CS540 Introduction to Artificial Intelligence

Lecture 24

Young Wu

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

August 16, 2022

Efficient Market Game

- 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.

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:

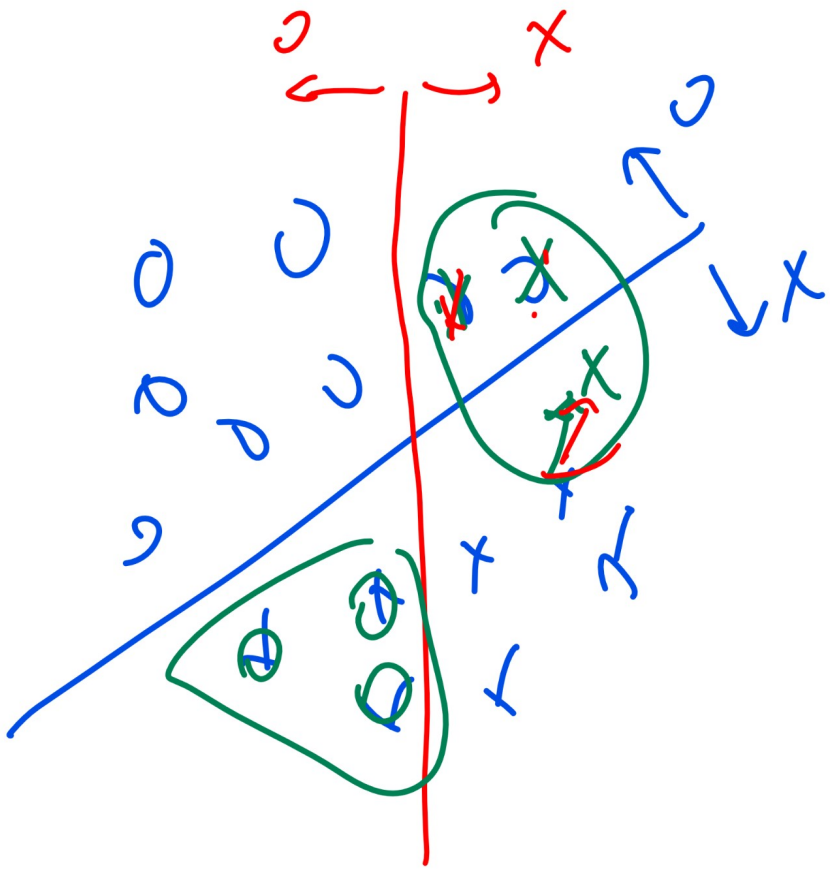
1. Adversarial attack.
2. Machine teaching.
3. Ethics: equality and fairness.

- Types of attack:

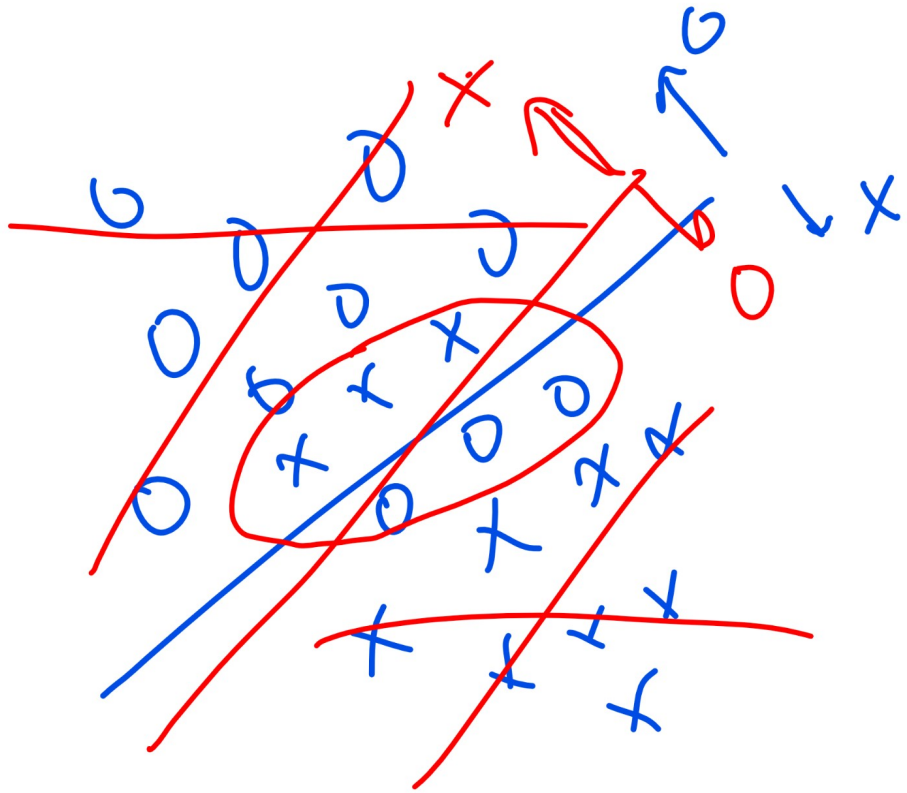
1. Test time.
2. Training time: misreport features or labels (misinformation).
3. Training time: select subset of data points (disinformation).

