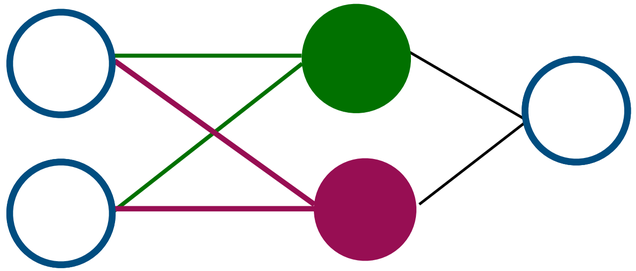
Neural Networks



- Website: https://www.cs.virginia.edu/~xay7te/courses/neuralnets/index.html
  - news, homework, and slides and …
- Office hours: Mondays 1:30-2:30 and Tuesdays 4-5pm in Rice Hall 105

Hadi Daneshmand

University of Virginia

Teaching Assistant

- We are fortunate to have Oishee Bintey Hoque in our teaching team.

https://oishee-hoque.github.io

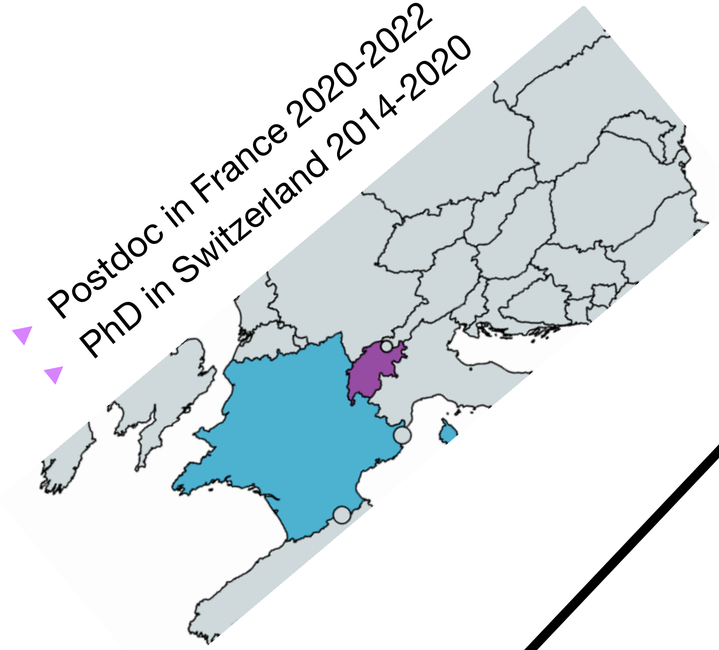
3rd PhD at Biocomplexity Institute

Introducing myself

- I joined UVA last December
- Background: +12 research in ML

