CS540 Introduction to Artificial Intelligence
Lecture 24

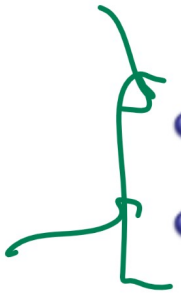
Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

August 10, 2021

Efficient Market Game

Q1

- The last two digits of your ID is your productivity (how much you can help a company produce). Choose between two companies to work for:
- A: you get paid how much you produce (your productivity).
- B: you get paid the average productivity of everyone working for this company.



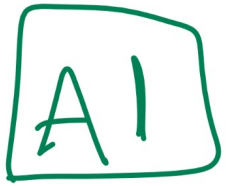
Announce Final update M8-11, P1-5, Q15-24 after midnight tonight.

P6 → Friday

Final

P1-6 → Sunday midnight → Monday submit

Remind Me to Start Recording



- Enter the last two digits of your ID.
- Add the last two digits of your ID to your Zoom name (for example, if your name is cat88 and the last two digits of your ID is 24, then change your name to cat88-24).

Mechanism Design Problem

- Players have hidden (private) information (type).
- Designer designs a game so that players with different types will choose different actions (thus reveal their type) in an equilibrium.

Adversarial Machine Learning

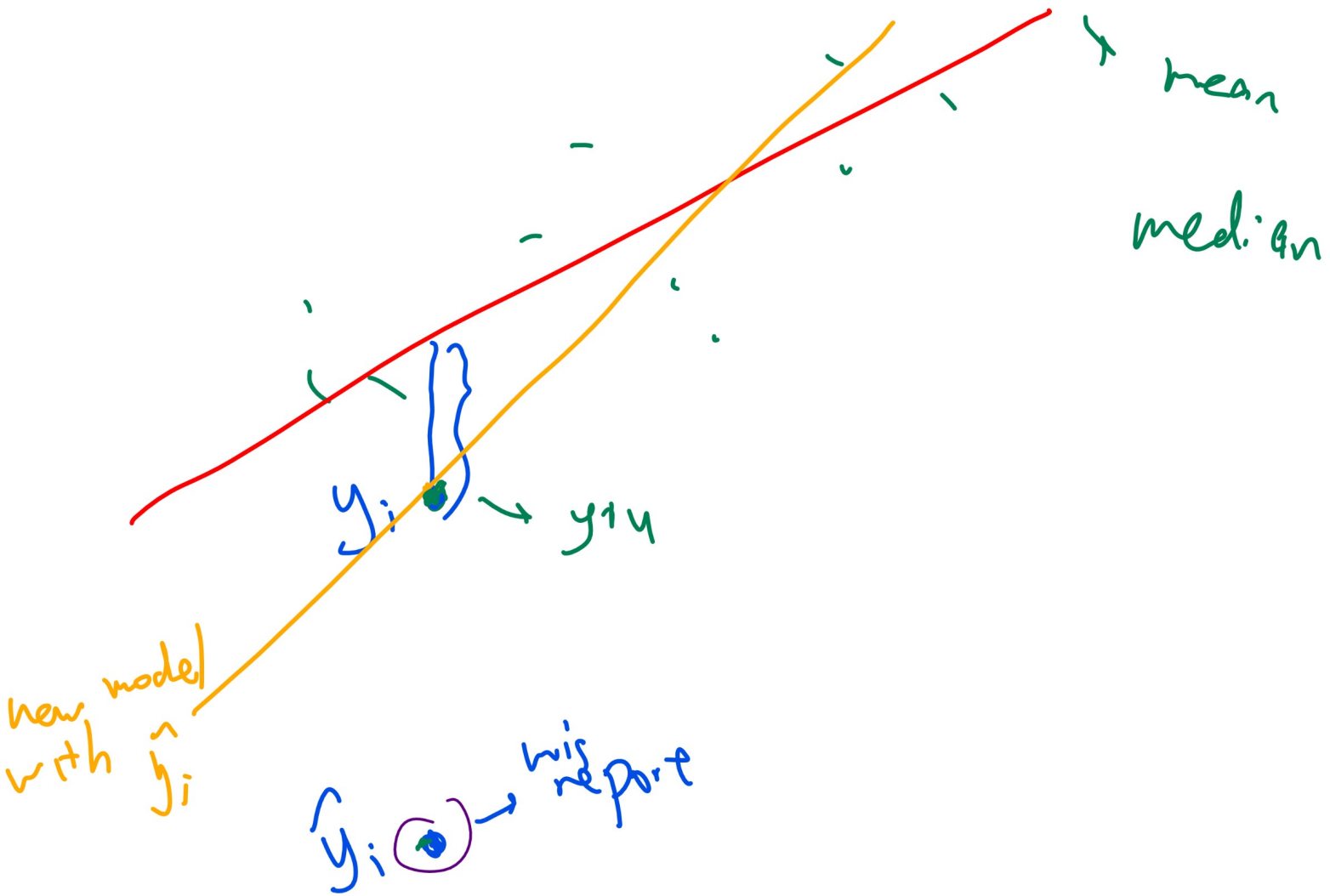
- Motivations:

