Data Mining

95-791

Team Project Rules

Artur Dubrawski

awd@cs.cmu.edu

Basics

- Do not co-operate between teams!

- Each team will receive specific data and (initial) objectives of analysis

- Each project will be assigned a "Designated Client"

- Use analytical tools of your choice: We <u>do not</u> insist on using any specific software tool (such as R) however we may be unable to assist you with technical issues that may arise when using exotic tools

Deliverables: Milestone reports

- Two milestone progress reports are required
- Each team will submit 1-2 page summaries, due on:
  - #1: Friday, February 10th, 3:00pm
    - Problem statement and specific goals
    - Approach and plan of work
    - Risks and contingencies intended to mitigate risks
  - #2: Friday, February 24th, 3:00pm
    - Status and accomplishments
    - Risks and contingencies intended to mitigate risks
    - Next steps

  Required form of filenames: "2017S_X_MY.pdf" – substitute your team letter code for X, and the milestone number (either 1 or 2) for Y

- The milestone reports should demonstrate:
  - #1: Understanding of business context of the assignment
  - #2: Understanding of technical approach to accomplish business goals

Deliverables: Final report

- Each team will prepare one written report of the results
  - **Up to 10 pages**, single-spaced, minimum font size 10pt, one inch margins on each side of the Letter size paper sheets)
  - Please number pages and place your team name on each page
  - List names of all team members on the cover page
  - Cover page and list of references cited (if any) do not count against the 10 page total

- Written reports due on Monday, March 06th, 3:00pm in <u>pdf format</u> via the course blackboard
  - Required form of filenames: "2017S_X.pdf" – substitute your team code for X

- Important parts of your written report include:
  - Executive Summary
  - Recommendations for Continuation Work

- Hints: rely on data, business intuition, logical thinking, be neat, convey competence

Auton Lab

Carnegie Mellon
Heinz College

Deliverables: Presentation

- Each team will produce a demo/movie presentation playable on a PC, **exactly 5 minutes** in duration
  - It could be simply a recording of a PowerPoint presentation with voiceovers, but more sophisticated solutions are welcome as long as they can be played on any PC

- Inspiration: "Lightning" talks, "Ignite" talks
  - http://perl.plover.com/lt/osc2003/lightning-talks.html
  - http://www.speakerconfessions.com/2009/06/how-to-give-a-great-ignite-talk/
  - Note: Unlike in e.g. Ignite format, we do not insist on exactly twenty slides

- Movies are due on Monday, March 06th, 3:00pm (please submit via blackboard)
  - Required form of filenames: "2017S_X.???" or – substitute your team code for X and use either wmv, mpg, flv, or swf for ???

- Presentations will be made available to other teams for viewing and peer voting
  - You will not be able to vote for your own presentation

Auton Lab

Carnegie Mellon
Heinz College

Deliverables: Client Interactions

- Do not hesitate to interact with your "Clients"!

- Required interaction: <u>at least 2 review sessions</u>
  (with the assigned Clients)
  - You are responsible for scheduling these meetings
  - Best to meet with both of your client representatives at each time

- Come prepared to these meetings
  - Professionalism and preparedness will be considered in grading

Deliverables: Client Interactions

- Do not hesitate to interact with your "Clients"!

- Required interaction: <u>at least 2 review sessions</u>
(with the assigned Clients)
  - You are responsible for scheduling these meetings
  - Best to meet with both of your client representatives at each time
  - Suggested scope:
    - <u>Interaction 1</u>: Verify understanding of the problem and goals, main takeaways of initial data exploration, execution plan and business questions
    - <u>Interaction 2</u>: Status, preliminary results, show value added to the business through your analysis, and future work
  - Come prepared to these meetings
    - Professionalism and preparedness will be considered in grading

Team rosters

Strong advice:

- Arrange a meeting with your teammates as soon as possible

- Identify your goals and approaches and have them reviewed as soon as possible with your Clients

- Analyze and understand before jumping into implementations
  - First: think, consider alternatives, plan

Grading

- The project is worth 30% of the total attainable course grade points

- Grades:
  - Based on reports, presentations, and the quality of the interactions with your Clients
  - Grades uniform across the team, unless at least one student requests otherwise
    - The request must be made in writing (an email is sufficient), justification is not required
    - In such case, a peer-review grade component will be implemented to reflect varied individual commitments and contributions of team members
    - Results of a peer review may not lead to grade increases

Data Mining Project Expectations

Recommendations for Meetings,
Presentations and Reports
by the
Data Mining TA Team

Data Mining Project Expectations

INTRODUCTION

Introduction

- Note that general **project rules** are published on blackboard under Assignments/

- This document discusses
  - How the project fits into the course
  - Expectations regarding client meetings
  - Expectations regarding final deliverables
    - Demo/Movie presentation
    - Final report

This Project Emulates Real Life

- May/June 2011 article from Analytics (INFORMS). Top challenges faced by Data Miners:

Key challenge No. 1: Dirty Data. Eighty-five data miners described their experiences in overcoming this challenge. Key themes were the use of descriptive statistics, data visualization, business rules and consultation with data content experts (business users).

