I remind you that you have to mail me your topic and teammates by 5/30, **with your date of preference to discuss your topic (read below)**.

Here are the specifications for your topic discussion & final presentation:

For topic discussion (5/31, 6/2),

- You have to present in three-page slides.
- The slides will include:
  * Topic
  * Methods you plan to use (either from course or from outside)
  * How you expect your final result would look like, doesn't have to be very specific

For your final presentation (6/7, 6/9),

- You have to present in 5-10 slides.
- The slides will include:
  * Problem definition
  * Data
  * Methods you used
  * Results
  * Discussion of your work

021
oject

Term project schedule

- 5/30
  - Submit your topic and teammates by email to TA
  - You can choose among 2 topics or devise a novel project and discuss
- 5/31-6/2
  - Term project discussion
  - We will discuss direction of your term project
- 6/7-9
  - Project presentation

Reminder of HW1

- Practice session on diabetes dataset from sklearn
  - linear regression (LASSO, Ridge), PCA, t-SNE
- HW1 using breast cancer dataset from sklearn
  - much the same
  - inspect coefficients
- Project topic 1 would be extension of it

Project 1: breast cancer

- Breast cancer dataset from HW1
  - 569 data (2 class: 212 malignant, 357 benign)
  - 30 numerical features
  - provided with sklearn.datasets package

- Task
  - Use dimension reduction techniques (PCA, t-SNE, ...) or deep learning techniques to classify malignant/benign cancers and interpret the analysis

Project 2: drug pattern mining

- DILIrank toxicity dataset from HW3
  - 835 FDA-approved drugs
  - skeleton code with augmentation and tokenization provided
  - target can be liver toxicity or severity

- Task
  - Use frequent pattern mining algorithm to find candidate motif sequence and classify drugs according to the found features, and interpret the analysis