1. Adversarial attack. → fake data. model
2. Machine teaching. → fake data. model
3. Ethics: equality and fairness. original data biased

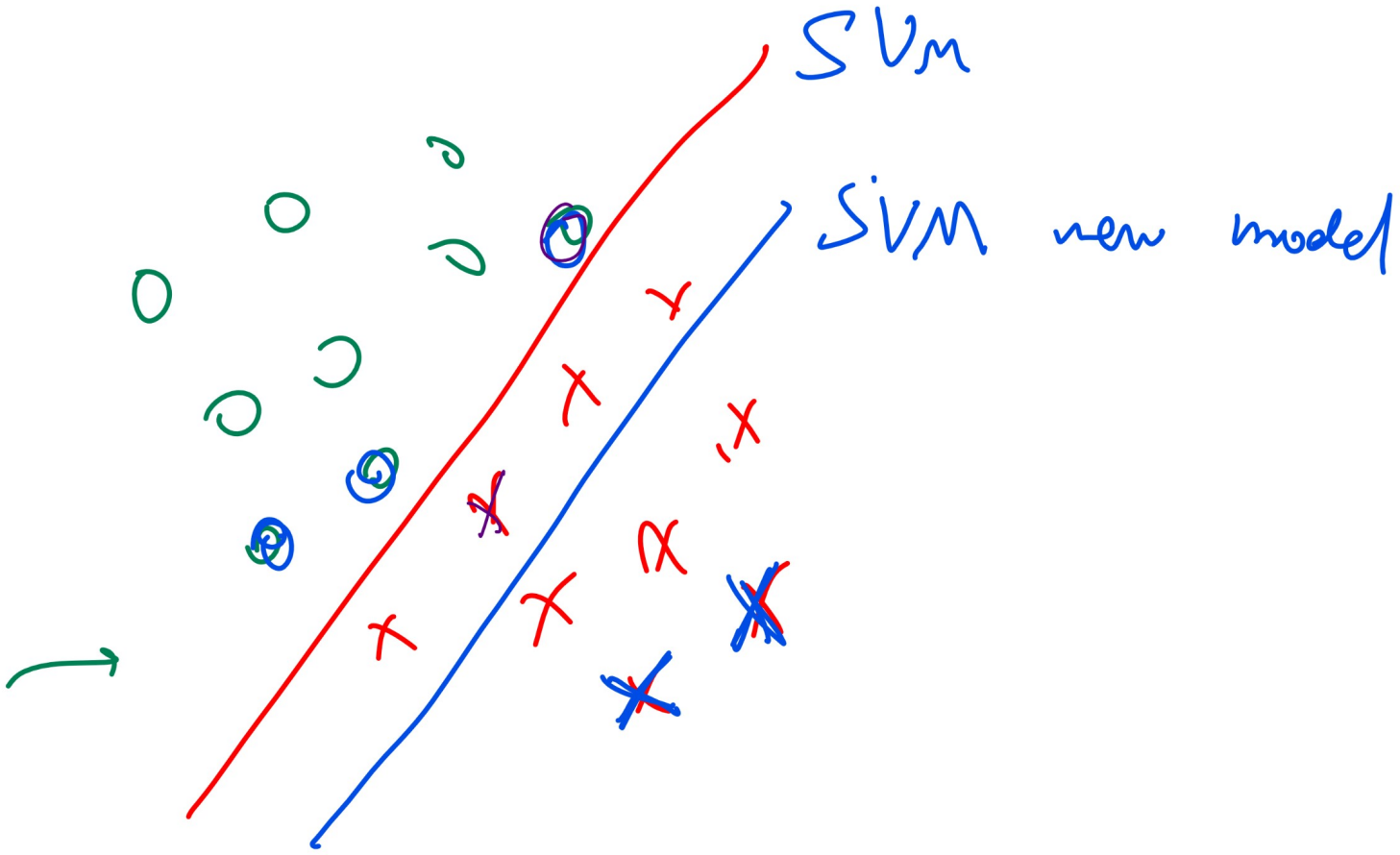
- Types of attack:

