6.046 Lecture 10 3/16

1

Admin:

→ PSet 5 is out today
→ We will have a post-quiz survey

Today:

→ Game theory
→ Two person zero-sum games
→ Min-Max theorem
→ How to get rich? (Stock market prediction)
→ (Rand.) Weighted majority alg.

Let's play a game!

Prisoner's Dillemma

Setting: → Two members of criminal gang are arrested
→ They are put in separate cells
⇒ no means to communicate!
→ Police does not have evidence to convict both of them on the principal charge
→ But: they are able to sentence them on a lesser charge for a year
→ They offer each one of them:
"If you testify against the other, you walk free, and him will go to jail for three years"

→ Catch: If both testify, they both go to prison, but "only" for two years

Compact matrix representation of possible outcomes:

Prisoner A: \ Prisoner B:	silent	testify
silent	1 year / 1 year	0 years / 3 years
testify	3 years / 0 years	2 years / 2 years

What will the prisoners do?

→ Intuitively: The "optimal" outcome for them is $\boxed{\text{silent/silent}}$
⇒ they get 'only' 1 year of jail each

→ But is that what will really happen?

Note: → If both prisoners are silent, then if any one of them switches to testify, he/she decreases his/her jail time from two to zero years!

→ Also, if exactly one testifies then the other one has incentive to testify as well (to go from three to two years)

⇒ The only "stable" outcome: (i.e., no one has incentive to deviate)

$\boxed{\text{Both prisoners testify}}$

Remark: → This behavior is somewhat hard to replicate in real world as this analysis ignores factors like "repeated" nature of the game
(e.g., if I "betray" you now, you will take revenge later)

Games (from game theory perspective):

Thought experiments to help us learn how to predict rational behavior in situations of conflict

Players aim to maximize their own utility (= gain) (Altruism, masochism, etc. are NOT modelled)

Every player's actions affect the other players' utility (not necessarily in negative way)

Convenient way to represent two-player game:

Utility matrices: A = utility matrix of Player A
B = —— " —— B

Rows of A (B) = actions of Player A (= Row player)
Columns of A (B) = —— " —— B (= Column player)

Given action i of Player A (row) & action j of Player B (column):

$$\boxed{\begin{aligned} A_{ij} &= \text{utility of Player A on outcome } (i,j) \\ B_{ij} &= \text{— " — B on outcome } (i,j) \end{aligned}}$$

Important class of games: (Two-person) zero-sum games

Two-person games in which:

$$\boxed{A_{ij} = -B_{ij} \quad \forall_{i,j}}$$

Note: Fully described by matrix A (or B) alone

⇒ Models direct conflict

(= "your loss is my gain")

⇒ Prisoner's Dilemma is NOT zero-sum

Example: Rock-Paper-Scissors game

Matrix A:
$$\begin{array}{c|ccc} & R & P & S \\ \hline R & 0 & -1 & 1 \\ P & 1 & 0 & -1 \\ S & -1 & 1 & 0 \end{array}$$

(Matrix B: (= −A)
$$\begin{array}{c|ccc} & R & P & S \\ \hline R & 0 & 1 & -1 \\ P & -1 & 0 & 1 \\ S & 1 & -1 & 0 \end{array}$$
)

Is there a "stable" outcome?

* $\boxed{\text{No}}$, if consider only pairs of single action (= deterministic strategy)

  If row player chooses rock → column player has incentive to switch to paper
  → But, then row player has incentive to play scissors
  → But, then column player has incentive to switch to rock...
  "infinite loop of switching"

* $\boxed{\text{Yes}}$ if we allow randomized strategies!

  ⇒ If both players randomize uniformly at random over rock, paper & scissors, no one can improve their expected utility by switching to different strategy

  ⇒ The pair of strategies $\boxed{\left(\left(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3}\right),\left(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3}\right)\right)}$ is a Nash equilibrium of R-P-S game

---

Nash Equilibrium: [Nash '50] (= "stable" outcome)
A state of game (= choice of (possibly randomized) strategies for each player) such that no player can improve his/her expected utility by unilaterally changing his/her strategy

"Fun" fact: "Beautiful Mind" movie got this concept WRONG!

