EE194/290S: ML for Sequential Decision under Uncertainty **Fall 2018**

Lecture 8: Online Optimization 1

Lecturer: Anant Sahai/Vidya Muthukumar *Scribes: Ashray Manepalli, Wesley Wang*

8.1 FTRL and FTPL Recap

As we have seen thus far in our discussions on sequential decision making, multiplicative weights for the Follow the Leader (FTL) scheme can be formalized in two ways:

1. A optimization problem utilizing entropy regularization, $\frac{1}{\eta}H(\vec{w})$, to pull us towards more 'random' solutions, and
2. A perturbation based approach that adds random perturbation $N_{t,x} \sim G(0, \frac{1}{\eta})$ to our loss.

We analyzed these alternate schemes, Follow the Regularized Leader (FTRL) and Follow the Perturbed Leader (FTPL) respectively, in lectures 6 and 7. We found that these schemes get us sub-linear total regret bounds, $R_T \in \mathcal{O}(\sqrt{T})$, which means that average regret is asymptotically pushed to 0.

Recall that in the analyses, regret bounds were proved using the concept of "Leader Change": $\mathbb{1}[\hat{x}_{t+1} \neq \hat{x}_t]$. This is certainly a powerful method of analyzing discrete problems in sequential decision making with relatively simple loss functions such as Hamming Distance. However there are many areas where this method fails, such as sequentially modeling and predicting continuous variables. In the case of predicting continuous variables, we have to look at other methods in optimization.

In this lecture scribe, we will begin to cover methods in Online Optimization. In particular, we will be covering the case of Online Optimization for Linear Loss functions.

8.2 Introduction to Online Optimization

Let's start off by defining a toy problem for us to analyze.

8.2.1 Defining our problem

Our goal is to make money off of the stock market by buying/selling stocks each day. Each morning when the market opens, we go over the list of stocks we're trading, and choose to either buy or sell some amount of each one, represented as vector $\vec{w_t} \in W$. For example $\vec{w_t}(i) = 3$ means we bought 3 shares of stock i on the morning of day t. If $\vec{w_t}(j)$ is negative, then we shorted stock j on the morning of day t. At the end of each day, we sell whatever stocks we bought, and buy back whatever we shorted. We also have a limit on the magnitude of stocks one could purchase or short in any given day. This is done to keep the problem realistic. We don't have an infinite buying power - the amount we can buy/short must be bounded.

We make/lose money depending on how much the value of the stock changed over the course of the day. If on day t we purchased k of stock i ($\vec{w_t}(i) = k$) and it increased in value by x dollars, then when we sell at the end of the day we make a profit of kx dollars on that stock. Similarly, if we shorted k shares of stock i ($\vec{w_t}(i) = -k$) and it increased in value by x dollars, then we'd lose money (profit $= -kx$). To that end, we define another vector

Z_t, representing the change in value of each stock. We observe that total profit is equal to the sum of element-wise product of the vectors $\vec{w_t}$ and Z_t.

Since loss is the opposite of profit, we define loss to be $l_t(w_t) = -\langle Z_t, w_t\rangle$. Observe that the loss term is a linear function of w_t.

8.2.2 Analyzing our problem

For a problem like this, Naive FTL would choose to do the action that is 'best in hindsight'. By this formulation, FTL would always pick to buy stocks that increased more than they decreased in the past, and would pick to short the opposite. Again, we set an arbitrary bound to the number of stocks trade-able each day. More formally, this allows us to fully utilize the size of Z. Now, a value of .1 will not be treated the same as .2 for Z. Notice for this problem, $\vec{w} = \vec{0}$ is a safe option, analagous to choosing half 1's and half 0's in the traditional FTL problem.

Similar to the '010101...' case in discrete sequence prediction, our naive FTL scheme can be led into a trap by an adversary. To mitigate the risk, we add on a regularization term like we did in FTRL: $R(w) = \frac{1}{2\eta}\|w\|^2$, giving us the new scheme:

$$w_{t+1} = \underset{w}{\operatorname{argmin}}[(\sum_{i=1}^{t} -\langle Z_i, w\rangle) + \frac{1}{2\eta}\|w\|^2],$$

where our weight at day $t+1$ is what would have minimized our regularized total loss over all previous days.

