Interactive Online Learning

Sanjiban Choudhury
Announcements (all on Ed!)

1. Assignment 0 (survey released)

2. Lecture 1 slides + notes up on website

3. Office hours available:

   Sanjiban (Tue/Thurs 11-12pm, Gates 413B)
   Dhruv (Mon/Wed 11-12 pm, Rhodes 400)
Learning Today!
How humans learn ...
Can’t we collect a LOT of data and train robots offline?
**SUPERVISED LEARNING**

1. Get Data
   - Input (s)
   - Output (a)

2. Train Policy
   \[ \pi : s \rightarrow a \]

3. Deploy!
Train ≠ Test
Activity!
Think-Pair-Share!

Think (30 sec): What are different sources of train-test mismatch?

Pair: Find a partner

Share (45 sec): Partners exchange ideas
Case 1: Data changes over time
Case 2: Data changes with robot behavior
Case 3: Data changes adversarially (game)
Challenge:
Don’t know the test distribution upfront

Learner  Collect Data
\[\text{\textbullet}\]
Interactive Learning

Learner  Adversary
Interactive Learning

Learner

Initialize policy

Update policy

Adversary

Chooses loss

\[ \pi_1 \text{ [policy]} \]

\[ l_1(\cdot) \text{ [loss]} \]

\[ \pi_2 \]

\[ l_2(\cdot) \]
Prediction with Expert Advice
Loss = 1.0
Let’s formalize!
Regret = \sum_{t=1}^{T} l_t(\pi_t) - \min_{\pi^*} \sum_{t=1}^{t} l_t(\pi^*)

(Learner) - (Best in hindsight)
How do we design algorithms that are no-regret?

Regret = \sum_{t=1}^{T} l_t(\pi_t) - \min_{\pi^*} \sum_{t=1}^{T} l_t(\pi^*)
At every round $t$, choose the best expert in hindsight

$$\pi_t = \arg \min_{\pi} \sum_{i=1}^{t-1} l_i(\pi)$$

(lowest total loss)
\[ \sum l_t \]

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>( l_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expert 2</th>
<th>( l_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expert 3</th>
<th>( l_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

Avg. Regret: - -
\[
\sum l_t \quad l_1 \quad l_2
\]

<table>
<thead>
<tr>
<th></th>
<th>(l_1)</th>
<th>(l_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Expert 3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Avg. Regret: \(0.80\)
$$\sum l_t$$

<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Expert 3</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Avg. Regret: **0.40**
\[
\sum_{t} l_t \hspace{1cm} l_1 \hspace{1cm} l_2 \hspace{1cm} l_3 \hspace{1cm} l_4 \\
\begin{array}{cccc}
\text{Expert 1} & 1.0 & 0.5 & 0.5 & 1.0 \\
2.0 & \\
\text{Expert 2} & 0.2 & 0.5 & 1.0 & 0.2 \\
1.7 & \\
\text{Expert 3} & 0.5 & 0.2 & 0.2 & 0.5 \\
0.9 & \\
\end{array}
\]

Avg. Regret: \textbf{0.53}
\[ \sum l_t \]

<table>
<thead>
<tr>
<th></th>
<th>( l_1 )</th>
<th>( l_2 )</th>
<th>( l_3 )</th>
<th>( l_4 )</th>
<th>( l_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expert 1</strong></td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>( 3.0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Expert 2</strong></td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.2</td>
<td>1.0</td>
</tr>
<tr>
<td>( 1.9 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Expert 3</strong></td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>( 1.4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Avg. Regret:** 0.40
<table>
<thead>
<tr>
<th></th>
<th>$l_t$</th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
<th>$l_5$</th>
<th>$l_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>3.5</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Expert 2</td>
<td>2.9</td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.2</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Expert 3</td>
<td>1.6</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

**Avg. Regret:** 0.32
<table>
<thead>
<tr>
<th>$\sum l_t$</th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
<th>$l_5$</th>
<th>$l_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>4.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.2</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Avg. Regret: 0.26
FTL appears to be no regret …
Let’s prove it!
Can you make FTL have high regret?
$$\sum l_t$$

Expert 1

Expert 2

Avg. Regret: --
\[ \sum l_t \quad \begin{array}{c} \text{Expert 1} \\ - - \\ \text{Expert 2} \\ - - \end{array} \quad l_1 \quad \begin{array}{c} 1.0 \\ 0.0 \end{array} \]

Avg. Regret:
\[ \sum l_t \]

<table>
<thead>
<tr>
<th></th>
<th>( l_1 )</th>
<th>( l_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Avg. Regret: 1.00
\[
\sum l_t
\]

<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Avg. Regret: 0.50
\[ \sum l_t \]

<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Avg. Regret: **0.67**
\[ \sum l_t \]

<table>
<thead>
<tr>
<th></th>
<th>( l_1 )</th>
<th>( l_2 )</th>
<th>( l_3 )</th>
<th>( l_4 )</th>
<th>( l_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expert 1</strong></td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Expert 2</strong></td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Avg. Regret: 0.50
<table>
<thead>
<tr>
<th>$l_t$</th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
<th>$l_5$</th>
<th>$l_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Avg. Regret: **0.60**
### Predictions not stable $\rightarrow$ High regret!