Test Time Attack Example

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

- 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:

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

- 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.

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

- 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.$250 dollars).
2. If you answer no: you pay a high insurance premium ($e.g.$100 dollars) with a low deductible of ($e.g.$50 dollars).

- 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.

X7 Q1

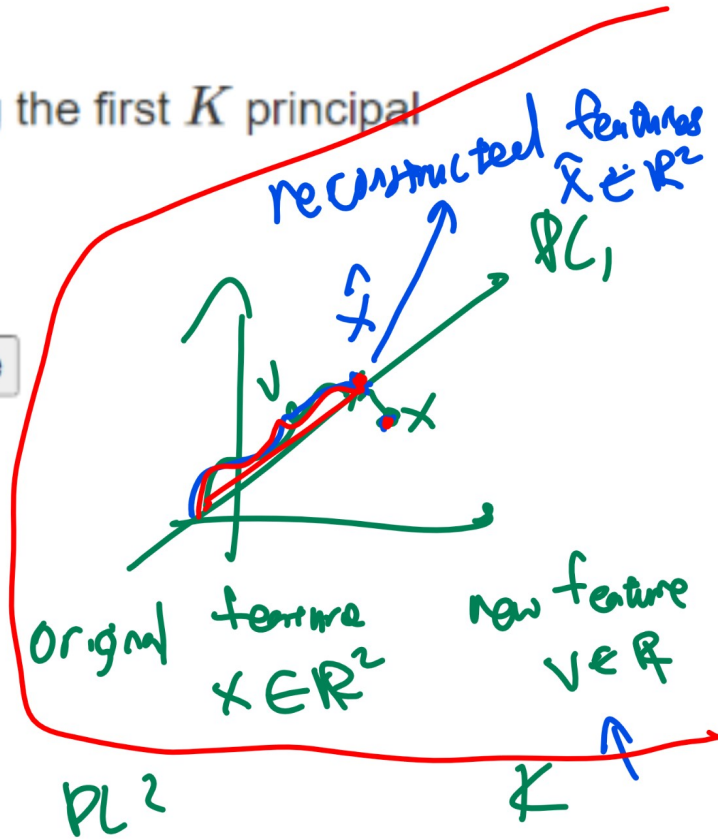
Question 1

- [3 points] Given the variance matrix $\hat{\Sigma}$ is a diagonal matrix, what is the smallest value of K so that the Manhattan distance between the vector

$$\begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}$$

with 38 1's and its reconstruction using the first K principal components is less than or equal to 8?

- Answer: 30 . Calculate



$\hat{\Sigma} = P \; D \; P^{-1}$

$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{pmatrix}$ ← one position

PC_1

$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 3 \\ 0 \\ 0 \end{pmatrix}$

PL^2

+ ...

$= \| \begin{pmatrix} \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow K \text{ 1's} - \begin{pmatrix} \vdots \\ \end{pmatrix} \|_1 = 38 - K$