Key challenge No. 2: Explaining data mining to others. Sixty-five data miners described their experiences in overcoming this challenge. Key themes were the use of graphics, very simple examples and analogies, and focusing on the business impact of the data mining initiative.

Key challenge No. 3: Difficulty accessing data. Forty-six data miners described their experiences in overcoming this challenge. Key themes were devoting resources to improving data availability and methods of overcoming organizational barriers. Some example responses:

http://viewer.zmags.com/publication/dd66abfd#/dd66abfd/28

Structure of the Project
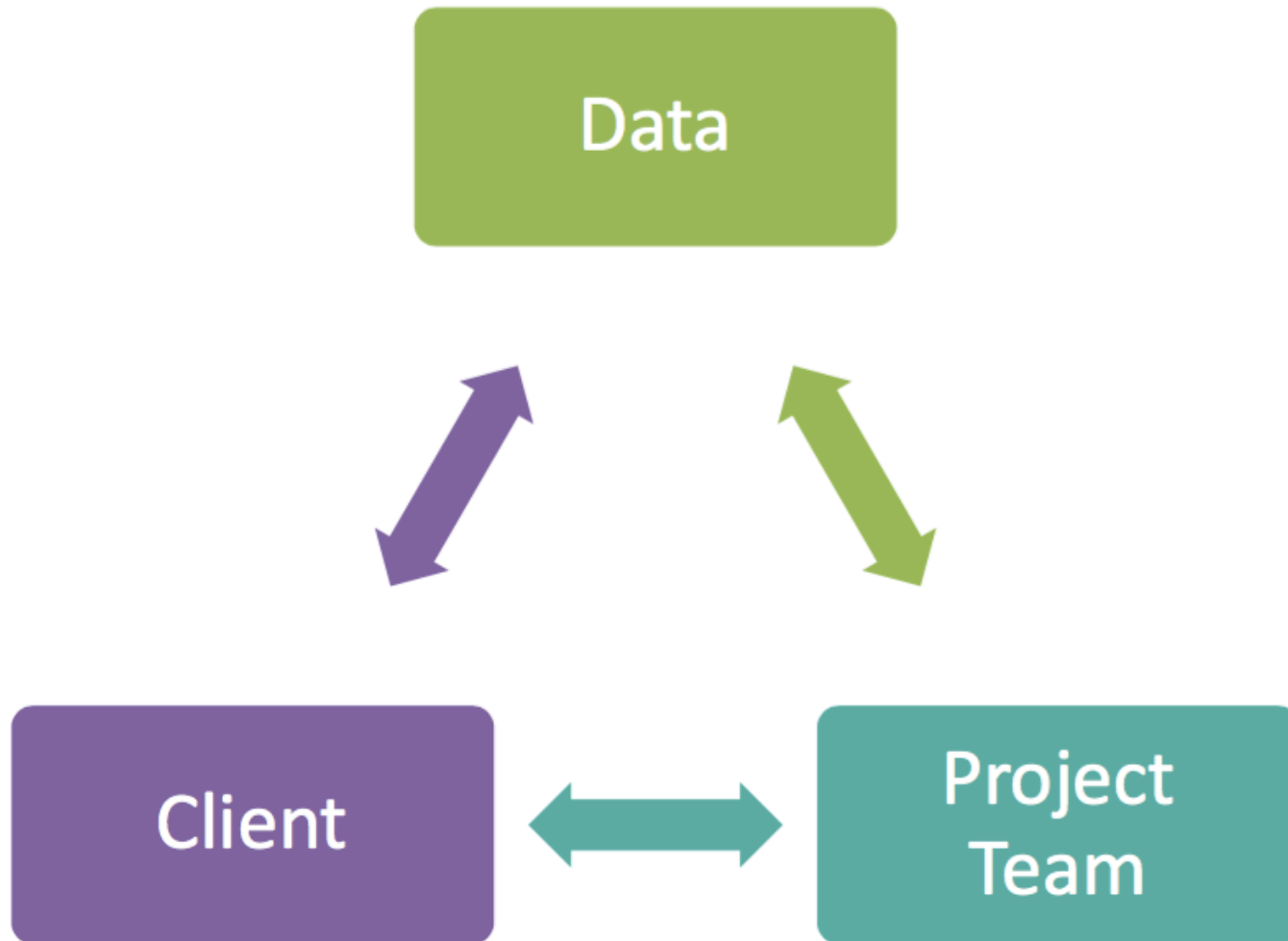
Project Management