Note: (testify, testify) = pure Nash equilibrium of prisoner's dilemma game

↑ deterministic

Fundamental question: Does Nash equilibrium always exist?

$P = \{x \mid x \geq 0 \ \& \ \sum_i x_i = 1\}$

$Q = \{y \mid y \geq 0 \ \& \ \sum_j y_j = 1\}$

Min-Max Theorem: [von Neumann '28]

For any matrix A, if

$$\boxed{V_R := \max_{x \in P} \min_{y \in Q} xAy} \quad \& \quad \boxed{V_C := \min_{y \in Q} \max_{x \in P} xAy}$$

then

$$\boxed{V_R = V_C = V}$$

Game theory interpretation:

→ V_R = expected utility of row player if she/he needs to go first (i.e., disclose his/her strategy before the other player discloses his/her)

→ V_C = expected negative utility of column player if he/she goes first

⇒ Row player wants to maximize xAy
Column player wants to minimize xAy

⇒ $V_R \leq V_C$ as V_R corresponds to row player playing with a handicap (column player can switch seeing x) & row player has an advantage in case of V_C

⇒ So, $\boxed{V_R = V_C}$ is surprising!

Key implication:

$$x^* := \underset{x}{\operatorname{argmax}} \min_y xAy$$

$$y^* := \underset{y}{\operatorname{argmin}} \max_x xAy$$

$\Rightarrow$ (x^*, y^*) is always a Nash equilibrium of the two-person zero-sum game described by A

(If there was an incentive to deviate from x^* or y^* it would contradict that $V_R = V_C$)

$\Rightarrow$ Nash equilibrium always exists for two-person zero-sum games

How about general games?

[Nash '51]: Any game with a finite number of players and finite number of possible actions per player has a Nash equilibrium

---

Proof of Min-Max Theorem:

Key idea: Let's express V_R (& V_C) as an LP!

Recall: $V_R = \max_x \min_y xAy$

$\Rightarrow$ if $z = V_R$ and x is the strategy corresponding to V_R, we have to have that, for any column action j, the expected utility we get if column player plays j has to be $\geq z$

(otherwise col. player could switch to playing j and thus contradict that $z = V_R$)

$\Rightarrow$ In other words: $\boxed{\sum_i A_{ij} x_i \geq z} \quad \forall j$

→ Also, as x is a randomization over actions
$$\boxed{\sum_i x_i = 1 \quad \& \quad x_i \geq 0 \ \forall i}$$

As, z is the expected utility we are guaranteed to get in this case, we can write:

$$\boxed{\begin{aligned} V_R = \max\ & z \\ & \left(\sum_{i} A_{ij} x_i\right) - z \geq 0 \quad \forall j \\ & \sum_i x_i = 1 \\ & x \geq 0 \end{aligned}} \qquad (R)$$

By similar reasoning (taking into account col. player wants to min. xAy)

$$\boxed{\begin{aligned} V_C = \min\ & u \\ & \left(\sum_{j} A_{ij} y_j\right) - u \leq 0 \quad \forall i \\ & \sum_j y_j = 1 \\ & y \geq 0 \end{aligned}} \qquad (C)$$

Key observation: (R) & (C) are dual to each other!

(See next page)

⇒ By strong LP duality:

$$\boxed{V_R \underset{\text{Def of (R)}}{=} OPT(R) = OPT(C) \underset{\text{Def of (C)}}{=} V_C}$$

⇒ Strong LP duality implies Min-Max Thm.

(In fact, one can show these two are equivalent!)