1. Test time. ✓ × y
2. Training time: misreport features or labels (misinformation).
3. Training time: select subset of data points (disinformation).

Misinformation Attack of Linear Regression



Disinformation Attack of Linear Classifiers



Attack Prevention

- Ways to prevent adversarial attacks on machine learning algorithms:

1. Regularization (train more general models)
2. Mechanism design (implement truthful report).
3. Competitive data provider.

VCG Mechanism

- Vickrey Clarke Groves Mechanism.
- Clarke Pivot Rule: players pay their externality.
- Example: Second Price Sealed Bid Auction.

First Price Sealed Bid Auction

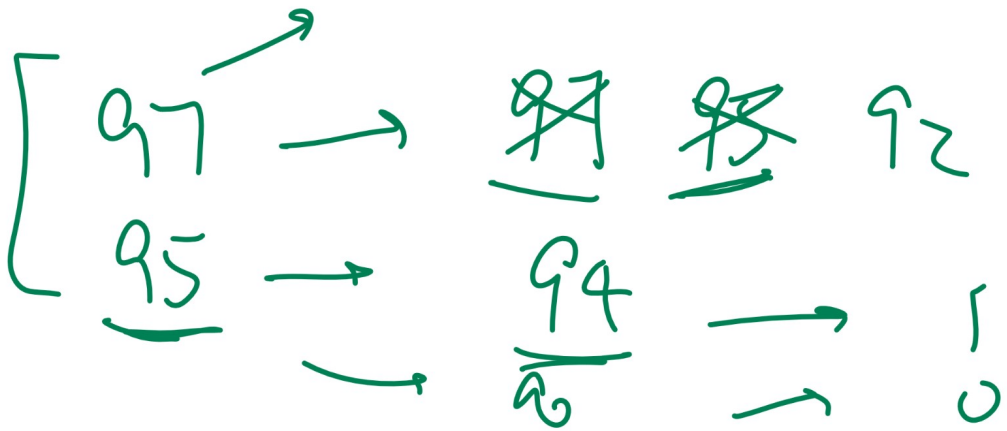
A2

- Enter a bid, the highest bidder gets the object and pay the bid.
- If the value of the object to you is v_i, and your bid is b_i, the (net) payoff is:

last two digit ID

17

1. $v_i - b_i$ if $b_i = \max_j b_j$.
2. 0 otherwise.



First Price Sealed Bid Auction Bid

- A: $b_i > v_i$
- B: $b_i = v_i$
- C: $b_i < v_i$
- D: $b_i = 0$

Second Price Sealed Bid Auction

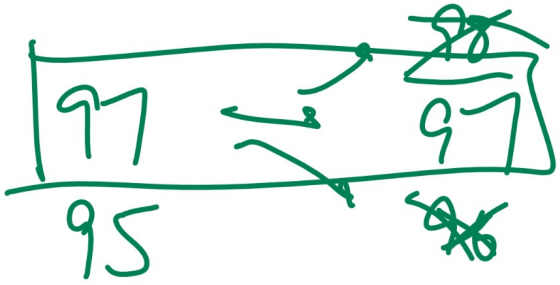
A3

$b_i = v_i$

- Enter a bid, the highest bidder gets the object and pay the second highest bid.
- If the value of the object to you is v_i, and your bid is b_i, the (net) payoff is:

1. $v_i - \max_{j \neq i} b_j$ if $b_i = \max_j b_j$.
2. 0 otherwise.