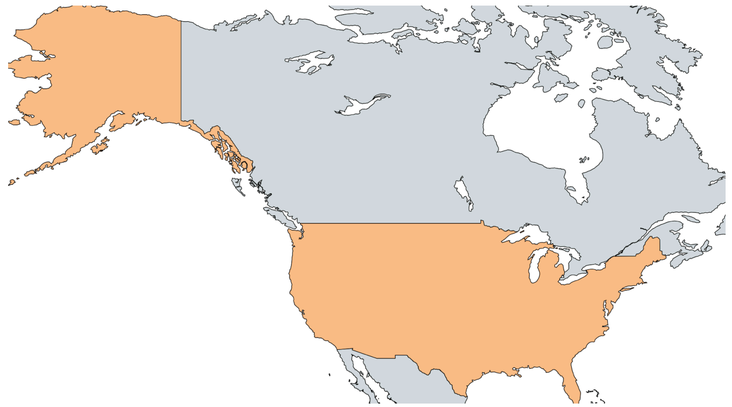
Europe

- Postdoc in France 2020-2022
- PhD in Switzerland 2014-2020



Asia

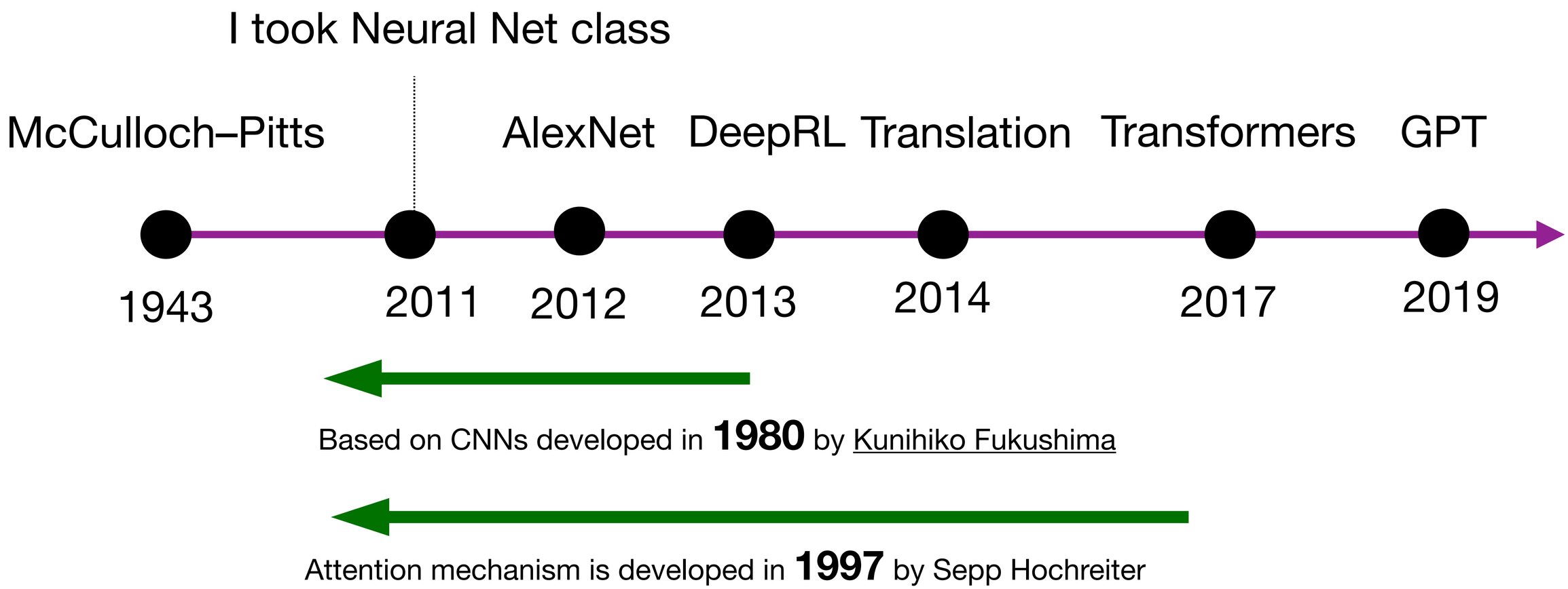
- Postdocs
  - Princeton University 2022
  - MIT and BU since 2022



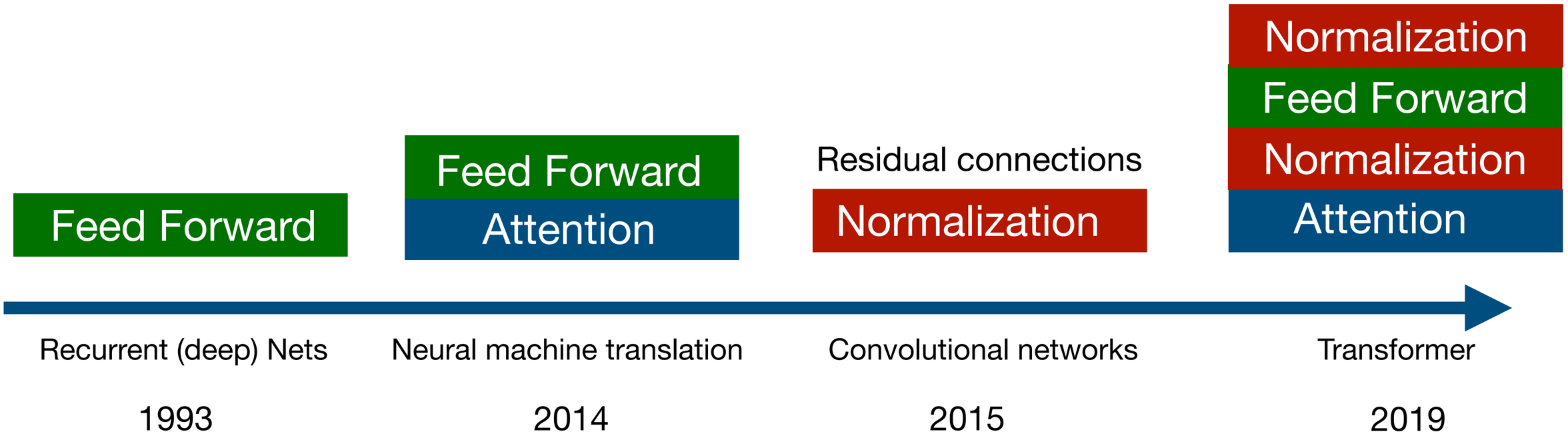
USA

- M.S. in Iran 2011-14

Neural Nets: New or Old?



Development of language models components



We want to look under the hood

Example 1: prompt engineering

- Prompt Engineering: How to talk to ChatGPT?

Which component does contribute to prompt engineering?