We observe that all terms of this seem to be convex and differentiable, so we can find the optimal minimizing w through differentiation with respect to w:

$$\frac{d}{dw}[(\sum_{i=1}^{t} -\langle Z_i, w\rangle) + \frac{1}{2\eta}\|w\|^2] = 0 \implies -\sum_{i=1}^{t} Z_i + \frac{1}{\eta}w_{opt} = 0$$

This gets us the following optimal

$$w_{t+1} = w_{opt} = \eta\sum_{i=1}^{t} Z_i$$

This can be rewritten:

$$w_{t+1} = \eta Z_t + \eta\sum_{i=1}^{t-1} Z_i$$
$$w_{t+1} = w_t + \eta Z_t$$

This looks like a single step of some sort of gradient descent. We observe that in this single step, we can control how much influence Z_t has on w_{t+1} by changing η. This regularizer allows us to control how 'stubborn' and resistant our weights w_t are to change. Choosing a small value for η will effectively limit how much our weights can change by in a single day.

Now that we have a way of modifying and optimizing our weights, how do we go about bounding regret?

8.3 Analysis of Continuous FTL

Ideas:

1. We treat this like a FTPL problem. We can let our regularizer from before, $R(w) = \frac{1}{2\eta}\|w\|^2$ act as a perturbation on initial loss l_0.
2. Then, similar to in FTPL analysis, we break regret up into 2 terms:
   (a) How our true loss behaves compared to perturbed loss (similar to estimation error). How do we, across time, behave relative to best in hindsight regularized loss?
   (b) How far perturbed loss in hindsight is from true loss in hindsight (similar to approximation error).

8.3.1 Big Ideas

We start by modifying our loss with a perturbation, as proposed in idea (1), loss defined as follows:

$$\tilde{l}_t(w) := \begin{cases} R(w) : t = 0 \\ l_t(w) : t > 0 \end{cases}$$

Let's also define u, the single 'best in hindsight' choice for weights:

$$u = \underset{u'}{\operatorname{argmin}} \sum_{t=1}^{T} \tilde{l}_t(u')$$

Now we want to try and analyze regret on our perturbed loss:

$$\text{Perturbed Regret} = \text{Our Perturbed Loss - Best in Hindsight Perturbed Loss}$$
$$\tilde{R}_T = \sum_{t=0}^{T} \tilde{l}_t(w_t) - \sum_{t=0}^{T} \tilde{l}_t(u)$$

In FTPL, we conducted our regret analysis using the concept of 'Leader Change'. That's an easy value to work with in discrete settings, but will require more work to figure out for our continuous setting here.

We observe that on day t, w_t is the trading decision we made. Since w_{t+1} incorporates information about the behaviour of Z_t, it essentially 'corrects' w_t. So, looking back, w_{t+1} is the decision we wish we could have made on day t. In terms of losses, we see that $\tilde{l}_t(w_t)$ is the loss that we actually incurred on day t. However $\tilde{l}_t(w_{t+1})$, the loss we would have incurred if we used w_{t+1} instead of w_t on day t, is like a 'cheating loss' as we incorporate information that hasn't really been revealed yet. From here, we can start to bound our Regret.

8.3.2 Leader Change Analogy

We're looking to find something analogous to 'Leader Changes' here. In FTPL, we measured the number of times we observed $\mathbb{1}[\hat{x}_{t+1} \neq \hat{x}_t]$. In a continuous setting, we instead find the comparison of w_{t+1} and w_t on day t to be the continuous analogy to our discrete version from FTPL. Our goal here, like the discrete case, is to bound our total perturbed regret, and we hypothesize that it can be bounded by the difference between our perturbed true loss and our perturbed cheating loss.
Like in FTPL, we claim that:

$$\tilde{R}_T = \sum_{t=0}^{T} \tilde{l}_t(w_t) - \sum_{t=0}^{T} \tilde{l}_t(u) \leq \sum_{t=0}^{T} \tilde{l}_t(w_t) - \sum_{t=0}^{T} \tilde{l}_t(w_{t+1})$$

Which is equivalent to saying

$$\sum_{t=0}^{T} \tilde{l}_t(w_{t+1}) \leq \sum_{t=0}^{T} \tilde{l}_t(u)$$

where $\sum_{t=0}^{T} \tilde{l}_t(w_{t+1})$ is our 'cheating loss' total.

How do we go about proving this claim? Let's try it out for some values.

