Shifting Experts on Easy Data

Manfred K. Warmuth      Wouter M. Koolen
Eternal Dilemma

Worst Case

IID
Huge Difference (Expert Setting Example)

$O(\sqrt{T})$

$O(1)$
Holy Grail

Adaptive
### Status Quo

<table>
<thead>
<tr>
<th>Experts</th>
<th>Hedge</th>
<th>FTL/ERM</th>
<th>FlipFlop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandits</td>
<td>EXP3</td>
<td>UCB</td>
<td>SAO</td>
</tr>
<tr>
<td>Shifting</td>
<td>Fixed Share</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Freund’s</td>
<td>Mixing Past Posterior</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Problem</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Shifting

Best model (expert) changes over time.

- Optimal algorithm for segment-wise IID data?
  Should pay $O(\ln \#\text{experts})$ per switch
- How to combine it with worst-case robustness?
Freund’s Problem: Long-term Memory

- Optimal algorithm for segment-wise IID data?
  - Should pay \( O(\ln \#\text{good experts}) \) per switch
- How to combine it with worst-case robustness?
Candidate Algorithms

For IID shifting:

- FL on the best partition
- FL on a shifting window
- FL on capped loss differences
- FL on exponentially decaying losses
- ... 

For IID long-term memory?
The Big Question

Single algorithm for shifting

▶ worst-case robust
▶ adaptive to IID data