Without prompt engineering	Engineered prompt
Prompt: compute 1-2+3*2	Prompt: compute 1-2+3*2 step by step
ChatGPT: 4 ✗	ChatGPT: 3*2 = 6, -2+3*2 = 4, 1-2+3*2 = 5 ✓

Example 2: in-context learning

In-context learning: learning from examples in the prompt

How does in-context learning work?

Example: swapping characters

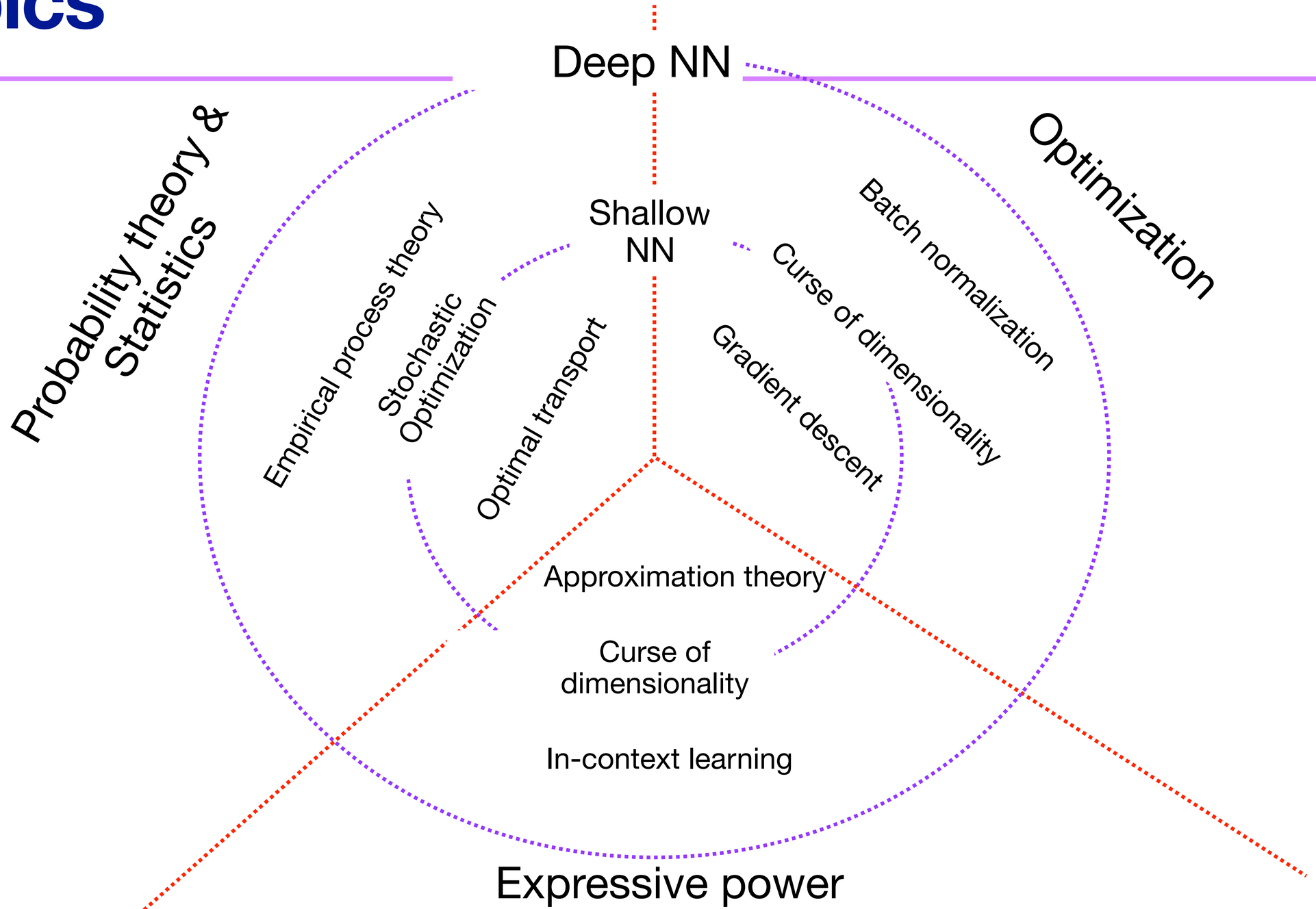
Prompt: ac ca ; qm mq ; mn ?

