6.7900 Fall 2024: Lecture Notes 09

1 Online Learning

In previous lectures, we assume the following.

a. All the training data points come at once (offline learning).

b. Each data point is sampled i.i.d. from a fixed distribution P.

c. After obtaining a well-trained ML model, we apply it to a test set where each data point also comes from the same distribution P.

In this lecture, we are going to make statistical assumptions as few as possible. We will focus on the situation where data comes sequentially one at a time. In each round, we have to make an irrevocable prediction (and incur a loss). The correct answer is revealed after each prediction and we can learn from these answers to improve our predictions. Basically, we are relaxing the three assumptions in the standard offline learning setting.

a. Each data point comes in a stream.

b. The data stream is governed by an arbitrary underlying process that we may not know — the nature may even act adversarially to make things unfavorable for us.

c. Each new data point is both a test case when we make a prediction and a training case after the loss is revealed.

Examples of such a sequential decision-making paradigm can be ubiquitous in real life:

1. Clinical trials. We decide which drug should be used each time after sequential observations.
2. Online advertising. We adjust advertisements each time after observing click-through rates.
3. Investment. We re-balance the portfolio each single day after we see price changes.

2 Online Prediction Game

In each round t,

- Nature reveals input x
- Learner makes a prediction on x
- Nature reveals the correct output
- Learner suffers loss
- Learner updates its model

Note again that we do not make any statistical assumptions about the sequence of inputs/labels — the nature may have access to the algorithm and choose labels adversarially. The problem is then *how we should evaluate the performance of an algorithm*.

Definition: We define the **regret** of an algorithm as the difference between the algorithm's performance relative to some benchmark class of methods.

We hope that regret can be growing as a sublinear ($o(T)$) term such that on average the regret goes down to 0 as T grows. Clearly, we have to be careful about both the benchmark as well as the nature. For example, assume we need to predict a label, and the nature is allowed to always choose the label to be the opposite of what we predict in each round. The learner would then suffer $\Omega(T)$ cumulative loss if we compare with the best algorithm with no restrictions.

2.1 Logistic model as an example

In each round t, Learner has a parameter $\theta^t \in B \subseteq \mathbb{R}^d$.

- Nature reveals input x^t
- Learner makes a prediction using $P(y = 1|x^t, \theta^t) = \sigma({\theta^t}^\top x^t)$
- Nature reveals the correct label $y^t \in \{0, 1\}$
- Learner suffers loss $L(x^t, y^t, \theta^t) = -\log P(y^t|x^t, \theta^t)$
- Learner updates the parameter θ^{t+1} for the next round

Learner's regret after T rounds is

$$R_T = \underbrace{\sum_{t=1}^{T} L(x^t, y^t, \theta^t)}_{\text{Learner's cumulative loss}} - \underbrace{\min_{\theta \in B} \sum_{t=1}^{T} L(x^t, y^t, \theta)}_{\text{cumulative loss of the best } \textit{fixed} \text{ predictor}} .$$

2.2 A warm-up exercise

Consider a restricted online classification task where we know at the beginning that a good solution must exist.

a. $\|x^{(i)}\| \leq R$, $y^{(i)} \in \{1, -1\}$ for all i.

b. All data points are linearly separated by a positive margin

$$y^{(i)} \frac{\theta^{*\top} x^{(i)}}{\|\theta^*\|} \geq \gamma.$$

In this case, the best linear predictor incurs a loss of 0. Note that the ratio R/γ characterizes the difficulty of the problem. A simple **Perception** algorithm (see Figure 1) dating back to 1950's solves the problem.

- Start with $\theta^1 = 0$ (vector)
- For t=1,2,3,...

  Nature gives x^t (outside of the margin)

  Learner predicts $\hat{y}^t = \text{sign}\left((\theta^t)^T x^t\right)$

  Nature reveals $y^t = \text{sign}\left((\theta^*)^T x^t\right)$ (nature has to abide by our assumptions)

  If $\hat{y}^t \neq y^t$ (mistake)

  $\theta^{t+1} = \theta^t + y^t x^t$

  else

  $\theta^{t+1} = \theta^t$

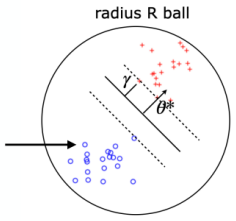

Figure 1: Perception

Theorem: Perception makes at most R^2/γ^2 mistakes and as a result incurs a bounded regret that does not depend on the dimension d.

The gradient update in LR also has a similar form of the update in Perception. **Q: What is the intuition that Perception works with large step-sizes?** In fact, as t grows, $\|\theta^t\|$ also grows, which makes the update $y^t x^t$ create smaller changes towards the decision boundary.

3 Online Convex Optimization

The logistic model and the warm-up exercise described above is a special case of *online convex optimization*. Let B be a set of parameters. For $t = 1, 2, \cdots$

- Learner selects $\theta^t \in B$
- Nature reveals a convex loss function $f_t(\cdot)$
- Learner incurs loss $f_t(\theta^t)$

Learner's goal is to minimize regret

$$R_T = \sum_{t=1}^{T} f_t(\theta^t) - \min_{\theta \in B} \sum_{t=1}^{T} f_t(\theta).$$

We make the following assumptions.

a. (Convex, Bounded) B is convex. $\|B\| \leq R$.

b. (L-Lipschitz) $|f_t(\theta) - f_t(\theta')| \leq L\|\theta - \theta'\|$ for any $\theta, \theta' \in B$.

