PROJECT

Handed out: 10 Feb 2025 (Mon)
Deadline 1 (Group Formation):   16 Feb 2025 (Sun), 11:59 PM
Deadline 2 (Project Proposal):   1 Mar 2025 (Sun), 11:59 PM
Deadline 3 (Final Report & Code): 13 April 2025 (Sun), 11:59 PM
Deadline 4 (Presentation Slides): 14 April 2025 (Mon), 5:30 PM
Deadline 5 (Presentation):   14 April 2025 (Mon), during lecture
Deadline 6 (Peer Review):   15 April 2025 (Tue), 11:59 PM

1 Types of Projects

You are free to propose a project on the application of machine learning methods to problems that you are interested in. The choice of the topic is up to you, provided that it is relevant to the course material; you are allowed to utilize advanced machine learning techniques that are not directly covered in the course but are related to the course material that have been taught. Below are some examples of potential course projects to spark your creativity.

(a) Prediction of types of tweets in a social network.
(b) Prediction of asset prices based on historical prices, variation, etc.
(c) Prediction of how customers would rate books, movies, products, etc.
(d) Prediction of results of sports events based on player statistics, team composition, etc.
(e) Discover hidden types in high dimensional biological data (e.g., data from high-throughput sequencing).
(f) Document classification (e.g., email and news)

In choosing your project, please bear in mind the computational and data resources that you would be able to garner. Your project must be implemented in Python 3.

It is crucial that you delve into and thoroughly analyze the workings of machine learning techniques within the context of your chosen problem. For example, if you merely take a pre-existing boosting package, rely on its default parameter optimization, and simply report its performance on a classification task, then you are likely obtain a mediocre grade (at best). Consider the peculiarities for your chosen problem: What are the alternative feature representations and what made you decide on which to choose? Are the class frequencies skewed and how did you resolve that problem? What makes your chosen machine learning algorithm(s) suitable for your problem? (Novelty and creativity are valued for the final project. If you simply use a run-of-the-mill Kaggle competition as your final project, then you will simply get a mediocre grade.)

Finally, remember to *enjoy* the project ... this is your chance to explore machine learning methods for real on tasks that you are interested in.

2 Group Size

You are to work on this project in **four-person** groups.

All members of each group are expected to contribute actively to the project. A peer review will be conducted at the end of the project in which each member confidentially assesses the contributions of the others. If a member is merely a 'passenger' throughout the project, cruising along without active participation, the peer review process will bring such behavior to light and result in a deduction of the offending member's individual grade.

3   Grade Breakdown

This group project is worth **40%** of your final IS5126 grade. The grade breakdown for the project is shown below. We reserve the right to exercise discretion on how points are allocated.

Item	Weightage (Total: 40%)
Group Formation	1%
Project Proposal	2%
Final Report & Code	21%
Presentation	16%

Each student is required to submit an individual peer review of his/her group-mates. The peer review is not worth any points, but if a student fails to submit his/her peer review by the stipulated deadline, an individual penalty of **5%** will be imposed (i.e., the most that student can get for the final project is 35/40).

3.1   Deliverables

Only one submission per group is required for each deliverable (except for the peer review).

(a) By Deadline 1 mentioned at the top of the handout, please write your group name and the names of your group members in a PDF document, and submit it on Canvas. (This is a 'freebie'; do not be late so that you can get the allocated points in full.)

(b) By Deadline 2, submit your project proposal via Canvas. Clearly, describe the problem you are addressing, the machine learning method(s) you would be using, and the data that you would be experimenting on. Also mention the difficulties you foresee and how you would overcome them. (In all likelihood, there will be several bugbears that you will *not* be able to anticipate, so start on your project early to flush these out as soon as possible.) Your proposal should be between half a page to a full page. (Like the first deliverable, this is a 'freebie', so do not be late to get the allocated points in full.)

(c) By Deadline 3, submit your final report and code via Canvas. The report should include, but is not limited to:

   - A detailed exposition of your machine learning algorithm(s). You have to demonstrate that you understand how your model works, be able to explain why it works well (or not), and explain you how arrive at your model. Please do not simply say, "We threw everything but the kitchen sink at the problem, and model X works best...but I don't know why."

   - Different approaches that you have tried (if any), and their respective performances.

   - Clearly state your data source(s) and how you obtained them. (If you use data from Kaggle or other online sources, clearly specify their URLs.)

   - A description of your manipulations of the data (if any).

   - A detailed description of how you train and tune your model on the training set. (You should **never** train or tune your model on the test set.) A step-by-step guide and data should be included as a *separate appendix*. This allows us to replicate how you obtained your final model. **(You will be penalized if your model and results cannot be replicated.)**

- Performance on the training data and test data. **(You will be penalized if these numbers cannot be replicated when we run your code.)** We reserve the right to test your algorithm on additional data.

The final report should not exceed 5 pages (Times 12 pt font; this does not include the appendix mentioned above).

(d) By Deadline 4, submit your presentation slides via Canvas.

(e) During the last lecture, give a 10-min presentation of your final project and field a 5-min question-and-answer session. Be polished, concise and clear in your presentation. **ALL** members of the group must participate in the presentation, and every member must be able to answer questions about the project. (The TAs and instructor may require a particular member of the group to answer any specific question.)