GPT3: nm

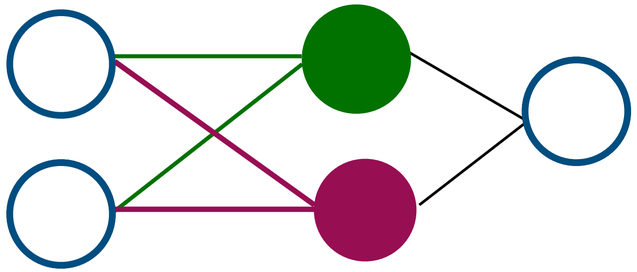
Goals

- Developing skills to conduct core research on neural networks
- Providing a comprehensive mathematical foundation for modern ML

Topics

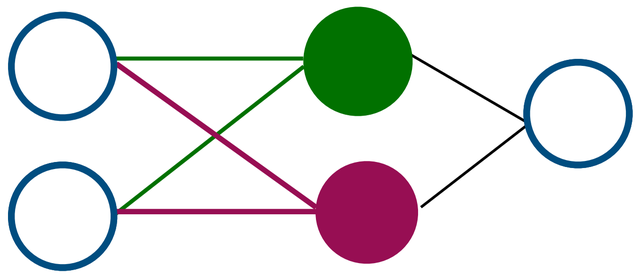

Deep NN
Shallow NN
Probability theory & Statistics
Optimization
Empirical process theory
Stochastic Optimization
Optimal transport
Batch normalization
Curse of dimensionality
Gradient descent
Approximation theory
Curse of dimensionality
In-context learning
Expressive power

What do we do in class?



We are detectives，investigating neural networks

Neural Networks : A Theory Lab

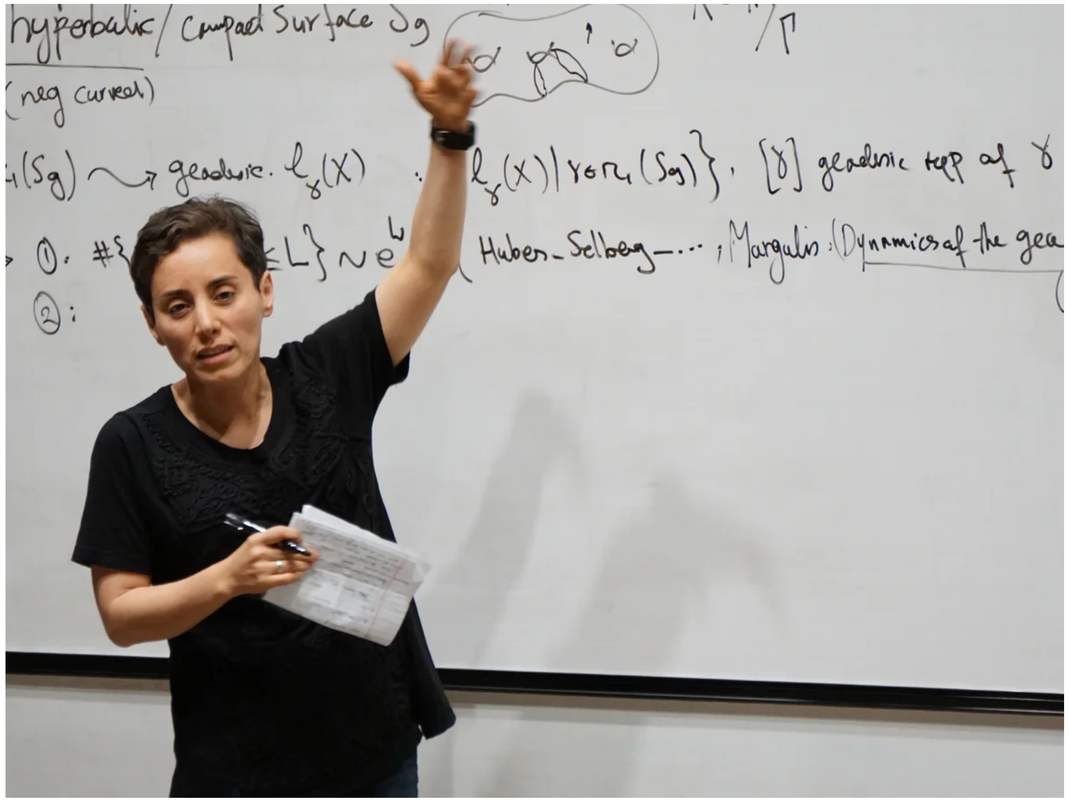


Hadi Daneshmand

University of Virginia

A Theory Lab ?