3.1 Follow the leader

A naive approach is to do **Empirical Risk Minimization (ERM)** in each time:

$$\theta^t = \arg\min_{\theta \in B} \sum_{s=1}^{t-1} f_s(\theta).$$

However, this may not work as opposed to the standard i.i.d. setting.

Example: Let $\theta \in [-1, 1]$, $\theta^1 = 0$. Nature's choices are: $f_1(\theta) = 0.5\theta$ and thereafter for even t, $f_t(\theta) = -\theta$, and for odd t, $f_t(\theta) = \theta$. Learner's losses (following ERM) become: $0, 1, 1, \cdots$. Competitor's cumulative loss is at most 0.5 (we can always choose $\theta^t = 0$).

The example above highlights that we need to come up with new methods to alleviate the situation where the first several wrong decisions lead to very bad outcome.

3.2 Follow the regularized leader

We now discuss a **Projected Gradient** approach (or follow the regularized leader). Initialize $\theta^1 \in B$. For $t = 1, 2, \cdots$

- Nature reveals a convex loss function $f_t(\cdot)$
- Learner incurs loss $f_t(\theta^t)$
- Learner updates the parameter according to projected gradient descent:

$$\hat{\theta}^t = \theta^t - \eta_t \nabla_\theta f_t(\theta)_{|\theta=\theta^t}$$
$$\theta^{t+1} = \arg\min_{\theta} \|\theta - \hat{\theta}^t\|$$

Theorem: Under the assumption that (a) B is convex, (b) $f_t(\cdot)$ is L-Lipschitz for all t, and (c) T is known, **Projected Gradient** with $\eta_t = R/L\sqrt{T}$ yields a regret bound of $RL\sqrt{T}$.

Proof. In class we mentioned the proof idea for the case without projection. Here we show a more formal proof. For simplicity we write $\nabla f_t(\theta_t)$ instead of $\nabla_\theta f_t(\theta)_{|\theta=\theta^t}$. We decompose the regret as:

$$\begin{aligned}
R_T &= \sum_{t=1}^{T} f_t(\theta_t) - f_t(\theta_*) \\
&\le \sum_{t=1}^{T} f_t(\theta_t) - \Big((\theta^* - \theta^t)^\top \nabla f_t(\theta_t) + f_t(\theta_t)\Big) \\
&= \sum_{t=1}^{T} (\theta^t - \theta^*)^\top \nabla f_t(\theta_t) \\
&= \sum_{t=1}^{T} (\theta^t - \theta^*)^\top \frac{\theta^t - \hat{\theta}^{t+1}}{\eta} \\
&= \sum_{t=1}^{T} (\theta^t - \theta^*)^\top \frac{(\theta^t - \theta^*) - (\hat{\theta}^{t+1} - \theta^*)}{\eta}. \qquad (1)
\end{aligned}$$

Meanwhile, we have

$$\begin{aligned}
-(\theta^t - \theta^*)(\hat{\theta}^{t+1} - \theta^*) &= \frac{1}{2}\Big(\|\theta^t - \hat{\theta}^{t+1}\|^2 - \|\theta^t - \theta^*\|^2 - \|\hat{\theta}^{t+1} - \theta^*\|^2\Big) \\
&= \frac{1}{2}\Big(\|\eta \nabla f_t(\theta_t)\|^2 - \|\theta^t - \theta^*\|^2 - \|\hat{\theta}^{t+1} - \theta^*\|^2\Big) \\
&\le \frac{\eta^2 L^2}{2} - \frac{\|\theta^t - \theta^*\|^2 + \|\hat{\theta}^{t+1} - \theta^*\|^2}{2} \\
&\le \frac{\eta^2 L^2}{2} - \frac{\|\theta^t - \theta^*\|^2 + \|\theta^{t+1} - \theta^*\|^2}{2}. \qquad (2)
\end{aligned}$$

The last inequality holds because $\|\theta^{t+1} - \theta^*\| \le \|\hat{\theta}^{t+1} - \theta^*\|$. (*Why? Try to convince yourself intuitively based on the fact that B is convex.*) Plugging (2) into (1) and setting $\eta = R/L\sqrt{T}$ yield

$$\begin{aligned}
R_T &\le \sum_{t=1}^{T} \frac{\|\theta^t - \theta^*\|^2}{\eta} + \frac{\eta L^2}{2} - \frac{\|\theta^t - \theta^*\|^2 + \|\theta^{t+1} - \theta^*\|^2}{2\eta} \\
&= \frac{\|\theta^1 - \theta^*\|^2 - \|\theta^{T+1} - \theta^*\|^2}{2\eta} + \frac{\eta T L^2}{2} \\
&\le \frac{R^2}{2\eta} + \frac{\eta T L^2}{2} = RL\sqrt{T}.
\end{aligned}$$

Exercise: What if we do not know T? Can you show that setting $\eta_t = R/L\sqrt{t}$ also leads to a regret bound of $C \cdot RL\sqrt{T}$ (C is an absolute constant)?