$\hat{x}$ ← reconstructed matrix

x

↑ Manhattan distance between x, $\hat{x}$

when $38 - K \leq 8$

$K \geq 30$

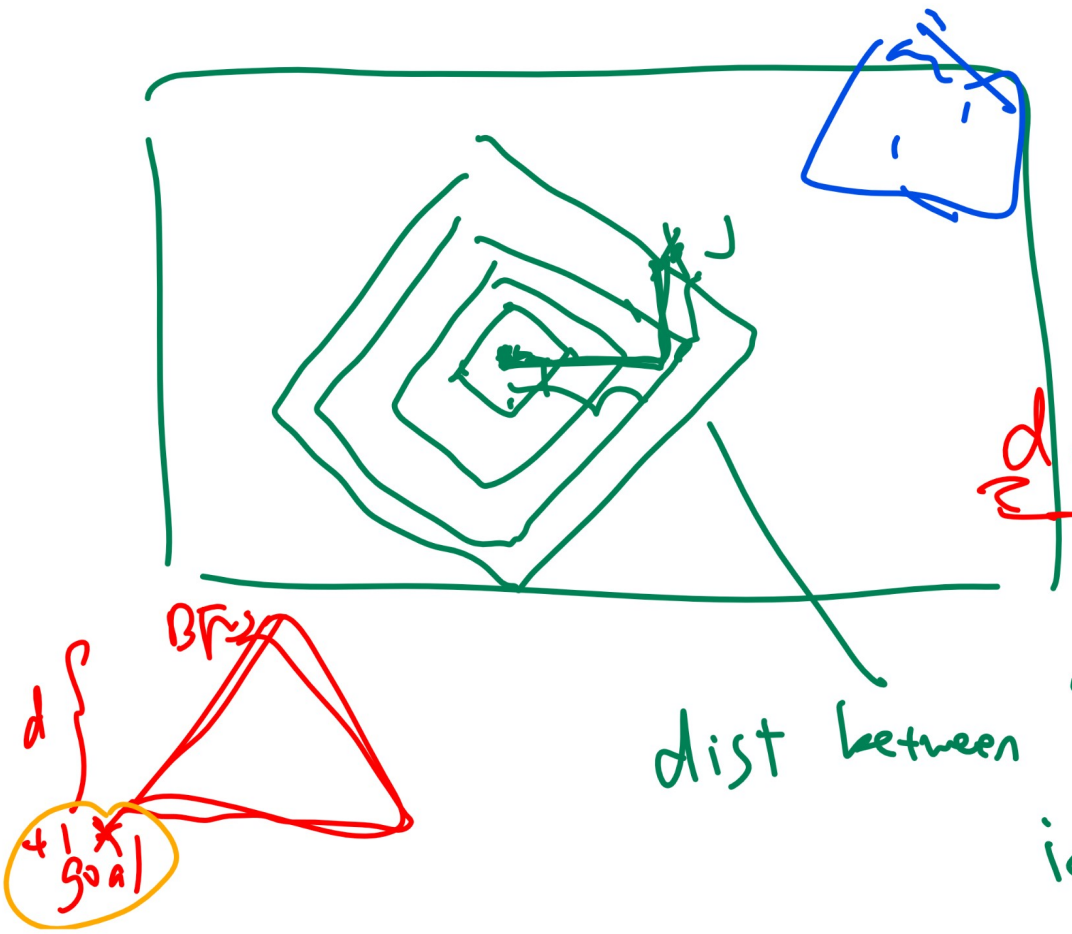
Question 2

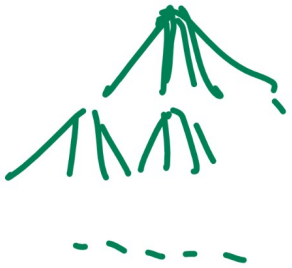
• [3 points] In a 200 by 140 grid, Tom is located at (98, 54) and Jerry is located at (112, 95). Tom uses BFS (Breadth First Search) to find Jerry and the successors of a state (one cell in the grid) are the four neighboring states on the grid (the cells above, below, to the left and to the right). What is the minimum number of states that need to be expanded to find (and expand) the goal state? The order in which the successors are added can be arbitrary. Include both the initial and the goal states.

• Answer: . Calculate

DFS

unique





states there are.

$d = |112 - 98| + |95 - 54|$ units

or legs from initial state.

dist between T and J

is Manhattan distance.

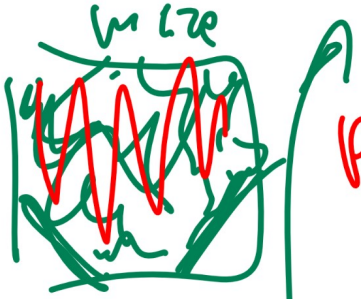
(1, 0) (2, 2)

