Week 3 Discussion

Made by: Sirui Tao, Samuel Lee, Parth Sandeep

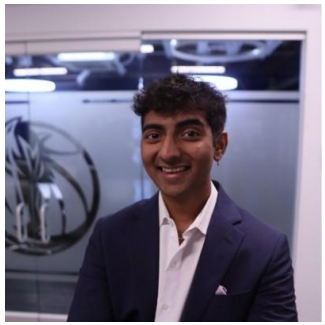
About us

Samuel Lee

- 2nd year
- Data science / Cog Sci in ML/NC

Sirui Tao

- 1st Year CSE Master
- B.S in Data Science & Statistics



Parth Sandeep

- 3rd Year
- Cog Sci In ML/NC

Forming team

Time: 5 mins

Form 4 person groups

Potential project directions

Time: 5 mins

- Discuss about research questions you are interested in
- Look for appropriate dataset
  - You can go to the course website to find example datasets and past students' work
  - Websites to find datasets: Kaggle, Google Dataset Search, Data.gov
- How you would analyze the dataset
  - You don't have to carry out the analysis
  - It's okay to change the question or dataset later

Question 1

1. Do you want to pursue a career in Data Science?
2. How do you want to pursue a career in Data Science?
    - Area of interest
    - Industry/ academia
    - …

Know the team & Discuss team projects

Time: 10 mins

1. **Have each other's emails**
2. **Create a chat channel**: Messenger, WhatsApp, Discord, Slack, …
3. Create a google folder
4. Setup a when2meet and decide a time that works best for the team before end of this week

(Scan QR Code to record your team - once per team)

Reading 1 Recap

- Any Comments?

- General Confusion?

Reading 1 Key Takeaways

- 6 Divisions of data science
- 2 Cultures in modeling: Prediction vs inference
- Common task framework: The secret sauce of the predictive culture
- Statistics vs Computer science vs Data Science

(Slide by: Matt Feig)

6 Divisions of Data Science

- Data exploration
- Data representation and transformation
- Computing
- Modeling
- Data viz
- The science of data science

(Slide By Matt Feig)

Prediction vs Inference

Prediction

- Accuracy with new data.
- "Black box" methods like neural networks.
- Located in computer science fields.

Inference

- Understanding input-output relationships.
- Creates interpretable, well-fitting models.
- Belief in a 'true' underlying data model.
- Central to traditional statistics.

(Slide By Matt Feig)

Common Task Framework

Key to predictive modeling:

- Reducing prediction errors over time.
- Best algorithms vary with dataset type.
- Focus on results, not the data's true source.

(Slide By Matt Feig)

Question 2

Any feedback? Suggestions for improvement?

Link to Monday Discussion slides

https://github.com/DylanTao/cogs9_wi2024