Intuition: Let $T = 0$, then $\tilde{l}_0(w_1) \leq \tilde{l}_0(u)$. Since w_1 is the term that minimizes $\tilde{l}_0(w) = R(w)$ and $R(w)$ has a minimum of 0, we have $\tilde{l}_0(w_1) = 0$. Therefore $\tilde{l}_0(w_1)$ must be less than or equal to $\tilde{l}_0(u) = R(u)$.

Let $T = 1$, then $\tilde{l}_0(w_1) + \tilde{l}_1(w_2) \leq \tilde{l}_0(u) + \tilde{l}_1(u)$. An important observation to make is that w_2 was chosen by the following scheme:

$$w_2 = \underset{w \in W}{argmin}[\tilde{l}_0(w) + \tilde{l}_1(w)].$$

So to actually get that term in the inequality we're faced with, we can add $\tilde{l}_0(w_2)$:

$$\tilde{l}_0(w_1) + \tilde{l}_0(w_2) + \tilde{l}_1(w_2) \leq \tilde{l}_0(u) + \tilde{l}_1(u) + \tilde{l}_0(w_2)$$

By definition,

$$\tilde{l}_0(w_2) + \tilde{l}_1(w_2) \leq \tilde{l}_0(u) + \tilde{l}_1(u)$$

Subtracting that away, we're left with:

$$\tilde{l}_0(w_1) < \tilde{l}_0(w_2)$$

We know this is true by the definition of w_1. Therefore, the inequality for case $T = 1$ has also been shown to be true.
From these base cases, we begin a proof by induction.

Proof: Assume this to be true for $T = k$.
Let's now try to prove it for $T = k + 1$

$$\sum_{t=0}^{k+1} \tilde{l}_t(w_{t+1}) \leq \sum_{t=0}^{k+1} \tilde{l}_t(u)$$
$$\implies \sum_{t=0}^{k+1} \tilde{l}_t(w_{t+1}) + \sum_{t=0}^{k} \tilde{l}_t(w_{k+2}) \leq \sum_{t=0}^{k+1} \tilde{l}_t(u) + \sum_{t=0}^{k} \tilde{l}_t(w_{k+2})$$
$$\implies \sum_{t=0}^{k} \tilde{l}_t(w_{t+1}) + \sum_{t=0}^{k+1} \tilde{l}_t(w_{k+2}) \leq \sum_{t=0}^{k+1} \tilde{l}_t(u) + \sum_{t=0}^{k} \tilde{l}_t(w_{k+2})$$

We know

$$\sum_{t=0}^{k+1} \tilde{l}_t(w_{k+2}) \leq \sum_{t=0}^{k+1} \tilde{l}_t(u)$$

because w_{k+2} is defined to be the minimizer of total losses from 0 to $k + 1$.

We also have

$$\sum_{t=0}^{k} \tilde{l}_t(w_{t+1}) \leq \sum_{t=0}^{k} \tilde{l}_t(w_{k+2})$$

since w_{k+1} is defined to be the minimizer of total losses from 0 to k.
Therefore,

$$\sum_{t=0}^{k} \tilde{l}_t(w_{t+1}) + \sum_{t=0}^{k+1} \tilde{l}_t(w_{k+2}) \leq \sum_{t=0}^{k+1} \tilde{l}_t(u) + \sum_{t=0}^{k} \tilde{l}_t(w_{k+2}).$$

Which was what we wanted.
Thus, we have proved our initial claim.

■

Now that we have this inequality, we can derive our regret bound.

8.4 Bounding Regret

From before, we showed that:

$$\tilde{R}_T \leq \sum_{t=0}^{T} \tilde{l}_t(w_t) - \sum_{t=0}^{T} \tilde{l}_t(w_{t+1}) = \sum_{t=0}^{T} \left[\tilde{l}_t(w_t) - \tilde{l}_t(w_{t+1}) \right]$$

We recall that at time $t = 0$, our perturbed loss term is just exactly equal to our perturbation $R(\cdot)$. As per our assumptions, this perturbation is exactly the approximation error that we need to 'pay to play the game'.
For all other nonzero time steps, our Regret is bounded by the difference between true and cheating loss. Similar to the 'Leader Change' formulation, the sum of these differences is just our estimation error.