ID



$91 - 97 < 0$

Second Price Sealed Bid Auction Bid

- A: $b_i > v_i$
- B: $b_i = v_i$
- C: $b_i < v_i$
- D: $b_i = 0$

All Pay Auction

A4

- Enter a bid, the highest bidder gets the object, but all players pay their bids.
- If the value of the object to you is v_i, and your bid is b_i, the (net) payoff is:

1. $v_i - b_i$ if $b_i = \max_j b_j$.
2. - b_i otherwise.

All Pay Auction Bid

- A: $b_i > v_i$
- B: $b_i = v_i$
- C: $b_i < v_i$
- D: $b_i = 0$

Incentive Compatibility

- In second price auction, bidders do not have incentive to lie about their value.

Public Good Provision

- Suppose the object is a public good (for example a highway, everyone can enjoy for free).
- The public good is provided if the sum of the bids is higher than the cost of providing the public good.
- Everyone pays the cost of the public good minus the sum of the other bidder's bids.
- The bidders do not have incentive to lie about their values.

Insurance Example No Mechanism

- Suppose the probability that you have an accident is proportional to the last two digits of your ID.
- You plan to buy an insurance, the insurance company asks if you are a safe driver.

1. If you answer yes: you pay a low insurance premium (e.g. 50 dollars).
2. If you answer no: you pay a high insurance premium (e.g. 100 dollars).

- A: YES
- B: NO

Insurance Example Indirect Mechanism

- Suppose the probability that you have an accident is proportional to the last two digits of your ID.
- You plan to buy an insurance, the insurance company asks you to select one of two contracts.

1. Contract 1: you pay a low insurance premium (e.g. 50 dollars) with a high deductible (e.g. 250 dollars).
2. Contract 2: you pay a high insurance premium (e.g. 100 dollars) with a low deductible of (e.g. 50 dollars).

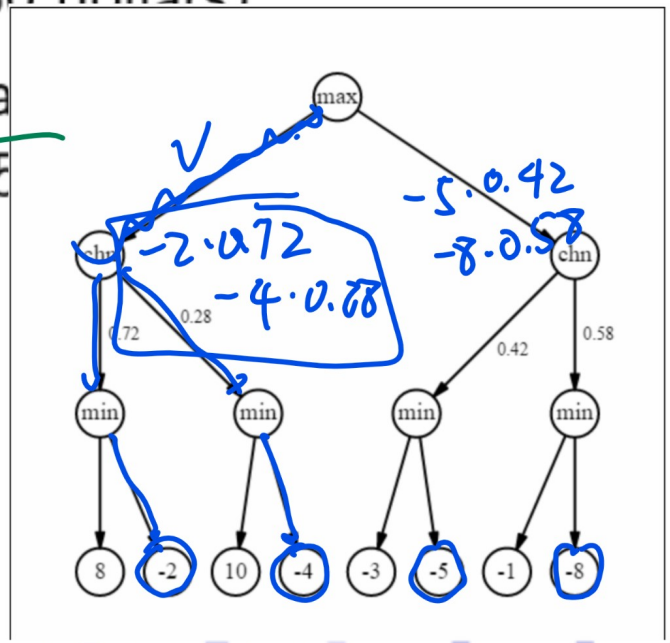
- A: Contract 1
- B: Contract 2

Insurance Example Direct Mechanism

- Suppose the probability that you have an accident is proportional to the last two digits of your ID.
- You plan to buy an insurance, the insurance company asks if you are a safe driver.

