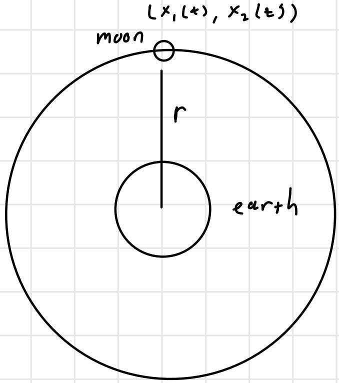
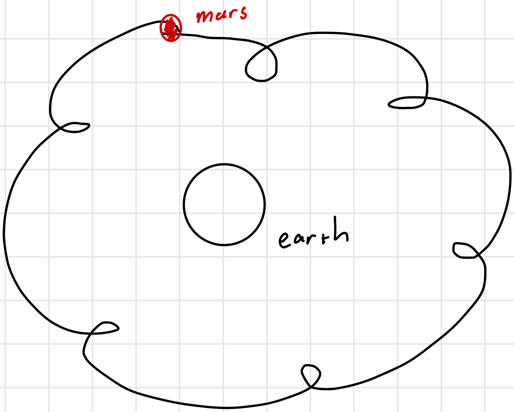
Lecture 1

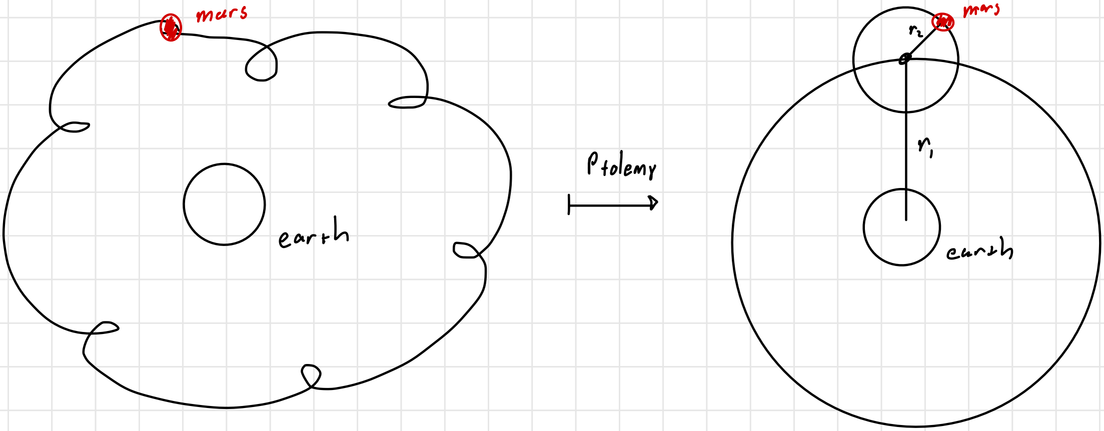
- Topics:
  - Regression
  - CART, adaboost, XGB
  - kernel regression, Gaussian Process, SVM
  - Deep learning: MLP, SGD, CNN, RNN, Transformer, GPT, BERT, generative: GAN, VAE, Diffusion
  - Reinforcement Learning: MDP, policy, value, AlphaGo, policy gradient, Q-learning, Decision Transformer.
- Coursework:
  - biweekly hw: coding & theoretical
  - coding Python / PyTorch

- Machine Learning in ancient time:
  - Astronomy: Observe positions of planets $\longrightarrow$ Predict motion
    - Ptolemy Epicycle:



$$X(t) = x_1(t) + i\, x_2(t)$$
$$= r e^{i\omega t}$$





$$X(t) = r_1 e^{i w_1 t} + r_2 e^{i w_2 t}$$

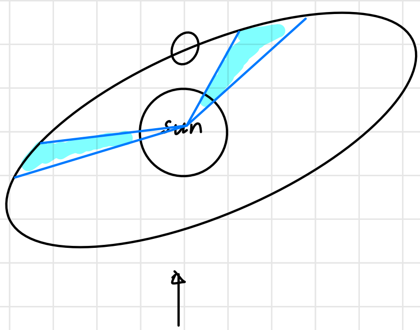
- Then for $r_1 > r_2 > \cdots$ we have

$$X(t) = r_1 e^{i w_1 t} + r_2 e^{i w_2 t} + r_3 e^{i w_3 t} + \cdots$$

boosting

  - We can draw an analogy with the MLP, i.e. adding a perceptron on top of a perceptron

