EE 660

Homework 6 (Week 8): Type 1 Project Proposal

Posted: Fri., 10/16/2020 Due: Fri., 10/23/2020, 5:00 PM PDT

**This proposal form is for Type 1 projects: Solve a ML problem by implementing a ML system of your own design, using real-world data.**

**Please fill in both the Project Proposal form (pp. 1-2) and the Dataset Information Form (p. 3)**. **This is required of everyone (each team submits one HW6 with all their names on it).** \*All fields except “other comments” are required. In each field, replace instructions (black text) with your descriptions. Preferred format is to enter your answers into the Word version of this form, then convert to pdf before submission. If you prefer to use another app instead of Word, then submit a typed version with each field labeled with its title (“Dataset”, etc.), and submit as a pdf file.

Please note that this proposal will not be graded like a regular homework. The primary purpose is to give you some feedback on your project topic and plans; the scoring on this homework will be primarily based on whether you put in a reasonable effort and whether the content makes good technical sense.

|  |
| --- |
| \*Insert Project Title Here |
| \*Project team: Your name(s) and email address(es) |
|  |
| \*Clear statement of the problem and/or goals. |
| This will guide your work throughout the project period, and help you define when you have completed your project work. (For example, is it to find the best performing system on a particular problem and dataset? Is it to compare different approaches to solve a problem, and draw conclusions from the comparison? Or, …) |
| \*A plan of preprocessing and feature extraction (if applicable) |
| Describe your plans, any preprocessing you expect to try, and similarly for feature extraction. |
| \*A plan of your approach |
| Describe at least a preliminary plan on how you will proceed. This could include, for example, methods and models to try for classification or regression; how you will use datasets (test sets, training sets, etc.); and how you will evaluate your system’s performance.  Your plan can be revised as you do the work, because you will learn more about the problem as you go. But having an initial plan for your approach is helpful. |
| \*A description of any other work of yours that is related to your class project |
| If none, state “none”. Otherwise, briefly summarize your related work here, and state how your EE 660 class project work will be different. Examples of related work you need to describe here: research work on the same or similar topic; work for hire (at your place of employment) on the same or similar topic; work for other classes (projects, etc.) on the same or similar topic. |
| \*If yours is a team project, roughly describe how work will be divided |
| Describe how you are planning to divide up the work and how you will collaborate. (Contributions to all aspects of the project (theory, coding, planning, analysis, writing etc.) will be stated in the final report.) |
| Other Comments |
| Anything else you think should be included. For example, if you see potential problems that might arise, you can mention them here with ideas of how you might deal with them. Or, anything that is particular to your project idea that wasn’t requested in the form fields above. |

EE 660 Dataset Information Form Fall 2020

*Include one form for each dataset you plan to use. (For each dataset’s form, you may continue onto an additional page if necessary.)*

\*Dataset or competition title:

\*Link:

**\*Problem type**: regression, classification/logistic regression

**\*Brief description of dataset and problem domain**:

**\*Number of data points**:

**\*Number of features or input variables**:

\*Feature or input-variable types: numeric (how many variables and what type), categorical (how many variables), other (describe and give how many variables)

\*Label (output) type: numeric, binary categorical, or multi-class categorical

**\*If Label Type is Categorical, is the number of samples significantly unbalanced (maximal variation of more than a factor of 2)**? Yes, No.

**If yes, \*rate as:**

significant (maximal variation is factor of 2 to 10)

major (factor of 10 to 100)

extreme (more than factor of 100))

**\*Has Missing Data**? Yes (give idea of how prevalent, if known), or No.

If yes, try to explain how you will deal with it in the “A plan of preprocessing and feature extraction” section in the Project Proposal

**\*Is the problem/dataset a Kaggle competition (current or past)?** yes or no

**If yes, answer:**

**\*(i) Is the competition current (give the end date), or past**?

**\*(ii) How much information is available on the Kaggle website (e.g., in “kernels” and links therein)**? Briefly describe what type of information and code is available.

**Any other comments on the dataset:**