DATA 37200: Learning, Decisions, and Limits (Winter 2026)

Lecture 15: Playing games with the exponentially weighted forecaster

Instructor: Frederic Koehler

Reference

Chapters 2 and 7 of Cesa-Bianchi and Lugosi textbook.

Intro

- Last time we introduced the idea of a zero-sum game between two players.
  - What one player wins is what the other player loses.
  - Can be formulated as a matrix game (single round), although multiple rounds (Markov game) is sometimes more natural.
  - Today we focus on the matrix game formalism.
- The most important fact about zero-sum games is the "minimax" theorem of Von Neumann.
- By revisiting and slightly generalizing the ideas for forecasting with expert advice, we can give a short proof of the minimax theorem.
- Illustrates connections between: concentration inequalities, online prediction, optimization, game theory.

Prediction with Expert Advice: Old vs New

- Previously: our goal is to predict a response $y_t \in \mathbb{R}$ in squared loss $(y_t - \hat{y}_t)^2$.
- Previously: each expert gave us a single number $f_{i,t} \in \mathbb{R}$ as their advice.
- NEW: our goal is to output a prediction vector in a convex set $\mathcal{D}$ to minimize a loss $\ell(\hat{y}_t, y_t)$ which is convex in $\hat{y}_t$.
- NEW: each expert gives us advice $f_{i,t} \in \mathcal{D}$ at round t.
- This generalized version of prediction with expert advice is a special case of "online convex optimization". Key assumptions are convexity of loss and decision space.

Prediction with Expert Advice/OCO

Setup: A forecaster predicts a sequence $y_1, y_2, \ldots \in \mathcal{Y}$ using advice from m experts. The advice of each expert is a vector in a convex set $\mathcal{D}$, and there is a fixed loss function ℓ

$$\ell : \mathcal{D} \times \mathcal{Y} \to \mathbb{R}$$

which is convex in its first argument.

Protocol at each time $t = 1, \ldots, T$:

1. Environment reveals expert advice $f_{1,t}, \ldots, f_{m,t} \in \mathcal{D}$.
2. Forecaster chooses prediction $\hat{p}_t \in \mathcal{D}$.
3. Environment reveals true outcome y_t.
4. Forecaster incurs loss $\ell(\hat{p}_t, y_t)$; experts incur loss $\ell(f_{i,t}, y_t)$.

Goal: Minimize the **regret** relative to the *best* expert in hindsight:

$$R_{i,n} = \hat{L}_n - L_{i,n} = \sum_{t=1}^{T} \ell(\hat{p}_t, y_t) - \sum_{t=1}^{T} \ell(f_{i,t}, y_t)$$

The Exponentially Weighted Average Forecaster

Intuition: Assign weights to experts based on their past performance. Shrink weights of experts who make mistakes[1]

Weights: Let $w_{i,t-1}$ be the weight of expert i at time t. We define it exponentially based on cumulative loss $L_{i,t-1}$:

$$w_{i,t-1} = e^{-\eta L_{i,t-1}}$$

(where $\eta > 0$ is a tuning parameter).

Prediction: The forecaster predicts the weighted average of the experts' advice:

$$\hat{p}_t = \frac{\sum_{i=1}^{m} w_{i,t-1} f_{i,t}}{\sum_{j=1}^{m} w_{j,t-1}}$$

[1] For squared loss, we gave a simple Bayesian motivation for this. Bayesian interpretation seems less helpful for today.

General EWA Guarantee

Theorem 2.2 (CBL book): Assume the loss function ℓ is convex in its first argument and takes values in $[0, 1]$. For any n and $\eta > 0$, and for all outcome sequences $y_1, \ldots, y_n \in \mathcal{Y}$, the regret of the EWA forecaster satisfies:

$$\hat{L}_T - \min_{i=1,\ldots,m} L_{i,T} \leq \frac{\log m}{\eta} + \frac{T\eta}{8}$$

Optimal Tuning: If we know the horizon T in advance, we can choose $\eta = \sqrt{8 \log m / T}$.

Resulting Bound: This yields a worst-case regret bound of $\sqrt{(T/2) \log m}$. (Not as good as we got for squared loss.)

Proof of Theorem 2.2 - Part 1

Proof Setup & Hoeffding's Lemma

- To prove the theorem, we analyze the sum of the weights: $W_t = \sum_{i=1}^{N} e^{-\eta L_{i,t}}$.
- The proof tracks the evolution of $\log(W_t / W_{t-1})$.

Tool - Hoeffding's Inequality: Let X be a random variable with $a \leq X \leq b$. Then for any $s \in \mathbb{R}$:

$$\log \mathbb{E}[e^{sX}] \leq s\mathbb{E}[X] + \frac{s^2(b-a)^2}{8}$$

We will apply this to a random variable Z that takes value $\ell(f_{i,t}, y_t)$ with probability $\frac{w_{i,t-1}}{W_{t-1}}$.

