Planning report for the FTEC final year project of Group O

Project goal:

* By applying the learned index on some synthetic data and compare its time efficiency with other indexs/methods, we will show how fancy the learned index is.

Significance of the project

* Learned Indexes have attracted widespread attention in the industry. Its core idea is to treat data search as a data address prediction problem with key as input, so that the research precipitation of machine learning on prediction problems can be used to optimize data search. Simply put, Learned Indexes is an index model combined with machine learning, which can optimize index structure and search efficiency according to data distribution. As a brand new index model, Learned Indexes may hopefully replace the traditional index of the current DBMS. And the project is mainly about applying the learned index.

Problem statement

* Learned Indexes itself is still in its infancy, and only a simple experiment has initially proved its feasibility. And we are trying to prove its efficiency in out project.

Proposed deliverables

* Apply the learned index on some synthetic data and compare its time efficiency with other indexs/methods

Proposed timelines:

* Semester 1:
  + Start from the toy example given by the professor. (1-d data, uniform/normal distribution, only existence query)
  + Implement the traditional methods and the trick method, and then test their time efficiency on some synthetic data.
  + Read the corresponding paper (LISA).
  + Try to apply it in this toy example, i.e., training a machine learning model to predict the position instead of using a formula.
  + Compare its time efficiency with above methods.
* Semester 2:
  + From the toy example to a more complicated case.
  + Choose one or some of the following directions to make the problem harder.
    - 1-d data -> 2d data
    - Efficient Data insertion / deletion
    - Range query
    - K-NN query
  + Test the time efficiency of the learned index and traditional data structures in the harder case.