8.4.1 Bounding estimation error

$R_t^{(A)}$, our estimation error, is bounded as follows

$$R_t^{(A)} \leq \sum_{t=1}^{T} \left[\tilde{l}_t(w_t) - \tilde{l}_t(w_{t+1}) \right]$$

Looking at the inner terms of our summation, we observe that we can simplify them.

$$\tilde{l}_t(w_t) - \tilde{l}_t(w_{t+1}) = -Z_t(w_t - w_{t+1}) = -Z_t(-\eta Z_t) = \eta Z_t^2$$

Therefore,

$$\sum_{t=1}^{T} \left[\tilde{l}_t(w_t) - \tilde{l}_t(w_{t+1}) \right] = \eta \sum_{t=1}^{T} Z_t^2$$

Therefore we have the following bound on the term resembling estimation error:

$$R_t^{(A)} \leq \eta \sum_{t=1}^{T} Z_t^2$$

8.4.2 Bounding approximation error

Now let's examine our approximation error. As previously stated, our approximation error is the same as the 'price we pay' at timestep $t = 0$ to 'play the game'. That's the same as the difference between modified loss in hindsight and true loss in hindsight.

$$R_t^{(B)} = R(u) = \frac{1}{2\eta}||u||^2$$

Therefore,

$$R_t = R_t^{(A)} + R_t^{(B)} \leq \frac{1}{2\eta}||u||^2 + \eta \sum_{t=0}^{T} Z_t^2$$

However, we still can't truly bound regret, since we have extra terms: u, Z_t. By formalizing our intuition of how 'markets' work, we can define bounds for these terms.

Assumptions: Let's define G and B such that:

- For all t, $||Z_t|| \leq G$
- Define B such that $w \in \{u, ||u|| \leq B\}$

Then, we have:

$$R_T \leq \frac{B^2}{2\eta} + \eta T G^2$$

Let's determine the value for η that minimizes our regret bound. Since our R_t is bounded by a convex equation, our minimum will occur if:

$$\frac{d}{d\eta}\left(\frac{B^2}{2\eta} + \eta T G^2\right) = 0$$

Then,

$$TG^2 - \frac{B^2}{2\eta^2} = 0$$

Which gets us the optimal value:

$$\eta = \frac{B}{G\sqrt{2T}}$$

Thus,

$$R_T \leq BG\sqrt{2T} \in O(\sqrt{T})$$

Observe that as B increases, the magnitude of 'bets' we're allowed to make increases. Intuitively, if we're capable of making large bets, we'll need to have a larger step size to utilize that capacity in reasonable time - we're willing to make large changes if we know we can afford it.

Similarly, if G is large, this tells us that the market is capable of swinging by large amounts every day and is more volatile. Intuitively, we'd want to be more conservative in such a volatile environment to not be led into a trap by an adversary. Consequently, we observe that for large values of G, our step size is small - we're stubborn to change in the face of volatility.

Thus, our total regret is sublinear.

Analysis

Do these assumptions mean anything physically? Are they needed?

Well, let's take a look at each assumption individually. Our first assumption tells us that Z_t is not too large because

$$||Z_t|| \leq G.$$

Then, we know that our Signal-to-Noise-Ratio (SNR) is bounded. It's good to limit our signal strengths as if we allow very strong signals, we open up the possibility that an adversary can send very strong signals. This can counteract the 'stubbornness' we want in our model and thus, we need a reasonable and bounded SNR to defend against adversaries. However, if we do increase G, then we can account for adversarial attacks by reducing our learning rate η.

Our second assumption limits the number of stocks we can trade per day:

$$w \in \{u, ||u|| \leq B\}$$

We can't just trade an infinite stocks for a 1 cent profit each. This is not realistic and it also breaks our model as we can accrue infinite profit. Thus, we bound the number of stocks we can trade. If B increases, however, then we can trade more and can approach that limit faster by increasing our learning rate η. Next time, we'll look over convexity and linear approximations for nonlinear losses.

8.4.3 Example

We perform a simple simulation of our stock-trading problem here, in a random but non adversarial setting. Over ten thousand days, we sample values for Z_t from several random variables, and compute w_t using our FTRL scheme.
We then plot the total and average regret, respectively, for this simulation run over several random variables of different parameters.
In all cases, we observe that total regret appears to behave sublinearly, and that average regret does seem to tend to zero, as expected.