Proof of Theorem 2.2 - Part 2

Bounding the Weight Ratio
Calculate the ratio:

$$\log \frac{W_t}{W_{t-1}} = \log \frac{\sum_{i=1}^{m} w_{i,t-1} e^{-\eta \ell(f_{i,t}, y_t)}}{\sum_{j=1}^{m} w_{j,t-1}}$$

Apply Lemma 2.2 (with $X = \ell(f_{i,t}, y_t) \in [0, 1]$, $s = -\eta$):

$$\log \frac{W_t}{W_{t-1}} \le -\eta \frac{\sum_{i=1}^{m} w_{i,t-1} \ell(f_{i,t}, y_t)}{\sum_{j=1}^{m} w_{j,t-1}} + \frac{\eta^2}{8}$$

Convexity Step: Because ℓ is convex in its first argument:

$$\frac{\sum w_{i,t-1} \ell(f_{i,t}, y_t)}{\sum w_{j,t-1}} \ge \ell\left(\frac{\sum w_{i,t-1} f_{i,t}}{\sum w_{j,t-1}}, y_t\right) = \ell(\hat{p}_t, y_t)$$

Result: $\log \frac{W_t}{W_{t-1}} \le -\eta \ell(\hat{p}_t, y_t) + \frac{\eta^2}{8}$.

Proof of Theorem 2.2 - Part 3

Telescoping and the Final Bound

Sum the inequality over $t = 1, \ldots, T$:

$$\log \frac{W_T}{W_0} \leq -\eta \hat{L}_T + \frac{\eta^2}{8} T$$

We also lower bound $\log \frac{W_n}{W_0}$ using the best expert ($W_0 = m$):

$$\begin{aligned} \log \frac{W_n}{W_0} &= \log \left(\sum_{i=1}^{m} e^{-\eta L_{i,n}} \right) - \log m \\ &\geq \log \left(\max_i e^{-\eta L_{i,n}} \right) - \log m \\ &= -\eta \min_i L_{i,n} - \log m \end{aligned}$$

Combine bounds and solve for $\hat{L}_n$:

$$\hat{L}_n \leq \min_{i=1,\ldots,m} L_{i,n} + \frac{\log N}{\eta} + \frac{\eta}{8} T$$

Next

Now that we have proved a bound for general prediction with expert advice, we return to game theory.

In game theory it is more customary to consider a “payoff/reward matrix” and the row player is trying to maximize the reward. But for today, we use the negated view that there is a loss matrix and the row player is trying to minimize their loss.

Repeated Zero-Sum Games

- **Setup:** Row player (forecaster) has N actions, Column player (environment) has M actions.
- **Matrix:** A loss matrix $\ell(i,j) \in [0,1]$.
- **Mixed Strategies:** Row player chooses a probability distribution $\mathbf{p}$ over N actions. Column player chooses distribution $\mathbf{q}$ over M actions.
- **Loss:** $\ell(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{N}\sum_{j=1}^{M} p_i q_j \ell(i,j)$.
  - Meaning: under mixed strategies $\mathbf{p},\mathbf{q}$ the expected loss of row player.
- **Connection to Prediction:** The row player can treat each of their N pure actions as an "expert".
- **Goal:** The row player plays the EWA strategy against the sequence of plays $y_1, y_2, \ldots$ chosen by the column player.

Applying EWA to Games

Bounding the Game Loss using Theorem 2.2

- Let the row player use EWA to generate mixed strategy $\mathbf{p}_t$ at time t.
- Let $\mathbf{q}_t$ be the column player's mixed strategy at time t.
- By Theorem 2.2, the sequence of mixed strategies $\mathbf{p}_t$ guarantees that for *any* sequence of opponent strategies $\mathbf{q}_t$:

$$\frac{1}{T}\sum_{t=1}^{T}\ell(\mathbf{p}_t,\mathbf{q}_t) \leq \min_{i=1,\ldots,N}\frac{1}{T}\sum_{t=1}^{T}\ell(i,\mathbf{q}_t) + \frac{\log N}{T\eta} + \frac{\eta}{8}$$

Note: Because the expected loss ℓ is linear in both arguments, the convex loss requirement of Theorem 2.2 is perfectly satisfied.