∎

"Missing" calculations:

8

→ (R) in standard form:

$$\boxed{\begin{array}{lll} \max & z^+ - z^- & \\ & -\sum_i A_{ij} x_i + z^+ - z^- \leq 0 \quad \forall j & (y_j) \\ & \sum_i x_i \leq 1 & (u^+) \\ & -\sum_i x_i \leq -1 & (u^-) \\ & x, z^+, z^- \geq 0 & \end{array}}$$

→ Dual of (R)

$$\boxed{\begin{array}{ll} \min & u^+ - u^- \\ & -\sum_j A_{ij} y_j + u^+ - u^- \geq 0 \quad \forall i \\ & \left.\begin{array}{l} \sum_j y_j \geq 1 \\ -\sum_j y_j \geq -1 \end{array}\right\} \sum_j y_j = 1 \\ & y, u^+, u^- \geq 0 \end{array}}$$

→ Taking $\boxed{u = u^+ - u^-}$ and multiplying $\geq$ by (-1) gives us (C) !

Σ

How to get rich (if you have good advice)

(Very simple) stock market model:

→ X_t = stock market index on day t (e.g., value of Dow Jones at day t)

→ Initially, $\boxed{X_0 = 0}$

→ Each day t:

→ You have to predict if $\boxed{X_t = X_{t-1} + 1 \text{ (UP)} \text{ or } X_t = X_{t-1} - 1 \text{ (DOWN)}}$

→ Only then, X_t is revealed

⇒ If your prediction was correct, gain \$1M; otherwise, lose \$1M

How well can you do if you don't know how X_t evolves?

→ Best we can do: flip a coin

⇒ Expected gain is 0

⇒ **Additional element:** n "experts"

"experts" as in "they have an opinion", but not necessarily know what they are talking about

⇒ Think: Financial advisors in CNN, WSJ, etc.

Each day t: Before you have to make a prediction, you get UP/DOWN advice from each expert

Note: This advice does not need to be consistent. Some might say UP, some might claim DOWN.

Problem: This advice might be wrong or even misleading!

Goal: Do well if at least one expert is consistently providing decent advice

Measure of quality:

$$\boxed{\text{Regret} = (\#\text{ of our mispredictions}) - (\#\text{ of mistakes of the best expert})}$$

→ Think: How much better off would we be, if we followed advice of the best expert from the start

Key difficulty: We know who the best expert is only in hindsight! (and then it is too late)

How to approach this problem?

① **Easier case:** The best expert makes NO mistakes

→ **Halving algorithm:**

- → Maintain a pool S of "trustworthy" experts
- → At first, S = all experts
- → At day t:
  - → Go with majority prediction of only experts in S
  - → After seeing X_t, remove from S, all experts that mispredicted

(Pruning of S)

Exercise: How good is this alg. if we never prune S?

Claim: Halving algorithms has regret $\leq \log_2 n$
(= makes $\leq \log_2 n$ mistakes) (does not depend on length of the sequence)

Can show: This is best possible bound here

Proof: Every time the alg. makes a mistake, we reduce the size of S by at least half (in pruning step)
$\Rightarrow$ it takes $\boxed{\leq \log_2 n}$ halvings to reduce size of S to 1 (which would be the (unique) perfect expert) $\square$

---

② **General setting:** Even the best expert makes some $\boxed{m^*}$ mistakes

→ **Iterated halving alg.:**
- → Run halving alg.
- → If S becomes empty after pruning
  - $\Rightarrow$ Replenish S by putting all experts back

Remember: Everyone might make a mistake now

→ **Can show:** Iterated halving alg. has regret

$$\boxed{\leq (m^* + 1) \log_2 n}$$

→ This is within $\approx \log_2 n$ of optimum

Can we do better?