goal

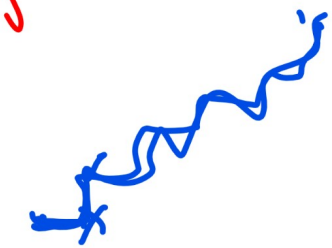
$d = 4$ = Manhattan dist

$d^2 + (d-1)^2 + 1$

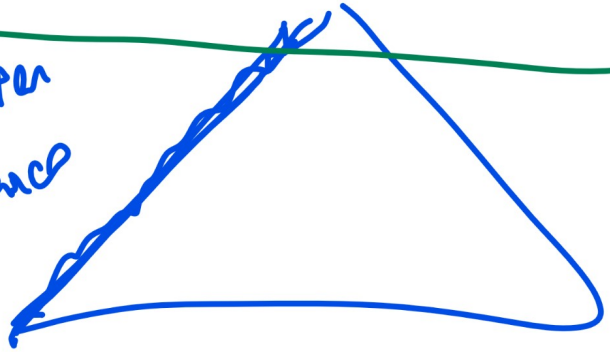
except $(d-1)^2 + (d-2)^2 + 1$



PJ



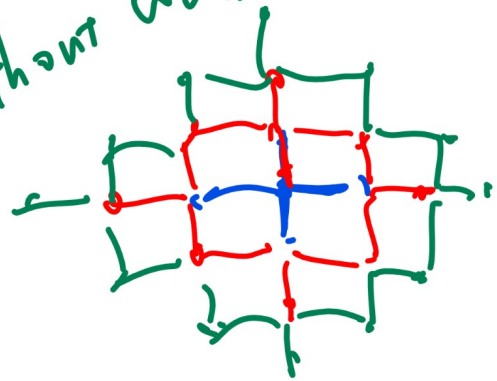
Manhattan distance

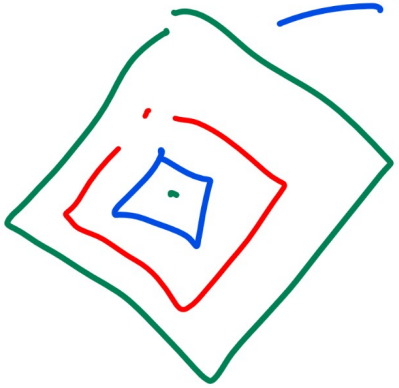


goal

$d+1$

PJ without walls.





Question 3

• [3 points] There are $n = 101$ students in CS540, for simplificity, assume student 0 gets grade $g = 0$, student 1 gets grade $g = 1$, ..., student $n - 1$ gets grade $g = n - 1$. The payoff for each student who drop the course is 0 , the payoff for the students who stay is $0.02\ g - 1.5$ if the student has the lowest grade among all students who decide to stay in the class, and the $0.02\ g - 1$ otherwise. If each student only uses actions that are rationalizable (i.e. survive the iterated elimination of strictly dominated actions), how many students will stay in the course? If there are multiple correct answers, enter one of them.

• Answer: 25 or ≥6 . Calculate

$0.02\ g - 1 < 0$

Stay Drop

$g < 50$

D is dominant strategy

$0.02\ g - 1.5 > 0$

S is dominant strategy

(stay is never best response)

$g > 75$

(drop is never best response)

if $g < 50 \longrightarrow D$

if $g > 75 \longrightarrow S$

rational

$50 \le g \le 75 \longrightarrow$ both D and S are rational.

↘ not rationalizable.

$\longrightarrow S = 50$

$0.02g - 1.5 < 0 \Rightarrow D$

$g = 51$

$0.02g - 1.5 < 0 \Rightarrow D$

$\vdots$

$g = 74$

$0.02g - 1.5 < 0 \Rightarrow D$

check $g = 75$

$0.02g - 1.5 = 0 \Rightarrow \begin{cases} D \\ S \end{cases}$ both rationalizable

$0.02g - 1.5 < 0 \Rightarrow D$ ✓

$0.32g - 1.5 > 0 \Rightarrow \underline{S}$ ✓

75 . 76 . . . 100 $\Rightarrow$ 26

76 . . . . 100 $\Rightarrow$ 25

① post Q29

② fix X7 add Q29

③ Zoom recording + slides

⋮

④ update grades.

Question 10

• [4 points] Given the following game payoff table, suppose the the row player uses a mixed strategy playing U with probability p, and column player uses a pure strategy. What is the smallest and largest value of p in a mixed strategy Nash equilibrium?

Row \ Col	L	R
U	9, 5	9, 0
D	9, 0	0, 10

$$\begin{cases} 5p \geq 10(1-p) \\ L \geq R \end{cases} \rightarrow p \geq \frac{10}{15} = \frac{2}{3}$$

$$\in [\tfrac{2}{3}, 1]$$

$$\begin{cases} U \geq D \\ U = D \\ U \leq D \end{cases} \quad q \geq 1, \quad q = 1, \quad q \leq 1$$

$$\begin{cases} L \geq R \\ L = R \\ L \leq R \end{cases} \quad p \geq 1, \quad p = 1, \quad p \leq 1$$