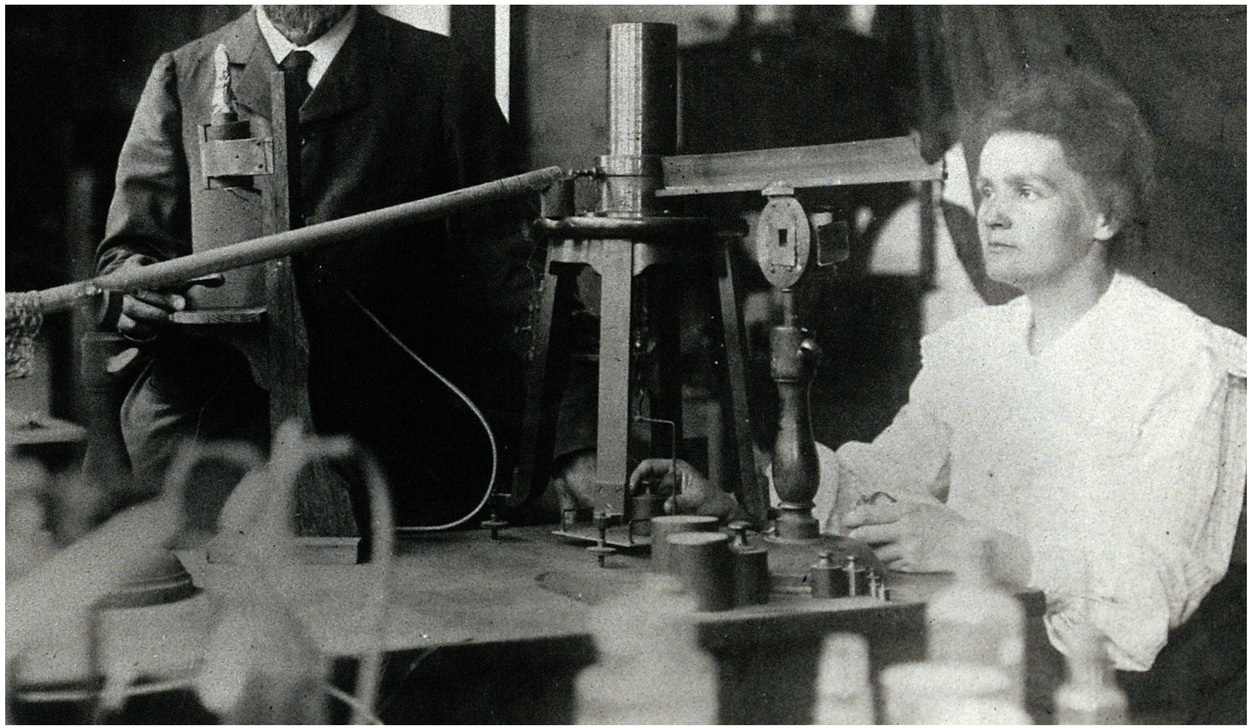
Theory



theguardian

Lab

Example



wikipedia: Marie Curie

Lab

Get ready for a data scientific puzzle

Group assignment

- Your table number =your birth day (as a numerical day of the month) modulo 10.
- Example: I was born in 26 December so my table number is 26 mod 10 = 6.

Advices

- Do **not** use ChatGPT
- Think-aloud
- Share your idea before coding
- Ask questions

A statistical puzzle

- Given: Two sets of points
  - $x_1, \ldots, x_n$ where x_i are 6 dimensional random vectors
    - Drawn independently from the same distribution (i.i.d: independent and identically distributed)
  - $y_1, \ldots, y_n$ where y_i are 6 dimensional random i.i.d. vectors
- **Question**: Are these points from the same distribution?

Where are points?

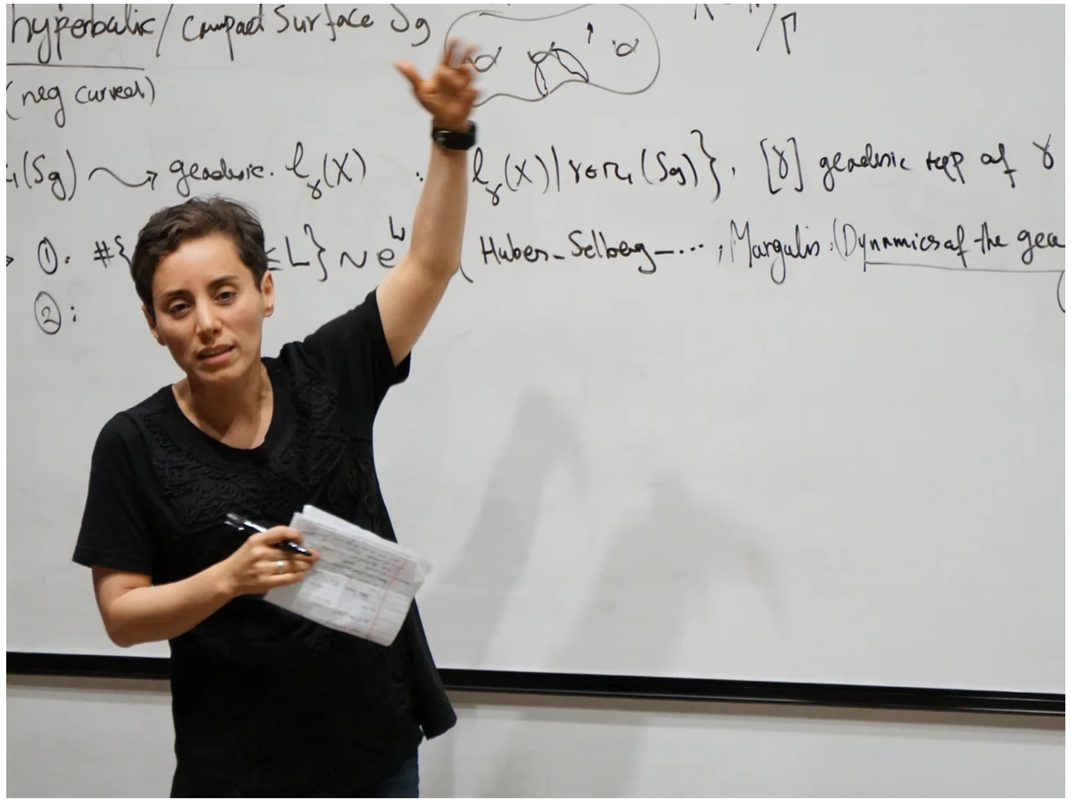
- Given: Two sets of points: $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$
- **Here:** https://shorturl.at/QfRTI