Proving the Minimax Theorem (Part 1)

Von Neumann's Minimax Theorem states:

$$\min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q}) = \max_{\mathbf{q}} \min_{\mathbf{p}} \ell(\mathbf{p}, \mathbf{q})$$

We know that $\min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q}) \geq \max_{\mathbf{q}} \min_{\mathbf{p}} \ell(\mathbf{p}, \mathbf{q})$ is always true.
We use the regret bound of EWA to prove the reverse inequality:

$$\min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q}) \leq \max_{\mathbf{q}} \min_{\mathbf{p}} \ell(\mathbf{p}, \mathbf{q})$$

Adversarial Environment: Let the column player play the *best response* to $\mathbf{p}_t$ at each step:

$$\mathbf{q}_t = \arg\max_{\mathbf{q}} \ell(\mathbf{p}_t, \mathbf{q})$$

Proving the Minimax Theorem (Part 2)

Define the average plays: $\bar{\mathbf{p}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{p}_t$ and $\bar{\mathbf{q}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{q}_t$.
Chain of inequalities:

$$
\begin{aligned}
\min_{\mathbf{p}}\max_{\mathbf{q}} \ell(\mathbf{p},\mathbf{q}) &\le \max_{\mathbf{q}} \ell(\bar{\mathbf{p}}_T,\mathbf{q}) \\
&= \max_{\mathbf{q}} \frac{1}{n}\sum_{t=1}^{T} \ell(\mathbf{p}_t,\mathbf{q}) \quad (\text{expand } \bar{\mathbf{p}}_n) \\
&\le \frac{1}{n}\sum_{t=1}^{T} \ell(\mathbf{p}_t,\mathbf{q}_t) \quad (\text{since } \mathbf{q}_t \text{ is best response}) \\
&\le \min_{i} \ell(i,\bar{\mathbf{q}}_T) + \frac{\log N}{n\eta} + \frac{\eta}{8} \quad (\text{EWA bound}) \\
&= \min_{\mathbf{p}} \ell(\mathbf{p},\bar{\mathbf{q}}_T) + \frac{\log N}{n\eta} + \frac{\eta}{8} \\
&\le \max_{\mathbf{q}}\min_{\mathbf{p}} \ell(\mathbf{p},\mathbf{q}) + \frac{\log N}{n\eta} + \frac{\eta}{8}
\end{aligned}
$$

Conclusion of the Proof

Convergence to the Minimax Value
We have shown that for any n and $\eta > 0$:

$$\min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q}) \leq \max_{\mathbf{q}} \min_{\mathbf{p}} \ell(\mathbf{p}, \mathbf{q}) + \frac{\log N}{n\eta} + \frac{\eta}{8}$$

- By setting $\eta = \sqrt{8 \log N / T}$, the error term becomes $\sqrt{\frac{\log N}{2T}}$.
- As $n \to \infty$, the error term goes to 0.
- Therefore: $\min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q}) \leq \max_{\mathbf{q}} \min_{\mathbf{p}} \ell(\mathbf{p}, \mathbf{q})$.

Conclusion: The EWA forecaster's guarantee that it will eventually do as well as the best pure strategy in hindsight directly proves that the Minimax value of the game exists.

Discussion

- Why did this work?
- EWA tries to compete with best pure strategy. However, we know from rock-paper-scissors that no pure strategy works against best-response adversary.
- So in (e.g.) rock-paper-scissors, if EWA converged to a single "best expert" it would fail badly.
- We used that EWA can compete with single best expert no matter how the adversary plays. If we force the adversary to fix their moves in advance, then pure strategies are optimal.

Approximate Nash Equilibria

Let

$$V = \min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q})$$

be the value of the game. In the proof we showed that

$$\max_{\mathbf{q}} \ell(\bar{\mathbf{p}}_T, \mathbf{q}) \leq V + o(1)$$

and

$$\min_{\mathbf{p}} \ell(\mathbf{p}, \bar{\mathbf{q}}_T) \geq V - o(1).$$

The pair $(\bar{\mathbf{p}}_T, \bar{\mathbf{q}}_T)$ is an approximate Nash equilibrium. This means that if the row player fixes mixed strategy $\bar{\mathbf{p}}_T$, the column player has a negligible incentive to deviate from $\bar{\mathbf{q}}_T$, and vice versa. In particular, we know

$$V - o(1) \leq \ell(\bar{\mathbf{p}}_T, \bar{\mathbf{q}}_T) \leq V + o(1).$$

Bayesian interpretation of Row-EWA (?)

- Pretend that there exists a "true but unknown" pure strategy $i^* \in [N]$. This is a strategy for the row player, but we model it as only known to the column player.
- Row player starts with a uniform prior over i^*.
- Pretend/model the column player's response is sampled[2]

$$\Pr(q \mid i^* = i) \propto \exp(-\eta \sum_j q_j \ell_{ij}).$$

- Provided $\eta = o(1)$ as $T \to \infty$, our posterior converges to minimax/Nash equilibrium strategy. Crucially, parameters are chosen so we may fail to determine/learn i^* completely if pure strategies are suboptimal.

[2] Interpretation (?): weight column player mixed strategies towards those where i^* would have been a good response by row player.