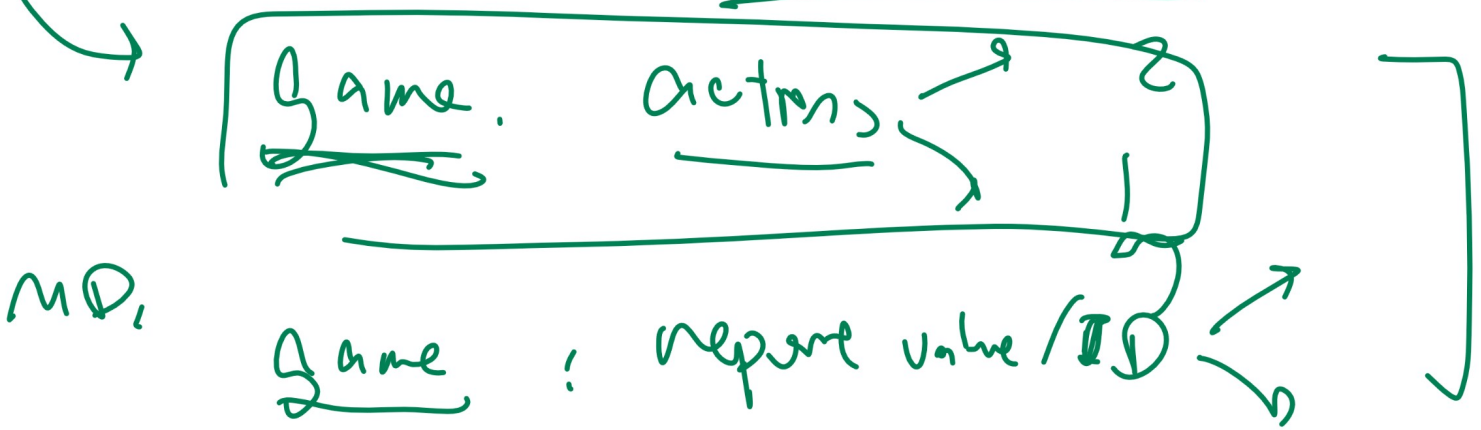
1. If you answer yes: you pay a low insurance premium (e.g. 50 dollars) with a high deductible (e.g. 250 dollars)
2. If you answer no: you pay a high insura… dollars) with a low deductible of (e.g. 5…

- A: YES
- B: NO



Revelation Principle

- Direct mechanism: ask the insurer to report their risk.
- Indirect mechanism: ask the insurer to select a contract.
- Revelation principle says, (under technical conditions), if there is an incentive compatible mechanism, there must be an incentive compatible direct mechanism.



Dating Model No Mechanism

- Suppose how much you enjoy spending time with the boy (girl) you like is proportional to the last two digits of your ID.
- The boy (girl) asks if you like him or her.

1. If you answer yes: he (she) will agree to be your boyfriend (girlfriend).
2. If you answer no: he (she) will say no.

Dating Model With Mechanism

- Suppose how much you enjoy spending time with the boy (girl) you like is proportional to the last two digits of your ID.
- The boy (girl) asks if you like him or her.

1. If you answer yes: you have to spend a lot of time with him (her), then he (she) will agree to be your boyfriend (girlfriend).
2. If you answer no: he (she) will say no.

Matching Problem

- Your preference over getting matched with another student is proportional to how close your ID is to that student's ID (for example, if your ID is 24, you prefer 23 to 21 and prefer 25 to 27). Also assume if you have an odd number ID, you can only be matched with someone with an even number ID.
- Enter the ID of the person you are matched with.
- If you are not matched with anyone, enter your own ID.

Deferred Acceptance Algorithm

- A stable matching is one in which no two people are will to break their own matches and get matched with each other.
- One efficient algorithm that guarantees a stable matching is the Deferred Acceptance Algorithm:

1. People with odd number IDs propose to their favorite choice (that hasn't rejected them before).
2. People with even number IDs accept the best proposal and reject the rest.
3. Repeat until stable.

Deferred Acceptance Algorithm Implications

- Also called Gale Shapley Algorithm for Stable Marriage Problem.
- The mechanism is incentive compatible for people with odd number IDs, but not people with even number IDs.
- The mechanism leads to the best possible match (among all stable matching) for people with odd number IDs, and the worst possible match for people with even number IDs.
- Rural Hospitals Theorem: unmatched people are unmatched in all stable matching.

Other Matching Applications

- Matching:

1. Assignment of graduating medical students to hospitals.
2. Assignment of servers to Internet users.