Obs 1: Replenishing the set S loses a lot of valuable info. No difference between very bad expert (tons of mistakes) and pretty good (few mistakes) but not perfect expert

Obs 2: Classifying experts as either "trustworthy" ($\in S$) or not "trustworthy" ($\notin S$) is too coarse

Idea: Use weights to capture "trustworthness" of experts in a more continious manner

$\Rightarrow$ Update this estimate of our trust by reducing the weight of an expert upon each mistake

$\Rightarrow$ Aggregate predictions by taking a weighted majority of answers

Weighted majority (WM) alg.:

- $\rightarrow$ Maintain a weight w_i for each expert i
- $\rightarrow$ Initially, $\boxed{w_i = 1}$, for all i
- $\rightarrow$ At day t:
  - $\rightarrow$ Predict by following weighted majority
  - $\rightarrow$ After X_t is revealed, $\boxed{w_i \leftarrow \frac{1}{2} w_i}$, for each expert i that was wrong

---

Claim: WM alg. has the regret of

$$\boxed{\lesssim 1.4 \cdot m^* + 2.4 \log_2 n}$$

Proof:

Let w_i^t = weight of expert i at the end of day t

Consider a potential $\boxed{W^t := \sum_i w_i^t} \Rightarrow \boxed{W^0 = n}$

Each time the alg. makes a mistake

$\Rightarrow \geq \frac{1}{2}$ of total weight of experts gets their w^i halved

i.e. $\boxed{W^t \leq \frac{3}{4} W^{t-1}}$ in such rounds

$\Rightarrow$ If T is the total # of days, and m is the # of mistakes of the WM alg., then

$$(*) \quad \boxed{W^T \leq \left(\tfrac{3}{4}\right)^m W^0 = \left(\tfrac{3}{4}\right)^m n}$$

Note: Even if the alg. makes no mistake on day t, we still have $\boxed{W^t \leq W^{t-1}}$

On the other hand:

If i^* is the best expert and m^* is the # of his/her mistakes, then

$$(**) \quad \boxed{w_{i^*}^T = \left(\tfrac{1}{2}\right)^{m^*}}$$

Putting $(*)$ & $(**)$ together, we get

$$\boxed{\left(\tfrac{1}{2}\right)^{m^*} \leq w_{i^*}^T \leq W^T \leq \left(\tfrac{3}{4}\right)^m n}$$

$$\Rightarrow \quad m^* \log_2 \tfrac{1}{2} \leq m \log_2 \tfrac{3}{4} + \log_2 n$$

(Taking $\log_2(\ )$ of both sides)

$$\Rightarrow \quad m \leq 2.4\,(m^* + \log_2 n)$$

$\left(\text{Use } -\frac{1}{\log_2 \frac{3}{4}} \approx 2.4\right)$

$$\Rightarrow \quad \text{Regret} = m - m^* \leq 1.4\, m^* + 2.4 \log_2 n$$

$\boxed{\xi}$

So, WM alg. gets performance that approaches (when $m^* \to \infty$) only const. overhead compared to best expert!

In fact: If we make weight reduction be by a factor of $(1-\varepsilon)^{-1}$ instead of 2, for some $\boxed{0 < \varepsilon \leq \tfrac{1}{2}}$

$$\Rightarrow \quad \boxed{\text{Regret} \leq (1+\varepsilon)\, m^* + \tfrac{2}{\varepsilon} \log_2 n}$$

Can we do even better?

Yes, but have to use randomness

Idea: (Deterministic) weighted majority vote

$\Downarrow$

Randomized weighted majority sampling

(= when making prediction, sample a single expert i proportionally to his/her weight w_i & follow his/her prediction)

$\leftarrow$ Pr[choose prediction of expert i] $= \dfrac{w_i}{\sum_j w_j}$

→ This helps with "tie breaking" when the advice is split

Can show: Randomize weighted majority alg. has regret

$$\boxed{\leq \varepsilon \cdot m^* + \frac{\log_2 n}{\varepsilon}}$$

→ This is best possible

→ As $\boxed{m^* \to +\infty}$ & $\boxed{\varepsilon \to 0^+ \text{ (appropriately)}}$

Regret becomes $o(m^*)$!

$\Rightarrow$ Asymptotically, regret is vanishing