- Start early and plan
  - Data exploration, preparation for client meetings, milestone reports, data analysis, final report, videos, etc.
  - It is easy to tell if the final report was completed in a rush.
- Split the tasks wisely
  - feature selection or parameter tuning for different algorithms may take some time
  - Everyone should agree on evaluation metrics of technical analysis
  - Everyone should know about the data, project goals

Why Milestone Reports?

- Keep students from leaving everything to the last minute.

- Develop a better understanding of the data for both client and team.

- Provide a real opportunity to wrestle with the challenges data miners face.

Data Mining Project Expectations

CLIENT MEETINGS

Client Meetings

- Be well prepared for client meetings.
  - Schedule in advance.
  - Know the data and domain.
  - Slides are recommended in order to report project progress.
  - Schedule at least 2 review sessions with Client.
  - Project progress is expected for each meeting. Expectations differ from project to project.
  - Everyone should participate in the discussions.
- **Preliminary results** are expected at your **second** client meeting.

Client Meetings

- Assume your clients have a **business** background
  - Listen to their input, take it into consideration.
- Clients have **minimal knowledge** in data mining techniques and statistics.
- You are responsible for **convincing** them that your work and analysis can add value to their business.
- Your results can impact their business considerably. Gain trust by being clear and precise.

Client Meetings

- Not all analyses are interesting from a business point of view.

- An analysis is interesting only if it is **motivated** and **well presented**.

- Be open to the business question changing over time.

- Have questions ready, for the client and TAs.

Data Mining Project Expectations

PRESENTATIONS - BASICS

Presentations - Basics

- Get the simple things right for meetings as well as the final presentation. Here are some guidelines
  - Be professional. Simple is better than overloaded.
  - When you show a new graph/table, explain what it represents before discussing specific details.
  - Use readable fonts/colors, proper resolution, especially for plots.
  - Don't forget about legends, axes and captions

Data Mining Project Expectations

FINAL DELIVERABLES

Some Quality Guidelines

- Evaluate models along several performance metrics, not just accuracy.

- Quality trumps quantity.
  - E.g. explore different (hyper-)parameter settings instead of trying 10 different classifiers.

- Work detail-oriented (check graphs, tables, typos, etc.)

- What is the bottom line of your presentation? Just because you ran a lot of tests does not mean you've added a lot of value.

Final Deliverables: Presentation

- General requirements in "Project Rules"
- Strictly no more than 5 minutes in duration
- Things to consider:
  - Avoid spending too much time talking about the dataset and how you dealt with missing values (!)
  - Try to keep speed and volume consistent.
- Heed the general advice given in the previous slide on presentations.

Final Deliverables: Report

- General requirements in "Project Rules"
- What your report should answer:
  - What are the goals of your project?
  - How does the business problem relate to specific data mining tasks?
  - What models have you proposed and tried? Why did you make those choices?
  - What are the metrics you used to measure performance of different algorithms?
  - For each algorithm, what different sets of parameters have you tried?
  - How do you compare the results of different models?

Report continued

- What is your final model suggestion and why?
- What are the final results and how do we interpret them? Are they aligned with project goals?
- What is the business implication? Is it interesting to client?
- What actions should your client take based on your results?
- What about risk and mitigation plans?
- Provide recommendations for continuation of work.

- Overall: show us that you can apply what you have learned in class to a new problem and that you can put it into a business context.

Teamwork

- Each team member will be asked to complete a Team Contribution Peer Review at the end of project.
  - It will NOT influence the project grade. Project grade will be the same for every team member, no exception.
  - But students with significant low contribution evaluated by other team members may get points deducted off the course grade (not project grade).
- Motivating others to contribute is also part of teamwork skills.
- Teamwork is not limited to technical, hands-on contribution
  - brainstorming, feedback, project management etc. are also important components