****

Butler University and Rose-Hulman Institute of Technology

Departments of Computer Science and Software Engineering

**Intro to Machine Learning - Online**

**Term Project Guide**

The separate course Syllabus describes, as three of the course outcomes:

* Possess a wide view of the places where ML systems vary in their results and in value.
* In a term (semester) project, be able to run different ML algorithms on data of your choice, data in some domain of personal interest to you.
* Exchange ideas and techniques with students in other disciplines.

This project description is to help guide you toward those outcomes, using a project as the means of achieving them.

The intent is for you to achieve these as a part of a team of about three students, total. This lets you build expertise working closely together, sharing experiences and helping one another, as you extend the basic lessons of the course to a slightly more difficult and interesting problem.

From the Syllabus:

**General:**  You will have a project, for which you get to pick the "data set" -- the data that you would like to learn about. You'll be expected to try this data out on many of the ML algorithms. Doing that, and writing up how it went, make up the term project. In the end, as you might imagine, the last assignment will be to talk about what worked best, and what didn't, and why you think that is. In case your team cannot find an available/interesting data set for the project (or the team has several choices but cannot decide) then please contact the instructor as soon as possible who can help you come to a resolution.

**Teams:**  We recommend that you do the term project as a team.  This way, if you get stuck on a particular aspect of using an algorithm on your project data set, there are more of you to help figure it out.  But, you will need to arrange to work together, and make the work load equitable.  Your instructor will help you set up these arrangements, as we start the course.

If, for various reasons, you will be doing this as an individual project, please contact us regarding how that should vary the discussion in this document!

Here are some of the pieces of the project that you will need to know:

**Grading:** As noted in the syllabus, overall, this project counts as 30% of the course grade.

**Expectations:**  The goals of the project are these:

* For you to try as many ML algorithms as you can, on data of your choice,
* Writing up the results in the Term Project Forums,
* Comparing each with your other results,
* Comparing with what other teams did on their projects, and analyzing why, and
* Coming up with a final presentation and report describing those outcomes.

**Ideas for the project:** At the start of the course, each person will be able to contribute ideas and datasets, for projects of interest to them.

**Process - Roles:** This will be a team-based project. You will evolve into some role on a team, where your expertise in some area becomes relied upon by the team. It is up to you and your teammates to decide on these areas for each person's contributions, so that the whole project moves forward, and each person feels like a relatively equal contributor. All roles should involve working with the ML algorithms, and with RStudio, in more than a trivial manner. You may elect to have specific roles with names, like:

* Project leader/manager,
* Shared repository owner,
* Forum contribution leader,
* Team meeting secretary, and
* Domain expert.
* Tester, Quality Assurance or Performance Evaluator

Setting up agreed-to roles is a part of establishing team coordination and logistics, by putting people in charge of things. It will be up to your team to make this work.

As this list suggests, likely most people will have multiple roles. And, for each person, overall, the project is to be a vehicle for them to learn about ML and R. Each person is expected to contribute to, and to be able to explain what is happening with, manipulating the data and coding, on your team project.

**Process -- Team and project formation:** We will define a process for joining teams and launching these. It is possible that more people will have project ideas than the number of possible teams, so be prepared also to work on a project conceived by someone else, besides the one you came in already wanting to do. Besides similarity of interests, team formation also will be related to ability to work together. E.g., having times and places where meeting with teammates is possible.

A part of team formation is agreeing on team rules, to go with the roles suggested, above. For instance, if a team meeting is agreed to, simply not showing up can be cause for notifying your instructor.

**Process -- Milestones:** This online course is divided into modules, which will be identified with a specific schedule. It is expected that your team will keep up with the schedule of modules, on a one-week-later basis, along the lines of this **example**:

* Week 5: Reading and homework for Milestone 5 are due. Thus, everyone on the team has now studied what the ML algorithms from Milestone 5 do.
* Week 6: In the Term Project Forum, and at the Synchronous Meeting for this week, the team can now describe how they applied the material from Milestone 5 in their project. Namely, you have an extra week to try applying the material from each week, as a team.

**Process -- How agile can we be?** If you are not a CS or SE person, you may not know much about "agile" software development. The trend is for such engineering work to be "light" on documentation, and "heavy" on actually trying things to make them work. The "milestones" for this development style come at regular intervals, with a small, well-defined set of goals for each interval.

For this class, we will use an "agile" approach. Your team's goal for each "Milestone" is, as shown above, to take what you learned the week before, and figure out a way to apply it, as much as possible, with your project data.

As noted in the Syllabus, this flow, of moving from individual work on new material, to using it on the team project, includes these ingredients:

1. **Try applying the algorithm to your term project data:** This is the real test -- can you make anything magical happen, on the data you are really interested in. The Moodle Term Project item for each Module is really a second Forum, for you and others to discuss the results of this as well as for your instructors to watch your progress and intervene when needed. It is designed to allow you to reflect on your own experience trying the algorithms, and also to compare that with how others fared.
2. **Contribute to our class "synchronous discussion":** Every week, at a mutually agreed time, we will have a short Google Hangout meeting, for everyone to discuss their progress. Like, around an hour. If enough people are local, we will have a local contingent for this, too. The primary goal of this meeting will be for you to describe your progress on the term project, the item *above*. By hearing how everyone else is doing, we hope you also are encouraged to try things you hear worked for them.

**Reflecting on results:** Starting around Week 4, you will have an additional team goal to compare results you got with results using previous ML algorithms. Your real goal, by the end of class, is to have a perspective on what worked, and what didn't. You will be expected to include that wisdom in your final presentation.

Additionally, toward the end of class, we stuck into the forum assignments a request that you compare your results with those of other teams. We want you to be able to reflect on what was different, about your data, the processes you used, or the field of endeavor that your project applies to, that made some differences in the value of those outcomes. We would like to hear about this aspect, as well, in your presentation.

**Frequently Asked Questions**

1. What do we need as a final project presentation? Our final Synchronous Meeting will be a short read-out, from each person or team, of what you accomplished. There is a place for you to turn in a copy of that presentation.
2. Do we have to achieve some specific goal, like testing data on some number of machine learning algorithms? No, you are not expected to have reached any particular state like that, by the end. Your work should reflect that you put a lot of effort into it. Some of the problems teams pick are, frankly, a lot harder than others. We want you to be able to work on something you like.