- Kepler:



Simpler model under this re-representation
But Ptolemy's model is much more general

- Newton:

$$F = ma$$

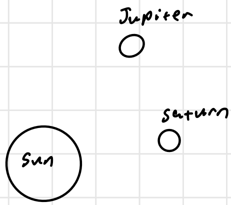
observed

force | mass | acceleration

latent variables

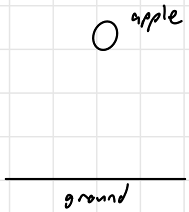
$$\underbrace{F = G\frac{m_1 m_2}{r^2}}$$

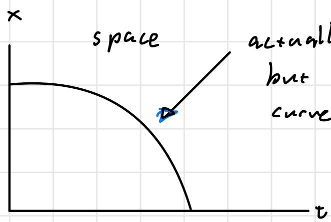
This is more general. Can explain 3-Body interactions.



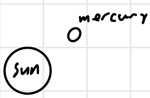
action at a distance as real as epicycle (unreal)

- Einstein:





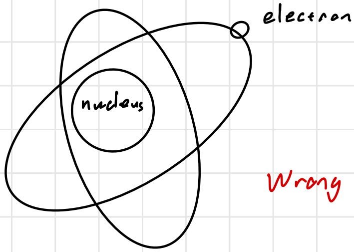
actually a straight line, but earth as a mass curves space time.



Einstein can model this better than Newton

But just as Ptolemy's epicycles are imaginary so is Einstein's notion of space-time

- Quantum



- Schrodinger :

$$V_{t+\Delta t} = (I + A\Delta t)\, V_t$$: Linear RNN (hidden layer)

$A = -iH/\hbar$ Hamiltonian, (quantized) discrete eigen values

embedding (state vector) thought vector

even more unreal than epicycle

query vector

key vector

- Born : $p_t(x) \propto |\underbrace{\langle u(x), V_t \rangle}|^2$: emission (output layer)

observed state

$\phi_t(x)$: wave function, not a physical wave

- Bohr, Heisenberg, Pauli : Copenhagen interpretation

<u>observer</u> outside system, collect data from system

Quantum mechanics is not responsible for explaining "reality" beyond observed data

- Machine Learning in Modern Time

  - Jan De leeuw : everything is regression

- regression (general meaning)
  - regression (special meaning) output: continuous
  - logistic regression (classification) output: categorical

more or less supervised learning

	input	output
1		
⋮		
i	x_i	y_i
⋮		
n		

- Gauss Paradigm :
  - inverted Least Squares Method : Used to predict Ceres
  - Gauss distribution : maximum likelihood
  - Gauss - Markov → optimality : LS is the best linear unbiased estimator

- Chat-GPT:
  - text: prompt

    $x_0, x_1, x_2, \ldots, x_{t-1}, x_t$ ← tokens (50 k of them)

    word
  - learn generative model $p(x_t \mid x_{<t})$, auto-regressive model.

    output input
  - $\max_{\theta} \underbrace{\mathbb{E}_{data}[\log p_\theta(x_t \mid x_{<t})]}$ maximum likelihood

    The negative of this gives us cross-entropy loss.

Similar to a super parrot → Oracle

- Diffusion model:

$$X_0 \longrightarrow \cdots \longrightarrow x_t \longrightarrow x_{t+\Delta t} \longrightarrow \cdots \longrightarrow X_t$$

$$X_{t+\Delta t} = X_t + \sigma\sqrt{\Delta t}\,\varepsilon_t, \quad \varepsilon_t \sim N(0, I)$$

learn $p(x_t \mid x_{t+\Delta t}) \sim N(f(x_{t+\Delta t}, t), \sigma^2 \Delta t I)$

$$\max_\theta \mathbb{E}_{data}[\log p_\theta(x_t \mid x_{t+\Delta t})]$$

$\updownarrow$

negative of likelihood gives us least squares loss.

$p_\theta(x_t \mid x_{t+\Delta t}, \text{text input}) \longrightarrow$ generating arts