7 minutes of in-group discussions

- **Question**: Are these points from the same distribution?

- https://colab.research.google.com/drive/1i9coc5Wnp6ODQ1SsDzNWj46Eg6aLJHZI#scrollTo=iyqFsOmB6pV4

Ideas?



theguardian: Maryam Mirzakhani

Theory

Get ready for math

Start Simple

Suppose we have an infinite number of samples

Step 1: what about checking averages?

- Recall central limit theorem: $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mathbb{E}[x]$
- If x_i s and y_i s are from the same distribution then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mathbb{E}[x] = \mathbb{E}\left[y\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i$$

- We can check averages are almost the same
- Can we conclude distributions are the same? - No

Step 2: moment matching

- Suppose that $\frac{1}{n}\sum_{i} x_i x_i^\top = \frac{1}{n}\sum_{i=1}^{n} y_i y_i^\top$, then are the distributions the same?
- Non-linear moment matching: given a non-linear function $\phi : \mathbb{R}^6 \to \mathbb{R}^m$,

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \phi(x_i) - \frac{1}{n}\sum_{i=1}^{n} \phi(y_i) \right\|^2 = 0$$

- then are the distributions identical?

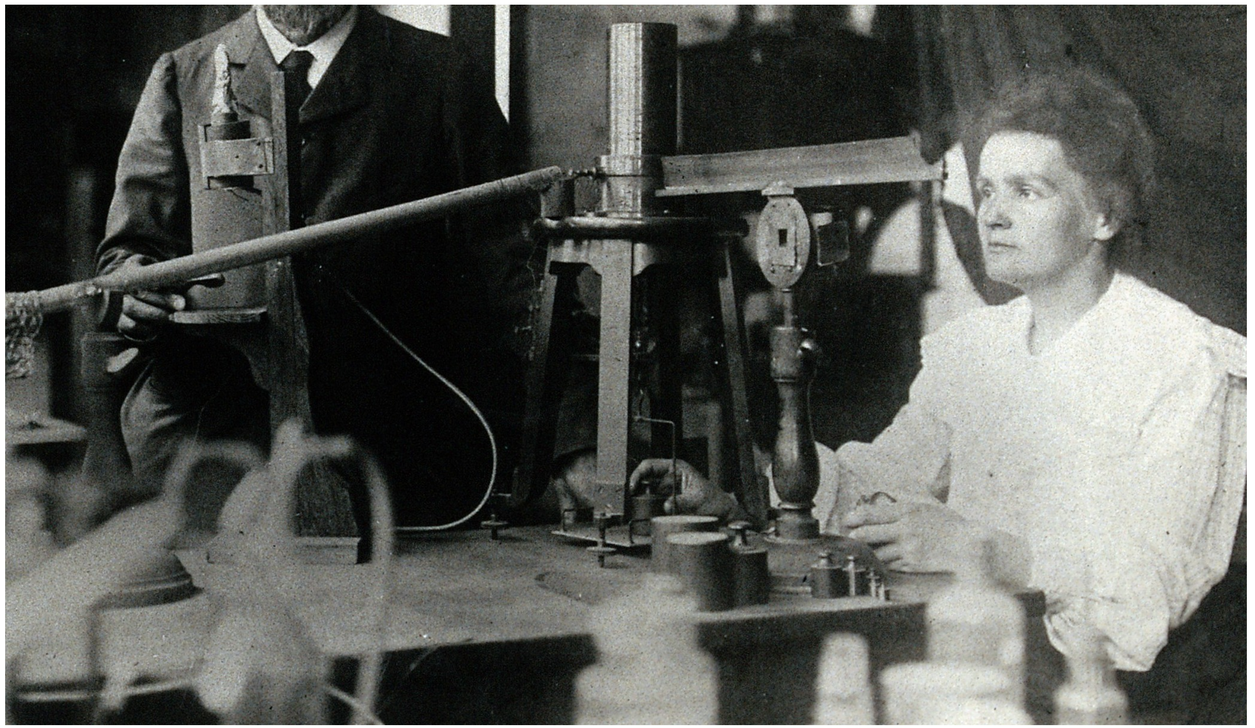
Step 3: Algebra

- $$\left\| \frac{1}{n}\sum_{i=1}^{n}\phi(x_i) - \frac{1}{n}\sum_{i=1}^{n}\phi(y_i) \right\|^2 = \frac{1}{n^2}\sum_{ij}\overbrace{\langle\phi(x_i),\phi(x_j)\rangle}^{k(x_i,x_j)} -$$
$$-\frac{2}{n^2}\sum_{ij}\langle\phi(x_i),\phi(y_j)\rangle + \frac{1}{n^2}\sum_{ij}\langle\phi(y_i),\phi(y_j)\rangle$$

Maximum Mean Discrepancy (MMD)

Key theorem

- Given two distributions μ and ν, define
