intermediate between online learning and multi-armed bandits.
   - We observe more than just the loss of the realized action,
     but less than the loss for all counterfactual actions.
   - Typically composite decision problems,
     where multiple actions are chosen in the same period
     with cross-constraints, e.g. budget constraints.
   - Each action has its own observed outcome.

What is observable? - continued

5. *Contextual bandits*
   - Similar to multi-armed bandits.
   - But additionally we observe predictors X_t, independently of actions D_t.

   $\Rightarrow$ Targeted treatment assignment.
6. *Reinforcement learning*
   - Similar to contextual bandits, with an additional state X_t observed in each period.
   - But X_t is endogenous to past actions.
     It develops according to a Markov transition kernel, given the previous action and state.
   - This framework leads to Bellman equations.
     Learning involves estimation of the value function.
   - Good actions don't just generate small loss now,
     but also good states next period, and down the road.

Practice problem

For each of these 5 settings
name some examples of economic settings where they might be applied.

Observability

How to evaluate algorithms

Optimal solutions versus the theory of heuristic algorithms

- In principle all of these frameworks can be combined with priors for the underlying parameters.
- This leads to dynamic stochastic optimization problems, where the "states" are posterior beliefs, which theoretically have optimal solutions.
- In practice, these solutions are impossible to compute.
- Economic theory in this space has focused on very stylized models, where solutions might be characterized.
- Modern machine learning has taken another approach: Construct heuristic algorithms for practically relevant settings, and develop (very sophisticated) theory to understand their behavior.
- This is the approach we will take in this class.

Decision theory and alternative evaluation criteria

- In decision theory, we saw different criteria for evaluating decision functions: Risk function, Bayes risk, minimax risk.
- These criteria translate into different theoretical approaches for evaluating online learning / active learning algorithms.
- There are some additional subtleties due to asymptotic approximations, and the dynamic nature of decisions.

1. "Stochastic" models assume that the Y_t are i.i.d. draws from some distribution and characterize behavior conditional on that distribution.
2. "Adversarial" models condition on the sequence of Y_t, and characterize behavior for any possible sequence.

How to evaluate algorithms (1)

1. *i.i.d. draws, fixed parameter*
   - Results characterize the rate of convergence of average regret toward $\mathbf{0}$.
   - Key tool: Large deviations theory.

   $\Rightarrow$ Good characterizations of bandit algorithms for the "high powered" case (large samples and/or large treatment effects).

2. *i.i.d. draws, worst-case parameter*
   - Results characterize the rate of convergence of worst case regret toward $\mathbf{0}$.

   $\Rightarrow$ Good characterization of bandit algorithms for the "low powered" case (smaller samples and/or smaller treatment effects).

How to evaluate algorithms (2)

3. *i.i.d. draws, drifting parameter*
   - Similar to approaches taken in the theory of weak instruments.
   - Key tool: Uniform central limit theorems.
   - Drifting parameter sequences allow to keep the problem equally hard, as sample size increases.

   ⇒ This gives a characterization of the risk function for the full range of parameter values.
4. *Worst-case sequence of outcomes*
   - There is no more probability involved, except possibly in the algorithm.
   - Similar to randomization inference, in this regard.
   - How could any algorithm possibly perform well for all sequences?
   - Key idea: Rather than restricting the data generating process we can restrict the comparison set of alternative decision functions.
   - Related to ideas we saw in PAC learning theory.

Practice problem

Discuss how these approaches for evaluating algorithms relate to the criteria we saw in decision theory.