<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
<th>$l_5$</th>
<th>$l_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Summing up the predictions for each expert:

- **Expert 1:** \(\sum l_t = 3.0\)
- **Expert 2:** \(\sum l_t = 3.0\)

**Avg. Regret:** 0.50
“A powerful enough adversary can drive the Regret of any deterministic online algorithm to $O(T)$ by anticipating its prediction and setting maximal loss”

How can we curb the power of the adversary?
\[ \pi_t = \arg \min_{\pi} \sum_{i=1}^{t-1} l_i(\pi) \]

FOLLOW THE LEADER!

Adversary breaks *any* determinism
The virtue of hedging
Choose probability over experts

\[ p = \frac{w_i}{\sum_{i} w_i} \]

- Expert 1: \( w^1 = 1.0 \)
- Expert 2: \( w^2 = 2.0 \)
- Expert 3: \( w^3 = 1.0 \)
Let’s formalize!
Let’s apply FTL again (but on the space of weights)

At every round $t$, choose the best weights in hindsight

$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w)$$
Choose $\pi^1$ if $w = 0.0$ and Choose $\pi^2$ if $w = 1.0$.

Loss $= 0.75$ Avg. Regret $= 0.5$
Loss = 1.0  Avg. Regret = 0.5

Choose $\pi^1$

Choose $\pi^2$
Loss \( w = 0.0 \)

Loss \( w = 1.0 \)

Choose \( \pi^1 \)

Choose \( \pi^2 \)

Loss = 1.0  Avg. Regret = 0.5
Choose \( \pi^1 \) when \( w = 0.0 \) and choose \( \pi^2 \) when \( w = 1.0 \).
Loss = 1.0  Avg. Regret = 0.5

Choose $\pi^1$ when $w = 0.0$

Choose $\pi^2$ when $w = 1.0$
Follow the leader is too aggressive …

Both in discrete and continuous settings!

Stability is the key problem!
\[ w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w) \]

Unstable predictions!
Be stable

Slowly change predictions
Follow the Regularized Leader

\[ w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w) + \eta_t R(w) \]

What are some choices for regularization?
GENERALIZED WEIGHTED MAJORITY

Episode IV

A NEW HOPE
1. At $t=1$, set weight for expert $i$ as $w_i^1 = 1$

2. At time $t$, choose expert $i$ with probability

$$\frac{w_t^i}{\sum_i w_t^i}$$

3. Update weight for expert $i$ (Bump down if loss is high)

$$w_{t+1}^i = w_t^i \exp(-\eta l_t(\pi^i))$$
<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>1.0</td>
<td>0.5</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>$w_2$</td>
<td>0.2</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>$w_3$</td>
<td>0.5</td>
<td>1.0</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>$w_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Choose $\pi^1$ when $w = 0.0$ and choose $\pi^2$ when $w = 1.0$.
Loss $w = 0.0$
$w = 1.0$

Choose $\pi^1$

$w = 1/\sqrt{e}$

Choose $\pi^2$

Loss = 0.6  Avg. Regret = 0.17
Loss \( w = 0.0 \) \( w = 1.0 \)

Choose \( \pi^1 \) \( \pi^2 \)

Loss = 0.78 Avg. Regret = 0.21

\( w = \frac{1}{\sqrt{e^3}} \)
Loss $w = 0.0 \quad w = 1.0$

Choose $\pi^1$

Choose $\pi^2$

Loss $= 0.6$  Avg. Regret $= 0.18$

$w = 1/\sqrt{e}$
Loss = 0.78  Avg. Regret = 0.2

\[ w = 0.0 \]

\[ w = 1.0 \]

Choose \( \pi^1 \)

Choose \( \pi^2 \)

\[ w = \frac{1}{\sqrt{e^3}} \]
Linear Programming

Games

Boosting

Soft-RL
Three Challenges

C1: Derive GWM from Follow the Regularized Leader

C2: Show that GWM is No-Regret

C3: Show that FTRL is No-Regret

(Share on Ed!)
Unstable predictions! 

\[
w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w) + \eta_t R(w)
\]

Regularization \Rightarrow \text{No Regret!}

\[
w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w)
\]