  $MMD_k(\mu, \nu) = \mathbb{E}_{x\sim\mu, x'\sim\mu} k(x, x') - 2\mathbb{E}_{x\sim\mu, y\sim\nu} k(x, y) + \mathbb{E}_{y\sim\nu, y'\sim\nu} k(y, y')$
- Gaussian kernel: $k(x, y) = \exp(-\|x - y\|^2/2)$
- Theorem (Gretton et al. 2012). For Gaussian kernel, $MMD_k(\nu, \mu) = 0 \implies \mu = \nu$
- Conclusion: There is a $\phi : \mathbb{R}^6 \to \mathbb{R}^\infty$ such that
- $\frac{1}{n}\sum_{i=1}^{n} \phi(x_i) = \frac{1}{n}\sum_{i=1}^{n} \phi(y_i) \implies$ distributions are the same



wikipedia: Marie Curie

Lab

Going back to experiments

Compute MMD distance

- For dataset at: https://shorturl.at/QfRTI
- Compute $\frac{1}{n^2}\sum_{i,j=1}^{n} k(x_i, x_j) - \frac{2}{n^2}\sum_{i,j=1}^{n} k(x_i, y_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(y_i, y_j)$

$$k(x, y) = \exp(-\|x - y\|^2/2)$$

5 minutes of group activity

- https://colab.research.google.com/drive/1i9coc5Wnp6ODQ1SsDzNWj46Eg6aLJHZI#scrollTo=iyqFsOmB6pV4

Compute MMD distance for identical distributions

- Given: Draw $x_i, y_i \sim \mathcal{N}(0, I_d)$
- Compute $\frac{1}{n^2}\sum_{i,j=1}^{n} k(x_i, x_j) - \frac{2}{n^2}\sum_{i,j=1}^{n} k(x_i, y_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(y_i, y_j)$
- Compare with the computed MMD distance in the pervious slide

Result

- For identical Gaussian distributions , MMD = 0.01
- For the given dataset, MMD = 0.37
- Conclusion: Data distributions are not the same (with high probability)

Summary

Theory

A kernel method approach

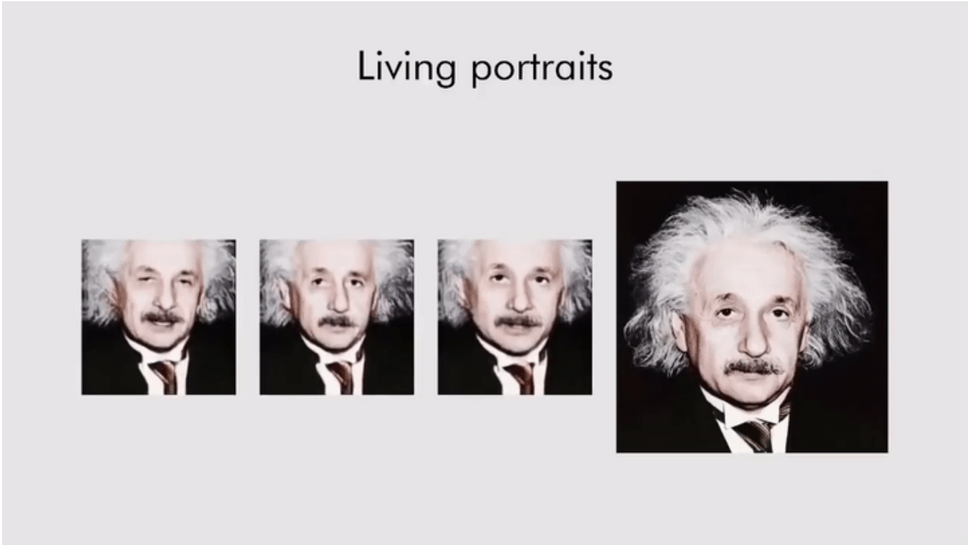
A Kernel Two-Sample Test Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, Alexander Smola; 2012.

