



CS 6375: Project Ideas

Rishabh Iyer

University of Texas at Dallas

Types of Projects



- **Project Type 1:** Take a specific problem, pose it as a machine learning problem (classification, regression, clustering, ..), train a ML model and study its performance.
- **Project Type 2:** Choose a specific topic (example, SVMs) and study it beyond what was covered in class.
- **Project Type 3:** Implement a machine learning family of algorithms from scratch (linear models, decision trees, neural networks).

Project Expectation



- The basic expectation of a project is that you have spent a significant amount of time on the project on your own trying to solve it.
- It is 25% of the grade.
- The amount of work expected is much more than what you would put in for a single assignments.

Ideal Project Type 1



- An ideal project type 1 would have the following characteristics:
 - Pick the dataset from a good online source.
 - A good study of the characteristics and properties of the dataset.
 - A decent amount of introductory data analysis done (correlations, label and feature distributions, missing values, imbalance, mean/variance of features, etc.)
 - A comprehensive evaluation of different ML algorithms (linear models, decision trees, boosted trees and random forests, neural networks, etc.)
 - Use of meaningful evaluation metrics
 - An analysis of what works and what does not work
 - Lessons Learnt from the Project and How the project helped you.

Ideal Project Type 2



- An ideal project type 2 would have the following characteristics:
 - A deep and non-trivial analysis of a family of approaches beyond what was studied in class.
 - For example, in class, we studied Support Vector Machines
 - But we only covered it for Classification and a little bit of analysis of duality and dual/kernel SVMs
 - For a project, we would expect a much more detailed analysis including:
 - Support Vector Regression
 - Detailed study on Kernel SVMs: what do different kernels accomplish
 - Outlier Detection

Ideal Project Type 3



- An ideal project type 3 would have the following characteristics:
 - A deep and non-trivial implementation of a family of algorithms from scratch.
 - The work here has to be much more than what you would put in for an assignment.
 - For example, consider you study Decision Trees and want to implement a Library for Decision trees and related approaches.
 - In the next assignment, you will implement a decision tree from scratch for classification.
 - But for the project, you will be expected to go beyond this and implement different splitting criteria, different stopping criteria for DTs, DTs for Regression and classification, Random Forests, Gradient Boosted Decision Trees, and so on.

How will we Evaluate



- The project will be due the last week of Nov 2024 (exact date TBA)
- You will submit:
 - A detailed project report (5 – 10 pages)
 - A set of slides describing the project
 - Detailed and commented code if applicable
 - A video (under 3 mins) describing the project in detail
- In the project report, make sure to discuss what you learnt from the project
- Be as detailed as possible. Make sure to write everything you worked on!

Please do not do the following



- Do not Submit a project on a complex research topic that we did not cover in class
 - For example, towards the end of the course, we will look at Neural Networks. It is easy to do a project on simple Neural Networks (maybe up to Convolutional Neural Networks) but not complex topics like Transformers.
- Do not Submit something that could have easily be an assignment (for example, you do a Type 3 project, and just implement a library on Linear Regression)
- Do not Work on the project the last minute and spend 3 days on the project. From the details of the report and slides, it is clear how much time you put into the project!
- Do not use the internet or chat-GPT to put together a project quickly. Again, it is easy for us to catch if you do this. There are also tools (<https://gptzero.me/>) to do that for us!