Lab

Two-sample test

Applications in neural networks

- Practical: Generative models such as generative adversarial networks



Demo credit Ram Sagar results by Egor Zakharov et al.,

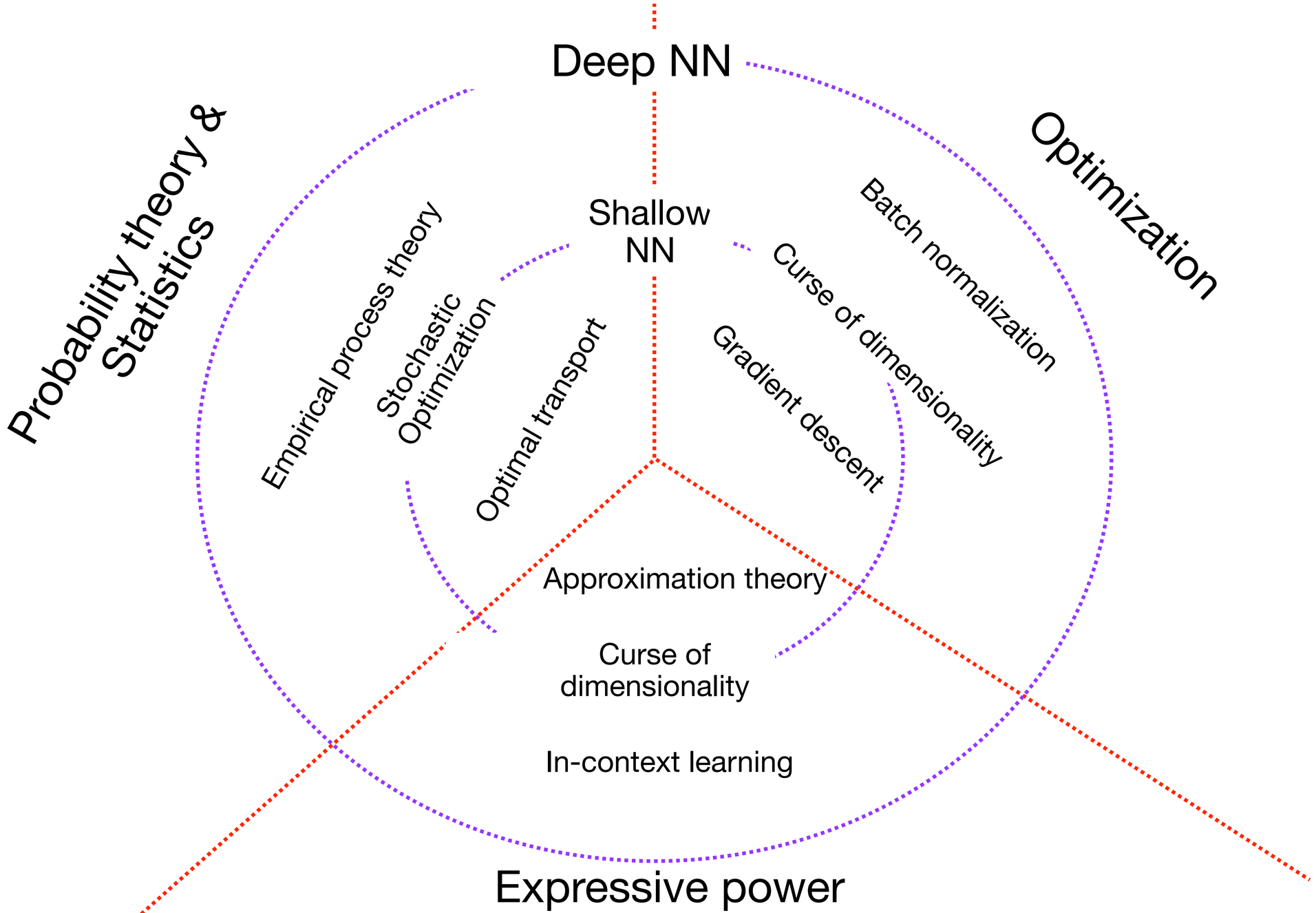
- We will discuss details

- Theoretical: Training analysis for single-layer neural networks

Evaluation and grading

- Homework (60%) : In groups of 2
  - Every 2-3 weeks
- Project (40%): in groups of 4-5
  - Literature review
  - Observations and implementations
  - Analysis
  - Final presentation and report
- Bonus (+20%): contributions to **wikipedia** and **scribing lecture notes**

Thank you very much!


Deep NN
Shallow NN
Probability theory & Statistics
Optimization
Empirical process theory
Stochastic Optimization
Optimal transport
Gradient descent
Curse of dimensionality
Batch normalization
Approximation theory
Curse of dimensionality
In-context learning
Expressive power